Articulatory Organisation in Japanese: an EPG study

By

Mitsuhiro Nakamura

A Thesis submitted to the University of London for the degree of Doctor of Philosophy in Phonetics

Department of Phonetics and Linguistics, University College, London

2001
Abstract

This instrumental phonetic study investigates lingual articulation and coarticulation in standard Japanese. A series of experiments were conducted using electropalatography (the Reading EPG2). The principal aim is to extract EPG correlates of spatiotemporal parameters relevant to the realisation of ten consonantal segments [t, d, n, p, r, s, ç, ts, tc]; two vocoids [i] and [j]; and four palatal(ised) segments ([n, ç, r̃, k̃]). Data for VCV (vowel-consonant-vowel) sequences was obtained from two Japanese speakers. Qualitative and quantitative analyses were performed to identify the basic features of various lingual articulations; the interdependency between the two components of the tongue; and the degree of coarticulatory effects together with their temporal extent. Of particular importance is the identification of the articulatory properties that are unexplained by the feature specification.

A particular focus is placed on the articulatory realisation of palatalisation. The spatiotemporal manifestations of intergestural coordination and timing between the primary articulatory gesture and the (secondary) dorsal gesture are examined in detail. Assuming that palatalisation is a specialised use of the raising gesture of the tongue body, we argue that: (i) the dorsal gesture for palatal(ised) consonants incorporates certain aspects of the articulatory nature of the high front vowel /i/ and the approximant /j/; (ii) there are two ways of resolving antagonistic gestures by the tongue, blending and sequencing, the choice between them depending on the primary articulation; and (iii) speakers employ two contrastive timing strategies for the raising gesture of the tongue dorsum to effect palatalisation. The discussion includes the issue of articulatory complexity.

Based on a detailed parametric specification of vocalic and consonantal gestures, the main discussion is devoted to the nature of the phonetic representation and the phonetic characteristics of standard Japanese. We develop a parametric analysis of certain phonetic processes: child phonology; synchronic sound changes; and vowel devoicing.
Acknowledgements

This dissertation would not have been possible without the guidance and support of John Wells, my supervisor. John suggested that I should teach Japanese phonetics, an occupation which made me aware that Japanese was a language I needed to study very carefully. The discussions we had, invariably went beyond Japanese to include other languages: this reminded me that Japanese does not stand alone but needs to be considered along with other world languages. During my graduate study John has continuously fostered my joy in the wonder of sound structure, and I owe him a great debt of gratitude.

Michael Ashby has been largely responsible for my interest in coarticulation, and has shown me, in his lecture on phonetic theory, that this phenomenon plays an integral part in speech production. In addition, Michael lent me a copy of Edwards (1903), a publication which is not readily available in Japan. I wish to express my thanks to Eva Estebas-Vilaplana for answering many questions about Catalan Phonetics; Hyunsong Chung for the Korean pronunciations; and Gordon Hunter who helped me solve various problems with statistical analysis. Discussions with the attendees of my lectures often inspired ideas which I then developed in this dissertation. Thanks are also due to Molly Bennett, Judith Crompton and Martyn Holland for their encouragement and moral support.

The late Dr. Katrina Hayward of SOAS will always have my warmest respect. This dissertation owes much to her thoughtful and helpful comments. Katrina taught me the nature and importance of the data in phonetic research. I am only sad that I cannot show her my completed dissertation.

I owe thanks to the laboratory technicians Bernard Howard and Jahan Latif for their assistance and for making my time in the lab enjoyable and productive. I also thank Ms. Taeko Maeda for her participation in my EPG experiments.

Many people helped me during the writing of this dissertation. I am most grateful to Dr. Kálmán Bollá of the Hungarian Academy of Science; Dr. Edward Flemming of Stanford University; Dr. Olle Engstrand of the University of Stockholm; Dr. John Esling of the University of Victoria; Dr. Alice Faber of Haskins Laboratories; Dr. Kuniko Kakita of Toyama Prefectural University; Dr. Mariko Kondo of Waseda University; Dr. Haruo Kubozono of Kobe University; and Dr. Daniel Recasens of the Universitat Autònoma de Barcelona. When asked, they all sent me their research papers which I would otherwise not have been able to study.

I also wish to thank my friends in London and in Japan for their encouragement. Special thanks are due to Elizabeth Swan who, as a native English speaker and a diligent student of Japanese, gave me many valuable comments and suggestions.

Finally, I would like to thank my wife, Mariko, for keeping my real life in order and enabling me to finish what I had started.
Table of Contents

Chapter 1 Introduction
1.1 Goals ....................................................................................................................... 16
1.2 Parameters and the Study of Lingual Articulation .......................................... 17
1.3 Phonetic Individuality in Phonetics and Phonology ....................................... 23
1.4 Towards Electropalatographic Specification of Lingual Articulation……. 35
   1.4.1 EPG and Its Limitations .......................................................................... 35
   1.4.2 General Experimental Procedure ........................................................... 36
       1.4.2.1 Zoning of the EPG Artificial Palate .......................................... 37
       1.4.2.2 Data Collection and Reduction .................................................. 40
   1.4.3 Statistical Inference ................................................................................. 40
1.5 Organisation of the Present Study .................................................................... 41

Chapter 2 Phonological and Articulatory Characteristics of the Japanese Sound System
2.1 Introduction .......................................................................................................... 43
2.2 The Architecture of Japanese Phonology ........................................................ 44
   2.2.1 Phonemes, Syllable Structure and Accent ........................................... 44
   2.2.2 Phonological Analysis of the Syllabary ............................................... 47
2.3 Allophonic Variation before the High Vowels ................................................ 51
   2.3.1 Palatalisation, Affrication and Labialisation ...................................... 51
   2.3.2 Linguistic Modelling of Japanese Palatalisation: a historical sketch .......... 54
   2.3.3 CyV Variability ........................................................................................ 58
2.4 Phonetic Segments in Loanword Pronunciation ............................................ 60
2.5 The Production of Vowels and Diphthongs .................................................... 64
   2.5.1 Vowels ....................................................................................................... 64
   2.5.2 Diphthongs ............................................................................................... 69
   2.5.3 Smoothing and Glide Formation: two sides of the same coin?.... 75
2.6 The Articulation of Two-Moraic Consonants ................................................ 81
   2.6.1 Uvular Nasal and its Allophones .......................................................... 81
   2.6.2 The Articulation of Moraic Obstruents ............................................... 87
2.7 Aspects of a Mora-timed Rhythm ................................................................. 90
   2.7.1 The Mora as a Unit of Duration and Temporal Compensation .......... 90
   2.7.2 The Mora as a Pattern of Articulatory Coordination ......................... 98
2.8 Summary and Conclusion ................................................................................. 102
Chapter 5 Lingual Articulation and Coarticulation of Voiceless Fricatives [s, ç, ç] and Affricates [ts, tc]

5.1 Introduction.................................................................................................... 198
5.2 Previous Studies............................................................................................ 198
  5.2.1 Fricative Articulations...................................................................... 198
  5.2.2 Affricate Articulations...................................................................... 202
5.3 Data Collection and Analysis..................................................................... 204
  5.3.1 Speech Items...................................................................................... 204
  5.3.2 Measurements.................................................................................... 205
5.4 Results of Experiment 1: fricative articulations...................................... 207
  5.4.1 Configurational Characteristics of [s, ç, ç].................................. 207
    5.4.1.1 [s]............................................................................................ 207
    5.4.1.2 [ç]............................................................................................ 208
    5.4.1.3 [ç]............................................................................................ 209
    5.4.1.4 Constriction Widths of [s, ç, ç]........................................... 210
  5.4.2 Coarticulatory Characteristics of [s, ç, ç]...................................... 215
    5.4.2.1 [s]............................................................................................ 215
    5.4.2.2 [ç]............................................................................................ 218
    5.4.2.3 [ç]............................................................................................ 221
5.5 Results of Experiment 2: affricate articulations...................................... 224
  5.5.1 Movement Characteristics of [ts] and [tc] ..................................... 224
  5.5.2 Coarticulatory Characteristics of [ts] and [tc] ............................... 230
5.6 Intergestural Coordination............................................................................ 238
5.7 Summary and Conclusion............................................................................. 240

Chapter 6 Gestural Coordination and Timing in Japanese Palatalisation

6.1 Introduction..................................................................................................... 243
6.2 The Phonology and Tongue Dynamics of Japanese Palatalisation............ 244
  6.2.1 Palatalisation as a Language-specific Coarticulatory Strategy... 244
  6.2.2 The ‘Complexity’ Debate................................................................. 248
6.3 Data Collection and Analysis...................................................................... 251
  6.3.1 Speech Items...................................................................................... 251
  6.3.2 Measurements.................................................................................... 252
6.4 Results and Discussion................................................................................... 255
  6.4.1 Experiment 1: the canonical EPG patterns of [i] and [j].............. 255
  6.4.2 Experiment 2: the target consonants
    in the palatalising environments....................................................... 265
List of Figures

Figure 1-1: EPG Artificial Palate of Speaker MN ............................................................ 38
Figure 1-2: EPG Artificial Palate of Speaker TM ............................................................. 39
Figure 2-1: Standard Japanese Five Vowels on a Polar Co-ordinate Vowel Diagram.... 65
Figure 2-2: Mean F1-F2 Values for the Japanese Five Vowels in Prose Passages (Keating & Huffman, 1984) ................................................ 67
Figure 2-3: Mean F2 Frequency Ranges for Five American English Diphthongs and Four Japanese Diphthongs at the Three Different Speaking Rates (Dolan & Mimori, 1986) ........................................ 73
Figure 2-4: Spectrograms of /banta/, /banda/, and /banra/ Produced by Speaker MN ................................................................................. 95
Figure 2-5: EPG Contact Patterns of /banta/, /banda/, and /banra/ Produced by Speaker MN ................................................................................. 96
Figure 4-1: Arrangement of Electrodes and Three Regions ............................................. 132
Figure 4-2: Measurement Points in Time for the EPG Data ............................................ 132
Figure 4-3: Mean Percentage Values of Linguopalatal Contact in the Three Regions by the Target Consonants for the Two Speakers......................... 135
Figure 4-4-1: Linguopalatal Configurations at the MAX Point for /t, d, n/ in the Vowel Symmetrical Contexts /i_i/, /a_a/ and /u_u/ (Speaker MN)... 138
Figure 4-4-2: Linguopalatal Configurations at the MAX Point for /t, d, n/ in the Vowel Symmetrical Contexts /i_i/, /a_a/ and /u_u/ (Speaker TM)… 139
Figure 4-5: Linguopalatal Configurations at the MAX Point for /r/ in the Vowel Symmetrical Contexts /i_i/, /a_a/ and /u_u/ ............................ 142
Figure 4-6-1: Anticipatory Effects in the Three Regions for the Consonants [t, d, n, p, r] Produced by Speaker MN......................................... 146
Figure 4-6-2: Anticipatory Effects in the Three Regions for the Consonants [t, d, n, p, r] Produced by Speaker TM......................................... 147
Figure 4-7-1: Carryover Effects in the Three Regions for the Consonants [t, d, n, p, r] Produced by Speaker MN......................................... 148
Figure 4-7-2: Carryover Effects in the Three Regions for the Consonants [t, d, n, p, r] Produced by Speaker TM......................................... 149
Figure 4-8-1: [t, d, r] at the Three Temporal Points of Articulatory Closure in /aCi/ and /iCa/ Contexts (Speaker MN)........................................ 153
Figure 4-8-2: [t, d, r] at the Three Temporal Points of Articulatory Closure in /aCi/ and /iCa/ Contexts (Speaker TM)........................................ 154

8
Figure 4-9-1: [t, d, r] at the Three Temporal Points of Articulatory Closure  
in /aCu/ and /uCa/ Contexts (Speaker MN) .................................................. 155

Figure 4-9-2: [t, d, r] at the Three Temporal Points of Articulatory Closure  
in /aCu/ and /uCa/ Contexts (Speaker TM) .................................................. 156

Figure 4-10-1: [t, d, r] at the Three Temporal Points of Articulatory Closure  
in /iCu/ and /uC/i Contexts (Speaker MN) ................................................... 157

Figure 4-10-2: [t, d, r] at the Three Temporal Points of Articulatory Closure  
in /iCu/ and /uC/i Contexts (Speaker TM) ................................................... 158

Figure 4-11-1: [n] at the Three Temporal Points of Articulatory Closure  
in the Vowel Asymmetrical Contexts (Speaker MN) .................................... 163

Figure 4-11-2: [n] at the Three Temporal Points of Articulatory Closure  
in the Vowel Asymmetrical Contexts (Speaker TM) .................................... 164

Figure 4-12-1: [n] at the Three Temporal Points of Articulatory Closure  
in the Vowel Asymmetrical Contexts (Speaker MN) .................................... 168

Figure 4-12-2: [n] at the Three Temporal Points of Articulatory Closure  
in the Vowel Asymmetrical Contexts (Speaker TM) .................................... 169

Figure 4-13: Mean Articulatory Closure Durations for [t, d, n, j, r]  
in the Vowel Symmetrical Contexts (Pooled-speaker Context) .................... 172

Figure 4-14: Mean Articulatory Closure Durations and Standard Deviations  
for [t, d, n, j, r] in the Vowel Asymmetrical Contexts .................................. 174

Figure 4-15: Mean Articulatory Closure Durations and Standard Deviations  
for [t, d, n, j] in the Vowel Asymmetrical Sequences  
(Pooled-speaker Contexts) ................................................................. 176

Figure 4-16: Mean Articulatory Closure Durations and Standard Deviations for [r]  
in the Vowel Asymmetrical Sequences .................................................... 176

Figure 4-17: Mean Front Region Durations for [t, d, n, j]  
in the Vowel Asymmetrical Contexts ..................................................... 179

Figure 4-18: Mean Front Region Duration for [r] in  
the Vowel Asymmetrical Contexts ...................................................... 179

Figure 4-19: Relationship between Closure Duration and  
Gestural Duration: all the consonants across the speakers ....................... 181

Figure 4-20: Relationship between Closure Duration and  
Gestural Duration: each consonant across the speakers ......................... 182

Figure 4-21: Relationship between Front Contact Degree and  
Articulatory Closure Duration: all the consonants across the speakers ...... 184

Figure 4-22: Relationship between Front Contact Degree and  
Articulatory Closure Duration: each consonant across the speakers .......... 185
Figure 4-23: [t, d, r]: Relationship between Front Contact Degree and Articulatory Closure Duration

Figure 4-24: Relationship between Front Contact Degree and Central Contact Degree (All the consonants across the speakers)

Figure 4-25: Relationship between Front Contact Degree and Central Contact Degree (Each consonant across the speakers)

Figure 4-26: Articulatory Trajectories of /iti/ and /idi/ Syllables

Figure 4-27: Articulatory Trajectories of /iri/ Syllables

Figure 5-1: Arrangement of Electrodes and Three Regions

Figure 5-2: Linguopalatal Configurations at the MAX Point for [s] in the Vowel Symmetrical Contexts

Figure 5-3: Linguopalatal Configurations at the MAX Point for [ç] in the Vowel Symmetrical Contexts

Figure 5-4: Linguopalatal Configurations at the MAX Point for [ç] in the Vowel Symmetrical Contexts

Figure 5-5: Regression Analysis of the Constriction Place and Width across Speakers

Figure 5-6: Spectograms of [s], [ç], and [ç] in the /a_a/ Symmetrical Context Produced by Speaker MN

Figure 5-7: Spectograms of [s], [ç], and [ç] in the /a_a/ Symmetrical Context Produced by Speaker TM

Figure 5-8: Linguopalatal Configurations at the MAX Point for [s] in the Asymmetrical Contexts

Figure 5-9: Mean Percentage Values of Linguopalatal Contact for [s] in the Asymmetrical Contexts

Figure 5-10: Linguopalatal Configurations at the MAX Point for [ç] in the Asymmetrical Contexts

Figure 5-11: Mean Percentage Values of Linguopalatal Contact for [ç] in the Asymmetrical Contexts

Figure 5-12: Linguopalatal Configurations at the MAX Point for [ç] in the Asymmetrical Contexts

Figure 5-13: Mean Percentage Values of Linguopalatal Contact for [ç] in the Asymmetrical Contexts

Figure 5-14: EPG Patterns of Japanese [atsa] Produced by the Two Speakers

Figure 5-15: EPG Patterns of Japanese [atçæ] Produced by the Two Speakers

Figure 5-16: Linguopalatal Configurations of [ts] at the Point of MAX, Closure Onset and Fricative Onset (Symmetrical Contexts)
Figure 5-17: Linguopalatal Configurations of [tc] at the Point of MAX, Closure Onset and Fricative Onset (Symmetrical Contexts) .......................... 229

Figure 5-18: Effects of the Preceding and Following Vowels on [ts] (Symmetrical Contexts) .................................................. 230

Figure 5-19: Linguopalatal Configurations of [ts] at the Point of MAX, Closure Onset and Fricative Onset (Asymmetrical Contexts; Speaker MN) .................. 231

Figure 5-20: Linguopalatal Configurations of [ts] at the Point of MAX, Closure Onset and Fricative Onset (Asymmetrical Contexts; Speaker TM) .................. 232

Figure 5-21: Effects of the Preceding and Following Vowels on [ts] (Asymmetrical Contexts) ........................................... 234

Figure 5-22: Linguopalatal Configurations of [tc] at the Point of MAX, Closure Onset and Fricative Onset (Asymmetrical Contexts; Speaker MN) .................. 235

Figure 5-23: Linguopalatal Configurations of [ts] at the Point of MAX, Closure Onset and Fricative Onset (Asymmetrical Contexts; Speaker TM) .................. 236

Figure 5-24: Regression Analysis for the Fricatives [s, ç] and the Fricative Elements [s', ç'] of the Affricates [ts, tc]: all the consonants across the speakers ................................................................. 238

Figure 5-25: Regression Analysis for the Fricatives [s, ç, ç] and the Fricative Elements [s', ç', ç'] of the Affricates [ts, tc]: each consonant across the speakers................................................................. 239

Figure 6-1: Three Regions and Patterns of their Interactions ............................................. 252

Figure 6-2: Linguopalatal Configurations at the MAX Point for /i/ and /j/ in the Vowel Symmetrical Contexts .................................................. 256

Figure 6-3: Means Values and Standard Deviations of Central Region Contact for /i/ and /j/ in the Pooled-speaker and Pooled-speaker-vowel Contexts.... 257

Figure 6-4-1: Front and Central Trajectories of /i/ and /j/ (Speaker MN) .................. 260

Figure 6-4-2: Front and Central Trajectories of /i/ and /j/ (Speaker TM) .................. 261

Figure 6-5: Means and Standard Deviations of the Central Region Duration for /i/ and /j/ ........................................................................ 262

Figure 6-6: Linguopalatal Configurations at the MAX Point for [p, ç, l, k] (Speaker MN) .................. 266

Figure 6-7: Linguopalatal Configurations at the MAX Point for [p, ç, l, k] (Speaker TM) .................. 267

Figure 6-8: Mean Percentage Values and Standard Deviations of the Central Region Contact for [p, ç, l, k] (Pooled-speaker Context) .......................... 268

Figure 6-9: COG Peak Values for the Four Palatal(ised) Consonants averaged across the Two Speakers ................................................................. 272
Figure 6-10: COG Contours averaged over Five Repetitions
for the Four Palatal(ised) Consonants ......................................................... 273

Figure 6-11: Articulatory Trajectories of the Syllables with [ɲ] .............................. 276
Figure 6-12: Articulatory Trajectories of the Syllables with [ç] ............................. 277
Figure 6-13: Articulatory Trajectories of the Syllables with [ɾ̝] ............................. 278
Figure 6-14: Articulatory Trajectories of the Syllables with [k̂] ............................. 279

Figure 6-15: Mean Sequence Durations and Standard Deviations
for the Four Palatal(ised) Consonants ......................................................... 280

Figure 6-16: Mean Front Region Durations and Standard Deviations
for the Four Palatal(ised) Consonants ......................................................... 282

Figure 6-17: Mean Central Region Duration and Standard Deviations
for the Four Palatal(ised) Consonants ......................................................... 282

Figure 6-18: Mean Peak Intervals and Standard Deviations
for the Four Palatal(ised) Consonants ......................................................... 286

Figure 6-19: Mean Total and Overlap Durations above the 50% Cut-off Point ....... 289

Figure 6-20: Mean Sequence Overlap above the Point of 50% Linguopalatal Contact .. 289

Figure 6-21: Mean Percent-anticipation of the Central Region for [ɲ, ç, ɾ̝] .............. 290

Figure 6-22: Mean Percent-anticipation of the Central Region for [k̂] ................. 290

Figure 6-23: Typical EPG Patterns of the [ak̂i] Syllable ....................................... 293

Figure 7-1: Spectrogram and EPG pattern of [acı] Produced by Speaker MN ....... 309

Figure 7-2: Spectrogram and EPG pattern of [aç̂u] Produced by Speaker MN ...... 310

Figure 7-3: Spectrogram and EPG pattern of [aci] Produced by Speaker TM ....... 311

Figure 7-4: Spectrogram and EPG pattern of [aç̂u] Produced by Speaker TM ...... 312

Figure 7-5: Spectrogram, Laryngogram and EPG Pattern of [stui,k̂i] ..................... 330

Figure 7-6: Spectrogram, Laryngogram and EPG Pattern of [stui,ɾ̝i] ..................... 331

Figure 7-7: Spectrogram, Laryngogram and EPG Pattern of ['stui,k̂i] ................. 332

Figure 7-8: EPG Patterns for the [stui,k̂i] Productions ......................................... 333

Figure 7-9: Spectrogram, Laryngogram and EPG Pattern of [k̂i,stu] ..................... 335

Figure 7-10: Spectrogram, Laryngogram and EPG Pattern of the Partially Devoiced [k̂i,stu] ................................................................. 336

Figure 7-11: Spectrogram, Laryngogram and EPG Pattern of [k̂i,zu] ..................... 337

Figure 7-12: Spectrogram, Laryngogram and EPG Pattern of ['k̂i,stu] ................. 338

Figure 7-13: Spectrogram, Laryngogram and EPG Pattern of [çi, to] ..................... 340

Figure 7-14: EPG Pattern of [çi, to] ................................................................. 341
**List of Tables**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1-1:</td>
<td>Articulatory Settings of Southern British English and French (Honikman, 1964)</td>
<td>28</td>
</tr>
<tr>
<td>Table 1-2:</td>
<td>Articulatory Settings of Southern British English and D&lt;sup&gt;e&lt;/sup&gt;h (Jenner, 1987)</td>
<td>29</td>
</tr>
<tr>
<td>Table 1-3:</td>
<td>Articulatory Settings of General American English and Danish (Collins &amp; Mees, 1995)</td>
<td>30</td>
</tr>
<tr>
<td>Table 1-4:</td>
<td>Articulatory Settings of Southern British English and Norwich (Trudgill, 1974)</td>
<td>31</td>
</tr>
<tr>
<td>Table 1-5:</td>
<td>Articulatory Settings of English spoken in Liverpool (Knowles, 1978) and in Cardiff (Collins &amp; Mees, 1990)</td>
<td>31</td>
</tr>
<tr>
<td>Table 2-1:</td>
<td>Phonological Inventory of Japanese</td>
<td>44</td>
</tr>
<tr>
<td>Table 2-2:</td>
<td>Basic Accent Patterns of Japanese</td>
<td>46</td>
</tr>
<tr>
<td>Table 2-3:</td>
<td>Examples of Basic Accent Patterns</td>
<td>46</td>
</tr>
<tr>
<td>Table 2-4:</td>
<td>Phonetic and Phonological Interpretation of the Japanese Syllabary</td>
<td>47</td>
</tr>
<tr>
<td>Table 2-5:</td>
<td>The Palatalised-vowel Representation of the Japanese Syllabary (Hattori, 1949)</td>
<td>55</td>
</tr>
<tr>
<td>Table 2-6:</td>
<td>The Palatalised-consonant Representation of the Japanese Phonological Inventory (Uemura, 1972)</td>
<td>55</td>
</tr>
<tr>
<td>Table 2-7:</td>
<td>The Distributional Characteristics of Sharpness (McCawley, 1968)</td>
<td>56</td>
</tr>
<tr>
<td>Table 2-8:</td>
<td>Emergent Segments in Loanword Pronunciation</td>
<td>62</td>
</tr>
<tr>
<td>Table 4-1:</td>
<td>Two-way ANOVA Results for the Effects of the Consonant Types and the Vowel Contexts on Linguopalatal Contact of [t, d, n, r]</td>
<td>135</td>
</tr>
<tr>
<td>Table 4-2:</td>
<td>Two-way ANOVA Results for the Effects of the Consonant Types and the Vowel Contexts on Linguopalatal Contact of [t, d, n, r] in /a_a/ and /u_u/ Contexts</td>
<td>135</td>
</tr>
<tr>
<td>Table 4-3:</td>
<td>One-way ANOVA Results for the Effects of the Consonant Types and the Vowel Contexts on Linguopalatal Contact of [t, d, n, r]</td>
<td>137</td>
</tr>
<tr>
<td>Table 4-4:</td>
<td>One-way ANOVA Results for the Effects of the Consonant Types and the Vowel Contexts on Linguopalatal Contact of [t, d, n, r] in /a_a/ and /u_u/ Contexts</td>
<td>144</td>
</tr>
<tr>
<td>Table 4-5:</td>
<td>Two-way ANOVA Results for Anticipatory and Carryover Effects on [t, d, n, r]</td>
<td>150</td>
</tr>
<tr>
<td>Table 4-6:</td>
<td>Two-way ANOVA Results for Anticipatory and Carryover Effects on [t, d, n, r] in /VCa/ and /VCu/ Sequences</td>
<td>151</td>
</tr>
</tbody>
</table>
Table 4-7: One-way ANOVA Results for Anticipatory and Carryover Effects on \( [t, d, r] \) at the MAX Linguopalatal Contact ............................... 159

Table 4-8: One-way ANOVA Results for Anticipatory and Carryover Effects on \([p]\) at the MAX Linguopalatal Contact ...................................................... 165

Table 4-9: One-way ANOVA Results for Anticipatory and Carryover Effects on \([n]\) in the /Vn\(a/\) and /Vn\(u/\) Sequences ...................................................... 170

Table 4-10: Mean Values and Standard Deviations of the Articulatory Closure Duration .............................................................. 171

Table 4-11: Two-way ANOVA Results and the Durational Relationship between the Consonants ................................................................................................. 173

Table 4-12: Mean Values and Standard Deviations of the Articulatory Closure duration in Asymmetric Contexts .............................................................. 173

Table 4-13: t-test Results for the Contextual Effects on Articulatory Closure Duration ........................................................................................................... 176

Table 4-14: Mean Values and Standard Deviations of the Duration of Consonantal Gesture ........................................................................................................... 177

Table 5-1: Mean Values and Standard Deviations of Constriction Width for \([s], [ç] \) and \([ç]\) ........................................................................................................... 211

Table 5-2: One-way ANOVA Results for Anticipatory and Carryover Effects on \([s]\) .................................................................................................................. 217

Table 5-3: One-way ANOVA Results for Anticipatory and Carryover Effects on \([ç]\) .................................................................................................................. 220

Table 5-4: One-way ANOVA Results for Anticipatory and Carryover Effects on \([ç]\) .................................................................................................................. 223

Table 5-5: One-way ANOVA Results for the Effects of \(V1\) on \([t']\) and \(ß\) those of \(V2\) on \([ß']\) ......................................................................................................... 233

Table 6-1: Two-way ANOVA Results for the Effects of Vowel Types and Flanking Vowels in Pooled-speaker Context .............................................................. 257

Table 6-2: One-way ANOVA Results for the Effects of the Contextual Vowels on Linguopalatal Contact for \(/i/\) and \(/j/\) .............................................................. 258

Table 6-3: Cross-speaker Differences for the Degree of Central Region Contact .............................................................. 258

Table 6-4: t-test Results for Differences in Region Duration between \(/i/\) and \(/j/\) .............................................................. 262

Table 6-5: One-way Anova Results for the Effects of the Vowel Contexts on the Region Duration in Pooled-vowel Context .............................................................. 263
Table 6-6: ANOVA and t-test Results for the Durational Differences in the Three Articulatory Stages between /i/ and /j/ ................................. 264
Table 6-7: Effects of the speaker and Vowel Contexts on the Contact Amount in the Central Region ................................................................. 268
Table 6-8: Differences between the Vowel Contexts and between the Syllable Pairs for the Contact Amount in the Central Region ...................... 269
Table 6-9: Comparison of the Contact Amount in the Central Region between /i, j/ and the Palatal(ised) Consonants ........................................ 271
Table 6-10: Effects of the Speaker and the Vowel Contexts on the COG Peak Values .................................................................................... 274
Table 6-11: Means and Standard Deviations for the COG Peak Values .......... 274
Table 6-12: Effects of the Vowel Contexts on Sequence Duration .................. 281
Table 6-13: Effects of the Vowel Contexts on the Front (Back) Region Duration ... 282
Table 6-14: Effects of the Vowel Contexts on the Central Region Duration ........ 283
Table 6-15: Effects of the Vowel Contexts on Peak Interval ......................... 286
Table 6-16: Mean Peak Intervals and Differences between the Syllable Types ...... 287
Table 6-17: Differences in Sequence Overlap between the Ci and CyV Syllables.... 291
Table 6-18: Differences in Percent Anticipation between the Ci and CyV syllables.... 291
Chapter 1

Introduction

1.1 Goals
This experimental phonetic study investigates lingual articulation and coarticulation in
standard Japanese (hereafter Japanese). It concerns the coarticulatory and coordinative
strategies used by a speaker in order to produce phonetic segments involving the tongue
tip/blade and dorsum. Previous studies of these aspects in Japanese are very limited in scope
and coverage. There are few systematic attempts to describe the spatio-temporal coordinative
pattern of vocalic and consonantal gestures and little empirical evidence. I believe there is a
crucial need for an explanation of the phonological system of the language to be established
from a quantitative and qualitative empirical basis. One path to an adequate explanation and
understanding of a spoken language is the systematic investigation of the events in the vocal
tract. Given the facts about the exact positions and movements of the articulators, it should be
possible to relate the physical processes in speech production to the linguistic knowledge of a
language.

This study, using the electropalatographic technique (hereafter EPG), examines
general and language-specific characteristics of lingual articulation, and explores the
relationship between the linguistic categories and the systematic articulatory controls
underlying them. We set out the five goals of the current study as follows:

(i) To present a reasonably thorough review of the descriptive and instrumental data
which provides potential indications of speakers’ control strategies used for the
production of a given phonetic segment;

(ii) To develop a parametric (dynamic) model for relating the phonological structure to
the articulatory organisation; an attempt is made to find out the articulatory basis of
the Japanese sound system; various segmental processes are represented in terms of
the recurring articulatory patterns observed;

(iii) To describe a series of EPG experiments in which the articulation and coarticulation
of various lingual consonants are systematically investigated as a function of vowel
context; an effort is made to extract EPG correlates of lingual activities relevant to the
realisation of ten consonantal segments [t, d, n, j, r, s, ç, ç, ts, tç]; two vocoids [i, j];
and palatalisation; and vowel devoicing;
(iv) To examine various spatio-temporal parameters for lingual gestures, previously proposed but given little attention to; of particular importance is the identification of the articulatory properties that are unexplained by the feature specification;

(v) To practise a parametric approach for analysing speech and describing the phonetic and phonological processes in Japanese.

The specific aims and objectives are: to determine the basic articulatory features of Japanese lingual consonants; to give a detailed account of the articulatory and linguistic nature of Japanese palatalisation; and to explore the relationship between articulatory gestures and linguistic structure. EPG experiments are designed not only to investigate the articulatory and coarticulatory characteristics of the consonantal articulations, but also to test various hypotheses that are formulated for a given class of phonetic segments and phonetic phenomena. Based on detailed parametric specifications of vocalic and consonantal gestures, the greater part of the discussion in this study is devoted to: the nature of phonetic representation; the phonetic characteristics of the individual language; and the place of coarticulation in the grammar of a language. The empirical data enables us to gain an insight, not only into the mechanisms of speech production, but also into some connected speech processes and phonetic variability.

The rest of this chapter discusses two important notions that serve as the starting point for this study. Section 1.2 discusses the nature of the parametric (dynamic) approach to spoken language. In order to describe the articulation from the perspective of speakers’ control strategy, we consider the significances and consequences of a parametric view of phonetic and phonological structure. Section 1.3 examines the nature and place of phonetic individuality, or ‘articulatory settings’ (Honikman, 1964), in the linguistic modelling of speech production. We argue that the articulatory settings have an important role in the language-specific patterning of coarticulatory activities. The discussion will show the importance of the detailed study of lingual articulation and will offer a very broad framework to further experimental investigations in the present study. After providing those preliminary considerations, the methodological procedures used in conducting the EPG experiments are explained in Section 1.4. We describe the general design of the instrumentation, data collection and statistical analyses. Finally, the organisation of this study is presented in Section 1.5.

1.2 Parameters and the Study of Lingual Articulation
Abercrombie (1965), Henderson (1965), Laver (1970) and Tench (1978) have discussed the importance of characterising speech as dynamic articulatory process. A parametric (dynamic) view of speech, as opposed to a postural (static or segmental) view, is characterised as ‘a view
of the study of speech sounds in which the continuous activity of the organs of speech in the production of a stretch of speech is highlighted' (Tench, 1978). In contrast, the traditional approach typifies speech as a series of static postures, or absolute segmental units, although the inherent dynamic nature of speech is implicitly, or explicitly, recognised. Although Laver (1970, 56) mentions that 'the dynamic concept of speech production is now one of the foundations of modern articulatory phonetics', it is necessary to describe the main points for the clarification of our position.

The segmental view of speech dominated over the dynamic view in the history of phonetics. There are two major reasons for this. First, the lack of appropriate instrumental techniques makes it difficult to study the dynamic activities of the speech organs (Laver, 1970; Tench 1978). That is, whereas the continuous nature of speech is acknowledged, there is no way to observe it effectively, and hence the static posture at a certain stage of articulation is emphasised with the alphabetical representation. Second, there is a close connection between phonetics and language teaching/learning (Tench, 1978). It is in fact practical to identify, analyse and isolate a particular problem of learners' pronunciation in terms of the segmental sounds, rather than the continuous flow of speech. Accordingly, this makes the postural view of speech unavoidable. It influences the methodological aspects of phonetics; analysing and describing the speech sounds by absolute categories (Abercrombie, 1965; Laver, 1970; Tench, 1978). Firth (1948; 857) criticises the strict segmental view of speech on the grounds that '[t]he alphabetic notation employed does not rest mainly on modern acoustic and physiological categories but largely on fictions, some of them very ancient, set up by grammatical theory and adapted for the statement of the findings by listening and looking, and by reference to the sense of posture and movement of the listener'. Firth acutely suggests that 'utterances are events, not facts' (ibid.). Similar observations can also be found in Abercrombie (1965, 1991) and Repp (1981).

The crucial difference between the parametric approach and the postural approach lies in the conceptualisation of time dimension that embodies a sequence of various phonetic segments. Abercrombie (1965; 123f.) mentions that '[t]he division of speech into phoneme-representing segments represents a division at right-angles to the time axis, whereas the division into parameters is a division parallel to the time axis'. While the parametric view of speech effectively shifts the main focus to the time-varying parameters of articulation (Laver, 1970), there remains the crucial question of what a parameter is; and how many parameters are necessary for the description of sound-producing activities.

Abercrombie (ibid.) offers the answer to the first question by defining a parameter as 'a variable, an ingredient which is continually present but changing in value'. More specifically Ladefoged (1979, 1980) suggests that the articulatory parameters are required to be controlled in speech production, rather than actual movements of the vocal organs: they
specify linguistically contrastive events in the vocal tract.

In contrast, the answer to the second question, the number of parameters, is not clear: there is little consensus for the number of parameters. For instance, Abercrombie (1965; 123f.) describes a tentative list as follows:

1. in the respiratory system
   (a) the syllable-pulse process
   (b) the pulse-reinforcing, or stress process

2. in the phonatory system
   (c) phonation-type control
   (d) on/off switching of voicing
   (e) voice-pitch variation

3. in the articulatory system
   (f) velic valve-action
   (g) tongue-body movements
   (h) tongue-tip movements
   (i) lip movements
   (j) jaw movements

Tench (1978; 37) simplifies the above list and adds different degrees of stricture: complete and partial (fricative) closure for supralaryngeal articulators; close, half-close, half-open and open for the tongue body; complete closure, partial closure (vibration) and complete opening for the laryngeal activity. Ladefoged (1980) proposes the provisional seventeen parameters:

4. (a) Front raising
   (b) Back raising
   (c) Tip raising
   (d) Tip advancing
   (e) Pharynx width
   (f) Tongue bunching
   (g) Tongue narrowing
   (h) Tongue hollowing
   (i) Lip height
   (j) Lip width
   (k) Lip protrusion
   (l) Velic opening
   (m) Larynx lowering
   (n) Glottal aperture
   (o) Phonation tension
   (p) Glottal length
   (q) Lung volume decrement

The lingual gesture is subcategorised into eight parameters. Here, we shall concentrate on the parameterisation of the tongue.

The tongue is ‘an extremely flexible structure’ and ‘has neither bones nor joints, yet it executes enormously complex movements such as bending, extension, retraction, torsion and leverage’ (Stone, 1991). The movements of the tongue surface, however, are commonly characterised as a small number of functional divisions that are activated in the horizontal and vertical dimensions. Hardcastle (1976; 100) suggests the two functional divisions of the tongue (i.e. tip/blade and body) and the four aspects of their movements. Their particular combinations yield seven parameters for defining the tongue configurations and motions: (i) horizontal forwards-backwards movement of the tongue-body; (ii) vertical upwards-

1) We may note, in passing, the following suggestion by Abercrombie (1965; 124):

...I am not suggesting that we normally listen to these parameters. We hear the medium as a single unanalysed continuing noise, fluctuating in quality. We listen, perhaps, in terms of three parameters representing three expressive systems: (i) articulatory patterns; (ii) intonation patterns; and (iii) register (‘voice quality’) variations. The three systems operate in speech quite independently of each other, and they are listened to as language, plus indications of mood, character, and so on.
downwards movement of the tongue body; (iii) horizontal forwards-backwards movement of the tongue tip-blade; (iv) vertical upwards-downwards movement of the tip-blade; (v) transverse cross-sectional configuration of the tongue body (i.e. convex-concave); (vi) transverse cross-sectional configuration extending throughout the whole length of the tongue, particularly the tip and blade (i.e. degree of central grooving); and (vii) surface plan of the tongue dorsum (i.e. spread or tapered). Stone (1991), in her three-dimensional model, divides the tongue into five crosswise functional segments that are further divided into five lengthwise segments: the sagittal segments are labelled as anterior, middle, dorsal, posterior and root; the cross-sectional segments are composed of one mid segment (i.e. midsagittal plane) and the two lateral segments on both sides. These cross-sectional segments involve the four movements to generate a particular tongue shape: elevation, lowering, medial compression and lateral extension. Thus, it is possible to identify the general configuration of the tongue based on the simple partitioning of the tongue surface and on the movements associated with it.

Support for this assumption comes from various studies in articulatory synthesis and in experimental phonetics. In his study of articulatory synthesis, based on x-ray tracings, Mermelstein (1973) considers the tongue blade and body as separate but interdependent articulators, the movements of which are specified in two distinct modes: the tongue movements correlate with the jaw; and they are relatively independent from the jaw.

Harshman et al. (1977) and Ladefoged et al. (1978) quantify the articulation of ten English vowels, spoken by five speakers, in terms of 13 superimposed grid lines on the tongue. Harshman et al. demonstrate that various shapes of the tongue surface can be decomposed into the two factors, upwards and downwards movements: the tongue root moves forwards as the tongue front is raised upwards; and the forward movements of the tongue root decrease as the tongue body moves backwards. Jackson (1988) reports similar results for Icelandic. Kröger (1993) applies a gestural model (Browman & Goldstein, 1989), in which specifications of ‘articulatory gestures’ serve as the inputs, and synthesises various realisations of /mɪt dɛm bɔt/ ‘by boat’. This model uses the combinations of the height and position of the tongue tip and body for the control parameters of the lingual gesture. Using an x-ray microbeam system to track the articulator movements, Lindau-Webb & Ladefoged (1989) show that the tongue configuration from the tip to just above the root of the epiglottis is reconstructed accurately from two pellets placed on the tongue surface. Also, in their study using an electromagnetic articulograph (EMA), Nguyen et al. (1994) demonstrates that the relevant spectral shapes of English [s] and [ʃ] are regenerated from the tongue positioning in the high-low and front-back dimension.

In the current study the articulation of various lingual consonants is characterised by the two-component model of the tongue physiology: the tip/blade component and the dorsum...
component. The EPG contact patterns obtained in our experiments are quantitatively and qualitatively analysed in terms of the two functional divisions of the tongue. This is supported by a notable contribution by Nguyen et al. (1996). In their EPG study (the Reading system) Nguyen et al. attempt to model the configuration of linguopalatal contacts in terms of a combination of a small number of EPG parameters. It was found that variation in the contact pattern was essentially due to tongue contacts in the alveolar and palatal regions. If we assume that the part of the tongue making a contact against the palate lies directly under that location, it is possible to interpret the results of Nguyen et al. as showing that the main control parameters are associated with those two regions. Also, this provides evidence for the hypothesis that the two components of the tongue, tip/blade and dorsum, are independently controllable articulators. The zoning of the EPG palate in the current study will be discussed later in section 1.4.

The two-component model of the tongue enables us to identify the lingual consonants as coordinative patterns between the tongue tip/blade and the dorsum. It has been common practice to specify the consonantal articulations by their primary constriction between the palate and the tongue. A voiceless stop [t], for instance, is described as having a complete occlusion by the tongue tip/blade at the alveolar region. Thus, the difference between [t] and [d] is the presence or absence of voicing. The activity of the tongue dorsum is not usually mentioned for these coronal consonants such as velars. The reverse is true for the dorsal consonants such as palatals. This style of the representation is also inherited in the phonological feature specifications of the articulator-based models: apart from the manner and voicing features, the distinctive feature is selected in terms of the articulator that forms the primary constriction or narrowing. In contrast, the two-component model characterises the interdependency between the two components in the production of the given consonant. Each consonant may exhibit different degrees of interdependency, or coupling, between the tongue tip/blade and the dorsum. This has important consequences for (lingual) coarticulation: the phenomenon by which 'the movements of articulatory organs overlap in time with those of different articulatory organs or different parts of the same organ' (Farnetani et al., 1989).

The magnitude and temporal extent of consonantal and vocalic coarticulation can be predicted from the degree of involvement of the tongue dorsum in the formation of complete occlusion or narrowing constriction. Since moving the tongue dorsum also produces the vowel, the degree of the vowel-dependent coarticulatory effects may vary with the given consonantal articulation: the variability in degree reflects the production constraint on the tongue dorsum. This idea, which is based on the concept of 'coarticulatory resistance' (Bladon & Al-Bamerni, 1976), is developed in a series of EPG studies by Recasens (1984a, b, 2)

---

2) One notable exception is the retroflex articulation.
c, 1985). Recasens et al. (1997) define articulatory constraints as various degrees of the involvement of the dorsum. The model predicts that the degree of the vowel-dependent coarticulatory effects is inversely related to that of the dorsum involvement: for Catalan, bilabials > dentoalveolars > alveolars, velars, dark /l/, /s/.

An implication of the fact that the consonantal gestures vary in the degree of coarticulatory resistance is that the speaker attempts to follow the particular production requirements of the given consonant under the influence of the vowel articulation. Because the lingual consonants are produced by the synergetic activities of the two components of the tongue, this naturally causes ‘articulatory antagonism’ (Farnetani, 1989) whenever competing and conflicting demands are made on the tongue. A question is how the antagonistic tongue gesture is resolved in order to articulate the given consonant successfully.

The current study examines the synergetic interdependency between the activities of the tongue tip/blade and the dorsum for the articulations of the coronal stops [t, d, n, r], the coronal and dorsal fricatives [s, ç, ç] and the coronal affricates [ts, tç]. The degree of interdependency is inferred from the correlation value ($r^2$) between the amount of contact in the front region and that in the central region (see section 1.4). In addition to the basic articulatory features of the consonantal articulations, the data of the correlation analysis enables us to gain important insights into the distinctiveness of the consonantal gestures in question. As Recasens’ model implicitly suggests, the ‘suppression’ of one component may be a necessary prerequisite to realizing the distinctive movement of the other component. We will expand this discussion using the examples from the phonological development of Japanese children.

Another question that arises from articulatory antagonism is how competing and conflicting demands are resolved. One possible response of the articulator to those commands is ‘gestural blending’ (Browman & Goldstein, 1991): the output gesture becomes an intermediate articulation between the two original gestures. This possibility can only account for the case where the given two gestures use the same set of articulators. If a certain lingual consonant strongly resists, or suppresses, the effects of the contextual vowels, then the two components of the tongue cannot be blended. Instead, they tend to be sequential: the realization of one gesture after the other. Furthermore, given the consonant-specific constraints on the tongue dorsum, the resolution of the antagonistic gesture (i.e. the choice of

---

3) Fowler and Saltzman (1993; 81) introduce the concept of blending strength.

...The tract-variables component of a given gesture are endowed with their own characteristic degrees of blending strength, according to the context-independent sets of blending parameters assigned to these gestural components. Roughly speaking, blending strength varies inversely with gestural sonority – stops are strongest and vowels are weakest. Blending of gestures with similar strength, e.g., two vowels, results in a simple averaging. The concept is intuitively appealing but Fowler & Saltzman make no attempt to describe a possible hierarchy of strength degree in detail.
blending or sequencing) is undoubtedly consonant-specific. We will examine these questions extensively in the EPG experiment of Japanese palatalisation. A systematic and detailed analysis will be presented for the pattern and timing of interarticulator coordination between the primary gesture and the (secondary) dorsum gesture during the production of the palatal(ised) consonants [n, c, r, k]. The discussion extends to the issue of articulatory complexity and that of the physical realisation of the relatively abstract phonological structure.

In this study we adopt a parametric perspective to characterise various patterns of lingual articulations and attempt to elaborate that perspective. Recent research of coarticulation and tongue dynamics has uncovered the inadequacy of previous articulatory descriptions and proposes various techniques to specify the movements of the articulators. It has been suggested that the consonantal articulations can be distinguished from each other by more properties than the three traditional ones: place, manner and voicing. There is an urgent need for an empirical study to examine the linguistically contrastive gestures at the physical level. The EPG technique produces articulatory information essential for the current study. Furthermore, the strategy we adopt leads us to a review of traditional explanations based on allophonic rules or phonological rules in the phonological component; the viewpoint offers new proposals based on organisation of articulatory gestures or rules in the vocal tract. Based on the results of our EPG experiment we will discuss some phonological processes: /r/ variability in child and adult phonology; CyV variability and de-palatalisation; and vowel devoicing and its related phenomena. All the phenomena have one characteristic in common: the process applies optionally and/or gradually. We will argue that the strategy enables us to approach the psychophysical factors underlying the phenomena and to consider the preconditions for the variability observed.

1.3 Phonetic Individuality in Phonetics and Phonology
One of the aims of the current study is to determine the basic articulatory feature of selected lingual articulations in Japanese. This task necessarily involves the issue of phonetic individuality. One persistent question regarding universal phonetics is concerned with the treatment of language-specific aspects of coarticulation, more commonly language-specific phonetic contents of the distinctive features. Crucial in the following argument is how the language-specific phonetic properties are incorporated in the grammar of a language. We discuss this issue by combining the concept of coarticulation and that of ‘articulatory settings’ (Honikman, 1964).

In the generative phonology framework the grammar of a language consists of several tiers, each of which represents a specific component. The phonological component functions largely as an interpretive component that assigns phonetic contents to the surface
structure by sets of phonological rules. The output of the phonological component is the input to universal phonetics or speech production. The features are binary at the phonological level where the phonological rules delete, add, and interchange the specification in a two-dimensional matrix. In contrast, the features at the phonetic level are interpreted as physical scales ‘describing independently controllable aspects of the speech events’ (Chomsky & Halle, 1968; 297). This model presupposes that coarticulation is largely predictable, and hence the grammar need not explain the phenomenon, and that it is assumed that the phonetic contents of the distinctive features are the same across languages. Halle (1983; 95) explicitly states that ‘the distinctive features correspond to controls in the central nervous system which are connected in specific ways to the human motor and auditory systems’. Thus, ‘phonology’ engages in language-specific rules, while ‘phonetics’ involves the universal automatic implementation of the phonology. There are two phonetics in the Chomsky & Halle’s model (Keating, 1985): one is the rule that supplies phonetic content and increases pronounceability of the phonological representation; and the other is the rule that converts the phonetic representation into a physical reality. The latter ‘phonetics’ is placed outside the grammar of a language: the physiologically-oriented explanations are no longer satisfactory.

Contrary to the assumption of universal phonetics, a number of cross-linguistic studies have shown that languages exhibit distinctive values for some phonetic properties that are considered as determined physically. Lisker & Abramson (1964), a frequently cited example, demonstrate that voice onset times (VOT) for initial voiceless stops vary systematically across languages. Gibbon et al. (1993) show that the degree of articulatory overlap between /k/ and /l/ in six European languages differs considerably in magnitude and timing. Furthermore, substantial phonetic variations can be found for regional accents of a given language. These studies suggest that the phonetic representation of a given phonological representation may be different from language to language; universal phonetics of the grammar needs refinements. A question then is how such language-specific phonetic properties are described in the grammar of a language.

The claim that coarticulation is a mechanical response of the vocal tract is not new. In traditional phonetics there is a distinction between ‘similitude’ and ‘assimilation’ (Jones, 1960; 217ff.): Jones’ example of the former is a partial devoicing of /l/ in words such as please and play; the latter involves both diachronic and synchronic sound changes: picture [piktjur]—►[piktʃə]; tells you [telz ju]—►[tel3 ju]. Wang & Fillmore (1961; 130) introduce another distinction between ‘intrinsic’ and ‘extrinsic’ allophones: the former are the ‘secondary cues which reflect the structure of speech mechanism in general’, whereas the latter are ‘those which reflect the speech habits of a particular community’. Their example of an intrinsic allophone is vowel shortening before voiceless consonants and their examples of an extrinsic allophone are vowel nasalisation before nasal consonants and the lip rounding of labial consonant before rounded vowels. It is true that a careful distinction is made depending on the origin of contextual variability. However, most of us would accept that the terminology is particularly confusing and the distinction tends to be arbitrary.
An obvious response is to devise a new system of phonetic implementation. Keating (1990a) suggests that the phonetic representation consists of the three discrete representations: the categorical representation generates the articulatory and acoustic parametric representations. It is assumed that 'speakers would use the categorical representation to arrive at an articulatory one, while listeners would use an acoustic representation to arrive at categorical one' (Keating, 1990a; 324). Instead of universal phonetics, Keating (1990b) has proposed that the 'window' model of coarticulation (Keating, 1990b) accounts for the qualitative and quantitative variability across contexts and languages. In other words, Keating attempts to promote the phonetic implementation to a part of the phonology and to derive more physical representation from the phonological representation.

A window is a range of variability within the target value of a feature or a combination of the features: the window width (i.e. a permissible variability) will be narrow when features are specified; but when underspecified, the width will be broad and greater variability is allowed. A sequence of windows represents the path for the posited articulatory trajectory varying over time in a given context. The given trajectory interpolates the target windows so that the narrow window provides a constrained path: the phonologically specified feature influences the unspecified features. Thus, the theoretical bases of the window model are: the concept of 'coarticulatory resistance' (Bladon & Al-Bamerni, 1976); and 'underspecification' (Archangeli, 1988; Keating, 1988). This model accounts for the different degrees of V-to-V coarticulatory effects observed for Russian, English and Swedish (Ohman, 1966). Since Russian has a phonemic distinction between palatalised and non-palatalised consonants, it is assumed that the tongue body feature [+high] is constrained in some way during the production of the consonant (Choi & Keating, 1991). When the intervocalic consonant in a given VCV sequence is underspecified for the tongue body feature, the trajectory from V1 to V2 is smoothly interpolated: the interpolation, or the fill-in rule of the underspecified feature, is language-specific in nature. This is the case for English and Swedish. In contrast, the movement path is constrained in Russian when the intervocalic consonant involves the specification of the tongue body feature: V-to-C coarticulation tends to be blocked. The window model suggests that the means of realising the language-specific effects should be dealt with by the grammar of a language.

The window model is a proposal about how the language-specific phonetic contents of the feature are represented, and how the sequencing of the segments is attained in terms of a single articulatory dimension. As described above, the difference in the coarticulatory effects is accounted for by assigning the narrower window to the feature [+high] in Russian, but not in English and Swedish. A similar explanation applies to the VOT values that differ across the languages (Lisker & Abramson, 1964): various window widths may be set up for the feature [±spread glottis] or [±constricted glottis] (Halle & Stevens, 1971). However, many
objections have been raised to Keating’s framework. We shall focus our argument on the
descriptive ability of language-specific characteristics\(^5\).

The studies of lingual coarticulation in particular do not support the idea of the
window model. Recasens’ model of lingual articulation (Recasens et al., 1997), cited in the
last section, shows that the tongue dorsum coarticulation varies across the consonants: for
Catalan the degree of coarticulatory resistance increases in the order bilabials<dentopalatals
<alveolars<velars<dark /l//s/. It is supposed that this hierarchy is, to a certain extent,
applicable to other languages. In contrast, the window model predicts that the movement of
the tongue dorsum is realised language-specifically for the coronal consonants specified by
the feature [+anterior]. Because of the underspecification of the dorsal feature, its value is
derived from that of the surrounding segments by the language-specific implementation of the
target interpolation: a particular width of the window is not provided for the dorsal feature.
Thus, there is no way to differentiate the various degrees of coarticulatory resistance that are
observed for the [+anterior] coronals. The hierarchy that may partly be similar across the
languages is not taken into account either. Therefore, Recasens & Farnetani (1999) suggest
that these differences among the consonants and similarities among the languages result from
the consonant-specific production constraints, rather than from the language-specific
realisation of phonological underspecification.

The core of the problem lies in the fact that the window model makes use of the
specified phonological feature alone. Whereas the window width of a given feature is
modified by the marked specification of the feature in the phonological representation, it is
left unconstrained for the underspecified feature, the phonetic content of which is filled in by
the language-specific phonetic implementation. Thus, ‘the physics’ of the phonological
representation is sought within the limits of the conventional ‘phonetics’ (Fowler, 1990).
Clements (1985, p. 226) states that ‘features are not entities…but categories to which entities are
assigned’. This is not to deny the correspondence between the feature and its posited event in
the vocal tract, but to imply the limitation of the feature representation. It is fictitious to
assume that the features signify all the events in the vocal tract.

The current study will argue that it is possible to characterise phonetic individuality
as the pattern of coordination and timing between the two components of the tongue.
‘Palatalisation’ in Japanese is chosen for the analysis. The idea that language-specific

---

\(^5\) See Browman & Goldstein (1995) for a severe criticism on the view of the grammar. In addition,
Keating (199a) assumes a natural correlation between the articulatory and acoustic parametric
representation, on the grounds that one representation guarantees the recoverability of the other
between speakers/hearers. This concept ignores the issue of motor equivalence: the similar acoustic
effect can be produced by various patterns of the articulatory gesture (e.g. Ladefoged et al., 1972). The
articulation of /r/ in American English illustrates the point. Delattre and Freeman (1968) show that
there are at least six types of the tongue shape that generate the consonant /r/.
characteristics are reflected in the coarticulatory patterning comes from the concept of 'articulatory settings' (Honikman, 1964). This is a well-exercised concept in traditional phonetics for capturing the phonetic individuality of a spoken language. Essential to this is that the systematic articulatory manoeuvres are language-specifically determined. We must distinguish the 'settings' of the articulators on the one hand, from the 'outcome' of the settings on the other. Both aspects are integrated into the specification of voice quality (Laver, 1980, 1994). We shall show how the concept of articulatory settings is defined; how it has been applied to the phonetic analysis; and why it has a link to the language-specific coarticulatory strategy.

It was Beatrice Honikman who first used the term 'articulatory settings' to mean 'the over-all arrangement and manoeuvring of the speech organs'. The parametric view of speech is the basis of the concept as Honikman (1964, 73) describes:

Articulatory setting does not imply simply the particular articulations of the individual speech sounds of a language, but is rather the nexus of these isolated facts and their assemblage, based on their common, rather than their distinguishing, components. The isolated articulations are mutually related parts of the whole utterance; they are clues, as it were, to the articulatory plan of the whole; the conception of articulatory setting seeks to incorporate the clues or to see them as incorporated in the whole. Thus an articulatory setting is the gross oral posture and mechanics, both external and internal, requisite as a framework for the comfortable, economic, and fluent merging and integrating of the isolated sounds into that harmonious, cognisable whole which constitutes the established pronunciation of a language.

This approach characterises speech of a language as coordination and regulation of articulatory movements that are determined language-specifically. Two aspects, 'external and internal', distinguish the lips and jaws from the tongue: the distinction is made relative to the oral cavity; whether the articulator is external or internal to the oral cavity. The parameterisation of the three articulators is specified by 'a gross oral posture and mechanics [i.e. biomechanical constraints]', 'muscular tension' and 'pressure exerted by the articulator'.

For the tongue setting in particular, Honikman (1964, 76) introduces the notion 'anchorage' that is the tethered area or position, as opposed to the 'free or operative part' of the tongue. It is assumed that the most frequently occurring sounds and sound combinations in a language determine the anchorage of a language. Honikman suggests that the tongue setting in English is influenced by the anchorage required for frequently occurring alveolar sounds [t, d, n, r, s, z]. In contrast, the free, or operative, part of the tongue gives support to the anchorage and is coordinated with it.

---

6) We do not discuss the historical development of the concept. See Kelz (1971) and Laver (1978) for a detailed survey.
7) Honikman (1964, 75) considers the internal articulatory setting as 'the over-all positioning of the internal mobile organs of the mouth for natural utterance'.

27
With regard to the free part of the tongue: for the most frequent English consonants [t, n]...the tip is the effective articulator: the tip is somewhat narrowed and tapered by lateral contraction. In [t, d, n, l, l] the tapered tip works energetically up and down as it touches, exerts some pressure on, and comes down away from the rim of the alveolar ridge to or towards the floor of the mouth, thus allowing some other part of the tongue to come comfortably into play for a following vowel or for a following consonant not requiring tip or blade articulation.

The distinction between the anchorage and the free part is compatible generally with the account of the two components of the tongue and their interdependency.

The articulatory settings have been studied for one language in comparison with the other. Also, one particular regional accent of a language is compared with a standard pronunciation of that language. Some instances of the systematic descriptions are helpful to see how the language-specific characteristics and intralanguage variability are captured. For effective comparison we use the descriptive framework proposed by Laver (1980, 1994) in its simplified form and apply it to the observations reported by Honikman (1964) for Southern British English (RP) and French; by Jenner (1987) for RP and Dutch; and Collins & Mees (1995) for General American English (GA) and Danish. For Japanese Edwards (1903) provides a thorough description of the articulatory settings (the articulation base in his terminology).

Table 1-1 shows the settings of RP and French. The information that is not available is left blank. The two languages involve distinctive use of the lips and the anchorage.

**Table 1-1: Articulatory Settings of Southern British English and French (Honikman, 1964)**

<table>
<thead>
<tr>
<th>Laryngeal Settings</th>
<th>Southern British English (RP)</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phonation Types</strong></td>
<td>Neutral; Moderately active</td>
<td>Rounded; Vigorously active in spreading and rounding</td>
</tr>
<tr>
<td>Laryngeal Tension</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Supralaryngeal Settings</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larynx</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lips</td>
<td>Loosely close but not clenched</td>
<td>Slightly open</td>
</tr>
<tr>
<td>Jaw</td>
<td>Tapered</td>
<td>Untapered</td>
</tr>
<tr>
<td>Tongue Tip/Blade</td>
<td>Slightly concave to the roof</td>
<td>Convex to roof</td>
</tr>
<tr>
<td>Tongue Body</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tongue Root</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anchages</td>
<td>Alveolarisation and velarisation</td>
<td>Dentalisation</td>
</tr>
<tr>
<td>Velum</td>
<td></td>
<td>Cheeks contracted</td>
</tr>
<tr>
<td>Overall Tension</td>
<td>Relaxed</td>
<td></td>
</tr>
</tbody>
</table>

8) The notion of 'the anchorage' agrees with the assumption that all skilled behaviour involves 'priming' of the relevant muscle groups (Lashley, 1951).

9) We describe the anchorage as the accommodating part represented by featural terms like dentalisation, alveolarisation and so on (Laver, 1980, 45).
### Table 1-2: Articulatory Settings of Southern British English and Dutch (Jenner, 1987)

<table>
<thead>
<tr>
<th>Laryngeal Settings</th>
<th>Southern British English (RP)</th>
<th>Dutch</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phonation Types</strong></td>
<td>Slightly breathy/relaxed</td>
<td>Hard and tense</td>
</tr>
<tr>
<td><strong>Laryngeal Tension</strong></td>
<td>Low (lax)</td>
<td>High (tense)</td>
</tr>
<tr>
<td><strong>Supralaryngeal Settings</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Larynx</strong></td>
<td>Slightly lowered</td>
<td>Raised</td>
</tr>
<tr>
<td><strong>Lips</strong></td>
<td>Neutral; lax spreading</td>
<td>Firmly held; drawn back at corners; tight inner rounding without protrusion</td>
</tr>
<tr>
<td><strong>Jaw</strong></td>
<td>Close but loose</td>
<td>Slightly open; tense</td>
</tr>
<tr>
<td><strong>Tongue Tip/Blade</strong></td>
<td>Tip (and blade)</td>
<td>Blade (and front); tip lowered</td>
</tr>
<tr>
<td><strong>Tongue Body</strong></td>
<td>†Slightly concave to the roof</td>
<td>Retracted and bunched</td>
</tr>
<tr>
<td><strong>Tongue Root</strong></td>
<td>Neutral or slightly advanced</td>
<td>Uvularisation/pharyngealisation nasalisation</td>
</tr>
<tr>
<td><strong>Anchorage</strong></td>
<td>Alveolarisation and velarisation</td>
<td>Firm closure; strong oral/nasal contrast; some secondary nasalisation</td>
</tr>
<tr>
<td><strong>Velum</strong></td>
<td>Firm closure; strong oral/nasal contrast; some secondary nasalisation</td>
<td>Firm closure; strong oral/nasal contrast; some secondary nasalisation</td>
</tr>
<tr>
<td><strong>Overall Tension</strong></td>
<td>lax</td>
<td>tense</td>
</tr>
</tbody>
</table>

†Honikman (1964).

The articulatory settings suggested for RP and Dutch (Jenner, 1987) are given in Table 1-2 above. Whereas the settings for RP are characterised overall as lax, those for Dutch are characterised as tense (Jenner, 1987, 133). The tongue settings are also opposite to each other: alveolarisation and velarisation for RP but uvularisation or pharyngealisation for Dutch. This diversity of the anchorage can be considered as one source of ‘phonetic transfer’ (Odlin, 1989) in pronunciation learning: it is observed that in English spoken by Dutch native speakers, the front vowels and consonants have a darker quality, while the sounds articulated in the back region have a heavily dark quality (Jenner, ibid.).

Table 1-3 summarises the settings of General American English (GA) and Danish suggested by Collins & Mees (1995). Citing Esling & Wong (1983), who describe American English as palatalised voice, Collins & Mees (1995, 418) claim that ‘we cannot detect in standard General American English anything analogous to the overall palatalisation effect heard in Danish’. The evidence for palatalised voice presented by Esling & Wong is the raising of the front vowels. This characteristic, however, would appear to be more obvious in non-standard American English, such as in the New York City accent and many southern accents, than in GA (Collins & Mees, ibid.).
In his classical, but innovative, work Edwards (1903; 52ff.) describes a number of observations, in which the articulatory settings of Japanese are discussed in terms of the movement of the lips and the tongue.

La position moyenne de la langue est avancée et relevée, sa forme un peu aplatie et ses muscles relâchés.

Les lèvres restent autant que possible dans la position neutre, qu'on pourrait décrire comme le premier mouvement de sourire.

L’arrière-langue est très peu active, donc les consonnes d’arrière sont peu nombreuses, et il n’y a pas de voyelle postérieure formée plus en arrière que celle intermédiaire entre (o) et (ə).

Les exemples comme kakatte (prendre) qui devient k*ok*atte, montrent la lenteur du mouvement de l’arrière-langue, elle se déplace si lentement qu’on a eu l’impression d’un frottement plutôt que de deux explosions. Ce changement de plosive en fricative est encore facilité quand la plosive est soufflée, parce que sa force d’expiration fait souvent du son transitoire un élément indépendant.

Le devant de la langue remue, au contraire, très vite, ce qui facilite les changements de place pour les sons voisins, et cause une certaine instabilité de la prononciation.

Pour fermer le passage de l’air, on emploie les parties plus larges de la langue plutôt que la pointe elle-même.

Les mouvements verticaux de la langue sont distincts et nettement arrêtés, et les mouvements horizontaux de fermeture leur sont subordonnés.

Il en résulte que les voyelles sont peu nombreuses et assez bien définies, tandis que pour les consonnes la langue change de position suivant les voyelles qui les accompagnent.

Sakuma (1929; 189f.) agrees with Edwards in that the tongue blade and pre-dorsum are the anchorages of the tongue settings (i.e. palatalisation); and that the tongue contact with the palate generally involves a larger area (i.e. the flattened shape and the preference for laminal articulation, rather than apical).

<table>
<thead>
<tr>
<th><strong>LARYNGEAL SETTINGS</strong></th>
<th><strong>GENERAL AMERICAN ENGLISH</strong></th>
<th><strong>DANISH</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PHONATION TYPES</strong></td>
<td>Relaxed (not tight)</td>
<td>Tight quality</td>
</tr>
<tr>
<td><strong>LARYNGEAL TENSION</strong></td>
<td>Low (lax)</td>
<td>Tense</td>
</tr>
<tr>
<td><strong>SUPRALARYNGEAL SETTINGS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LARYNX</strong></td>
<td>Relatively lowered</td>
<td>Lowered</td>
</tr>
<tr>
<td><strong>Lips</strong></td>
<td>Weaker rounding for /u, u, u:/; rounding and protruding for /j, 3, tf, ð, r, w/</td>
<td>Active; to distinguish front rounded and unrounded vowels</td>
</tr>
<tr>
<td><strong>Jaw</strong></td>
<td>Relatively relaxed</td>
<td>Tip (and blade); tense</td>
</tr>
<tr>
<td><strong>TONGUE TIP/BLADE</strong></td>
<td>Tip (and blade); tense</td>
<td>Tip (and blade); lax</td>
</tr>
<tr>
<td><strong>TONGUE BODY</strong></td>
<td>Concave to the roof and bunched</td>
<td>Lax</td>
</tr>
<tr>
<td><strong>TONGUE ROOT</strong></td>
<td>Slightly advanced</td>
<td>Tense and retracted towards the lower pharynx</td>
</tr>
<tr>
<td><strong>ANCHORGES</strong></td>
<td>Alveolarisation and uvularisation</td>
<td>Palatalisation and laryngopharyngealisation</td>
</tr>
<tr>
<td><strong>VELUM</strong></td>
<td>Weakly closed; semi-continuous secondary nasalisation</td>
<td>Lax</td>
</tr>
<tr>
<td><strong>OVERALL TENSION</strong></td>
<td>Lax</td>
<td>Lax</td>
</tr>
</tbody>
</table>
To show the articulatory settings studied for some regional accents in comparison with the settings of a standard pronunciation, we shall summarise the observations of English spoken in Norwich (Trudgill, 1974), in Liverpool (Knowles, 1978) and in Cardiff (Collois & Mees, 1990). It is reasonable to assume that the intra-language variability of the settings is planned. Even if the surface realisation of the given phonetic segment varies considerably, the settings behind it share much in common with what we may call the base settings. If we consider the RP settings as the base for British English, the voice characteristics of the regional accents are interpreted as systematic deviations from the base settings. The settings of the three regional accents are given in Tables 1-4 and 1-5, with the RP settings by Jenner (1987) included as a reference.

### Table 1-4: Articulatory Settings of RP (Jenner, 1987) and Norwich (Trudgill, 1974)

<table>
<thead>
<tr>
<th>Laryngeal Settings</th>
<th>Southern British English (RP)</th>
<th>Norwich</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonation Types</td>
<td>Slightly breathy/relaxed</td>
<td>Creaky</td>
</tr>
<tr>
<td>Laryngeal Tension</td>
<td>Low (lax)</td>
<td>High (tense)</td>
</tr>
<tr>
<td><strong>Supralaryngeal Settings</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larynx</td>
<td>Slightly lowered</td>
<td>Raised</td>
</tr>
<tr>
<td>Lips</td>
<td>Neutral; lax spreading</td>
<td></td>
</tr>
<tr>
<td>Jaw</td>
<td>Close but loose</td>
<td>Slightly open; tense</td>
</tr>
<tr>
<td>Tongue Tip/Blade</td>
<td>Tip (and blade)</td>
<td>Blade (fronted)</td>
</tr>
<tr>
<td>Tongue Body</td>
<td>†Slightly concave to the roof</td>
<td>Lowered</td>
</tr>
<tr>
<td>Tongue Root</td>
<td>Neutral or slightly advanced</td>
<td></td>
</tr>
<tr>
<td>Anchorage</td>
<td>Alveolarisation and velarisation</td>
<td></td>
</tr>
<tr>
<td>Velum</td>
<td>Firm closure; strong oral/nasal contrast; some secondary nasalisation</td>
<td>Lowered; nasality; high muscular tension of the pharynx walls</td>
</tr>
<tr>
<td><strong>Overall Tension</strong></td>
<td>Lax</td>
<td>Tense</td>
</tr>
</tbody>
</table>

### Table 1-5: Articulatory Settings of Liverpool (Knowles, 1978) and Cardiff (Collins & Mees, 1990)

<table>
<thead>
<tr>
<th>Laryngeal Settings</th>
<th>Liverpool</th>
<th>Cardiff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonation Types</td>
<td>Adenoidal</td>
<td>Husky, breathy, hoarseness</td>
</tr>
<tr>
<td>Laryngeal Tension</td>
<td>Tense</td>
<td>Tense</td>
</tr>
<tr>
<td><strong>Supralaryngeal Settings</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larynx</td>
<td>Upwards (raised)</td>
<td>†Tense; retracted corner of the mouth; overall lack of lip rounding</td>
</tr>
<tr>
<td>Lips</td>
<td>Spread</td>
<td></td>
</tr>
<tr>
<td>Jaw</td>
<td>Held close to the upper jaw; even for open vowels</td>
<td>Tense front/blade; close to the alveolar ridge</td>
</tr>
<tr>
<td>Tongue Tip/Blade</td>
<td>Backwards</td>
<td>Lax; posterior of the tongue</td>
</tr>
<tr>
<td>Tongue Body</td>
<td>Upwards</td>
<td></td>
</tr>
<tr>
<td>Tongue Root</td>
<td>(Velarise all consonants)</td>
<td>(Alveolarised)</td>
</tr>
<tr>
<td>Anchorage</td>
<td></td>
<td>Lowered semi-continuous nasality; enlarged pharyngeal cavity</td>
</tr>
<tr>
<td>Velum</td>
<td>Tightened pharynx</td>
<td>Tense</td>
</tr>
<tr>
<td><strong>Overall Tension</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† Honikman (1964); † Lip-rounding is determined socially: the pronunciation of middle-class Cardiff English tends to have neutral lip setting (Collins & Mees, 1990; 88).
Of most interest here are the settings of laryngeal tension, larynx position and the anchorage. For the former two settings, the three accents all have a tenser and more raised location of the larynx in RP. For the latter, the tongue settings show changes in the degree of modification. In the Norwich settings the phonetic space of the vowels is effectively fronted and lowered, as the anchorages are alveolarisation and possibly dentalisation. In the Liverpool settings the habitual position of the tongue shifts backwards and upwards, the consequence of which is the velarised realisation of all consonants. In contrast, the tongue positioning tends to be advanced and alveolarised in the Cardiff settings. Thus, these three regional accents reveal systematic deviations from the base settings, i.e. the settings of Southern British accent.

We have observed the articulatory settings analysed for Southern British English, Dutch, Danish, and Japanese; and for English spoken in Norwich, Liverpool and Cardiff. It can be seen that the concept of articulatory settings is a proposal to describe the production of speech as a set of reciprocal physiological actions of the articulatory organs, a composite movement that is organised language-specifically and is constrained biomechanically. Also, it is understood that the settings constitute an important component of spoken language. Although the voice characteristics are typified by various individual parameters of the settings, it is important to emphasise that the articulatory settings are the totality, or coordination, of the movements executed by each parameter, rather than a regular positioning of a given single articulator. If we assume that the articulatory settings described above are prototypes of each speech community, it is reasonable to suppose that a native speaker of a given language (or accent) must have acquired the particular settings as they acquire knowledge of the spoken language. Then, what is the status of those language- and accent-specific characteristics of the vocal tract controls and configurations?

When the above articulatory settings are related to the articulation of the individual phonetic segments, it is expected that some segments will be more influenced by a given setting than other segments. Laver (1980, 20) calls this ‘a principle of segmental susceptibility’: ‘the effect of a setting on a segment will be proportional to the distance between the articulatory locations of the setting and the segment’; and ‘the effect of the setting can be either to shift the articulatory location of the segment closer to that of the setting, or to add a secondary articulation to the production of the otherwise unperturbed segment’. According to this principle, the articulatory settings are defined as ‘the aspect of the performance of a segment that it shares with other susceptible segments’ (Laver, ibid.).

The consequence of the above characterisation is that the articulatory settings function as a filter in the linguistic modelling of speech production. Trudgill (1973, 1974) suggests that it might be possible to explain the regional differences at the phonetic realisation level by incorporating the articulatory settings, rather than by formulating a set of complicated variable rules. Trudgill (1973; 161f.) outlines a possible form of ‘general setting
rules’ for English spoken in Norwich.

\[(5) \quad \text{(a) Vocal Organs } \rightarrow x <\text{Setting 1 } \sim \text{ Setting 2}> \quad x = f (\text{Class})
\]
\[(b) \text{ Setting 2 } \rightarrow \text{ High Muscular Tension}
\]
\[(c) \text{ High Muscular Tension } \rightarrow x \begin{cases} \text{Raised Larynx} \\ \text{Tense Vocal Tract} \end{cases} \quad x = f (\text{Class, Sex, Style})
\]
\[(d) \text{Tongue } \rightarrow \text{ Fronted and Lowered / Raised Larynx}\]

These setting rules effectively produce the more open vowels found in pronunciations such as beer /biːə-/beə-/beə/, seeing /siːə-/seə-/seə/n/ and he have /hiːə-heə-heə/ (Trudgill, 1973; 157). Knowles (1978) also discusses the effects of the settings upon the realisation of the phonetic segment: the Liverpool settings (the Scouse setting in his terminology) prefer a velarised laminal articulation, so that the alveolar fricative [s] tends to be realised as the postalveolar [ʃ]. These two cases represent the ‘susceptibility’ and ‘compatibility’ (Laver, 1980) of the articulatory settings. On the one hand, the lingual gesture is susceptible and compatible in the Norwich case: there are few conflicts between the settings and the production requirements for a given phonetic segment. On the other hand, the Scouse setting is not compatible with the articulatory movements required for the [s] production: while the anchorage of velarisation and the preference of laminal articulation effectively raise the tongue dorsum, the tip tends to be directed downwards; this gesture, when applied to [s], generates a tongue shape similar to [ʃ], with the appropriate groove width maintained\(^\text{100}\). Knowles (1978; 90) suggests that ‘[i]n a ‘thick’ accent the setting might be given priority’.

These observations reveal that the realisation of the phonetic segments is conditioned by the settings. However, it seems that current phonological theories might have abandoned the commitment to incorporate the above idea and the commitment to enrich the linguistic modelling of speech production.

The current study will attempt to describe some aspects of the articulatory settings of Japanese within the framework of Articulatory Phonology (Browman & Goldstein, 1986, 1989, 1990a,b, 1992). Because the articulatory settings are defined as ‘a gross posture’ (Honikman, 1964) of a given articulator, gestural coordination and timing are less emphasised in the description and explanation of the settings. It is the fact that, although the articulation of a given segment is transcribed by the same phonetic symbol across languages, the

\(^{100}\) Our explanation is basically a restatement of Knowles: [the Scouse tongue setting] reduces the efficiency of the grooving, and allows the air to escape more diffusely, thus making [s] more like [ʃ]. The word ‘efficiency’ is somewhat ambiguous in this context. Therefore, we attempted to improve the explanation by distinguishing the tongue shape generated by the Scouse settings from the groove width similar between the two fricatives. We will discuss the relationship between the constriction place and the groove width for the sibilants in chapter 5.

Knowles (1978; 90) also mentions the lack of overlap between the segments: shrink [ʃrɪŋk] \(\rightarrow\) [ʃrɪŋk]. This necessarily involves temporal aspects of gestural coordination in connection with the language-specific settings.
manifestation of the language-specific characteristics differs spatially and temporally. Furthermore, it is supposed that gestural coordination and timing between the two components of the tongue reflect the characteristics of the settings.

In contrast to the generative phonology framework and Keating's window model, it is possible to suppose that the cross-language variability arises from the language-specific coarticulatory patterning, rather than the different manifestations of the feature. Browman & Goldstein (1986, 1989, 1990a,b, 1992) have proposed that articulatory gestures are the unit of phonological representation. The fundamental hypothesis is that a small number of dynamically defined articulatory gestures constitute basic phonological units, which provide a principled link between phonological structure and articulatory movements: the model assumes no intermediate levels. Gestures are 'autonomous structures that can generate articulatory trajectories in space and time' (Browman & Goldstein, 1986*, 223). The duration is specified inherently for the unit; no external mechanisms of timing are assumed. A gesture is specified by a set of articulators and their dimensions: constriction location (CL); constriction degree (CD); and stiffness (duration of a gesture)\(^1\). There is a broad distinction for the oral constriction gestures, vocalic and consonantal: the consonantal gesture has a greater degree of constriction and a shorter time constant, than the vocalic gesture. Browman & Goldstein propose that the phonological phenomena are naturally accounted for by assuming a single mechanism: the changes in the magnitude and timing of gestural overlap. The gestural analysis is particularly successful in modelling assimilation, insertion and deletion in connected speech and in historical sound change. In contrast, the model does not explicitly raise the issue of the treatment of language-specific characteristics.

The EPG experiments in the current study will give us information necessary for the precise description of the spatiotemporal aspects of various phonetic segments. The variables in the vocal tract, as Browman & Goldstein acknowledge, are incomplete at the present stage. It is important to examine the validity of the variables currently assumed and to seek better descriptions in the context of the particular language, Japanese. At the same time, this task involves the specification of the habitual positioning of the tongue. Given that the tongue settings of Japanese generally require dentalised and palatalised laminal articulation (Edwards, 1903), the principle of susceptibility and compatibility predicts that the settings have more

\(^{1}\) The term 'stiffness' needs clarification. The gestures are implemented in terms of task dynamics (Saltzman & Kelso, 1987) in which a damped mass spring model with constant mass is assumed. Thus, the parameter stiffness refers to the state of the spring in the model. The stiffer the spring becomes, the faster it returns to its resting position. Phonetically, changes in the stiffness manifest different durations. Hawkins (1992) discusses the problem of the abstract status of the stiffness in the model. Barry (1992) discusses the inadequacy of the hypothesis that the mass is always constant for all the articulators. In his EPG experiment of Russian palatalisation Barry shows that, while coronal gestures undergo weakening, dorsal gestures do not. This suggests that there are differences between the masses of the two gestures.
influence upon the consonants articulated at the posterior region: the more posterior the (neutral) constriction place becomes, the more susceptible to the anchorages. This prediction will be examined in the analyses of the articulation of the individual consonants: coronal stops, fricatives and affricates. Furthermore, the EPG analysis of palatalisation will highlight important implications for the spatial and temporal effects of the settings. Because the movement of the tongue tip/blade is, by hypothesis, ‘anchored’ at the front region of the palate, it is reasonably supposed that this effectively constrains the magnitude and timing of the dorsum movement, as well as the fronted (or advanced) realisation of the consonantal articulations, alveolopalatals in particular. These analyses of the standard pronunciation allow us to consider the different phonetic process in standard Japanese and regional accents. CyV variability and de-palatalisation are chosen for the analysis. The former process in standard Japanese produces the change in vowels (e.g. [ç'^u,ktudai]—>[çi,ktudai] ‘homework’), while the latter in regional accents produces the change in consonants (e.g. ['ç'^u:ziN]—>['siiiziÜN]). We will argue that the diversity comes from intra-language differences in the tongue settings, as well as from the production requirements for the sibilants.

In summary, this study adopts a parametric approach to investigate the articulatory and coarticulatory characteristics of Japanese. A parametric approach differs from a postural approach in that the production of speech is seen as continuous, consisting of composite movements in the vocal tract, rather than a sequence of discrete postures of independent articulation. The empirical goal of this study is to undertake a series of EPG experiments to determine the basic articulatory features of the consonantal and vocalic segments; to describe them with relevant parameters; and to examine the effects of linguistic variables upon the articulatory timing of the segmental articulations and palatalisation. The crucial idea is that, while the vocal tract is constrained biomechanically, it is organised phonologically. Based on the results of our EPG experiments, the theoretical goal of this study is to apply the parametric, or gestural, analysis to various phonetic processes in standard accent and regional accents of Japanese. We share the fundamental aim of current research in speech production; to describe dynamic patterns of articulatory movements and to uncover principles of coordination during the production of speech.

1.4 Towards Electropalatographic Specification of Lingual Articulation

1.4.1 EPG and Its Limitations

The EPG system is a safe and well-established technique to detect the articulatory pattern and timing of the tongue contact with the hard palate. The subject wears a thin acrylic palate designed to fit his or her hard palate. It is modelled from a dental impression. The artificial palate is embedded with 62 electrodes and is connected to monitoring electronics. A low voltage signal derived from an oscillator passes through the subject’s body and the electronic
processing unit detects the on-off signal of contact between the tongue and the electrodes on
the surface of the palate (Hardcastle, 1972, 1984; Hardcastle et al. 1991a,b; Hardcastle &
Gibbon, 1997). The EPG system used in this study is the Reading system (EPG2
Electropalatograph, described in Hardcastle et al., 1989), which tracks how the tongue makes
and breaks contact with the hard palate during speech at the rate of 261-302 frames per
second (1 frame = 3.31-3.83msec.).

There are two potential limitations to the EPG technique. First, the technique
typically involves an artificial palate during articulation. There is a concern that the presence
of an artificial palate might seriously interfere with normal speech production (e.g.
Abercrombie, 1957). Opinions differ on this point. To examine the influence of the thickness
of the artificial palate, Matsuno (1989) makes a comparison between articulatorily controlled
and auditorily controlled versions of Japanese /ni/ and /ki/, using the Reading EPG system
(EPG2). For the former version, the speaker focuses on the muscular control, and for the latter
the target consonants are monitored by the ear. It is reported that, while the thickness of the
palate is generally assumed to be negligible for practical purposes, the influence is not always
ignored for two selected consonants in the palatalising environment. However thin an
artificial palate is, the subject has to get used to the feeling of the strange object in their mouth.
Nevertheless, the subject can make slight adjustments of articulatory behaviour in order to
speak naturally with the artificial palate. Generally a short period of practice is enough for the
accommodation (Hardcastle et al., 1989). All the same, it is important to be aware that there
might be a (bare) possibility of affecting the motor organisation of articulation and its acoustic
results. Secondly, there may be difficulty in interpreting articulatory gestures from the contact
data. Because the EPG technique records the facts about lingo-palatal contact, the actual
tongue shape can only be inferred from the location and timing of the contact, the anatomy of
the tongue; and other information (Hardcastle et al., 1989). In this study data taken by other
experimental techniques, both in Japanese and in other languages, was frequently consulted to
interpret the EPG contact pattern and to infer the overall shape of the tongue (e.g. Bollá, 1981,

1.4.2 General Experimental Procedure

The EPG experiments described in the following chapters examined the spatiotemporal
characteristics of lingual activities for the realisation of ten consonantal segments [t, d, n, ŋ, r,
s, ŝ, ç, ts, ŏ]; two vocoids [i, j]; and palatalisation; and vowel devoicing. In this section the
basic analytical procedures are explained in three ways: (i) zoning of the EPG artificial palate;
(ii) data collection and reduction; and (iii) statistical inference. Further details of the
measurement criterion specific to the target articulation will be presented in each chapter.
1.4.2.1 Zoning of the EPG Artificial Palate

In the Reading EPG system 62 electrodes embedded in the surface of the custom-made artificial palate detect the tongue-palate contacts. The electrodes are arranged in eight horizontal rows, where the frontmost row (row 1) has six electrodes and the other seven rows have eight electrodes. While electrode positions in each row are equidistant, the spacing between the front four rows (rows 1-4) is smaller than that between the back four rows (rows 5-8). The electrodes of the frontmost row (row 1) are positioned above the central maxillary incisor teeth and those of the backmost row (row 8) do not extend over the soft palate. The shape of the artificial palate used in this study is shown in Figure 1-1 for speaker MN and Figure 1-2 for speaker TM (see next pages).

It is particularly important for EPG research to divide the surface of the artificial palate into several zones relevant for the analysis of linguopalatal contact. Also, a clear understanding of the subdivisions of the tongue is necessary: it is generally taken for granted that the part of the tongue making a contact against the palate is that which lies directly under that location. Researchers, however, differ in the way they divide the surface of the artificial palate (e.g. Barry, 1992; Farnetani et al., 1989; Hardcastle et al., 1991; Recasens, 1984, 1990). This is mainly related to the fact that an artificial palate for the Reading EPG system is custom-made. Different EPG systems employ technically different arrangements of electrodes on the surface of an artificial palate. The potential problem is that the results from different EPG analyses are not directly compatible.

In the current study the EPG palate was divided into the zones shown in (6). The zoning was identified during the course of the present experiments and was also based on several written sources (e.g. Catford, 1977; Firth, 1948; Keating, 1991; Nguyen et al., 1996; Recasens, 1990).

![Table of tongue and palate regions]

The palatal morphology was studied from the impression of the hard palate and the electrode locations on the EPG palate were tape-measured. It was found that the corner of the alveolar ridge (i.e. the intersection where convexity gives place to concavity (Catford, 1977, 141)) lies between row 4 and 5 on the EPG palates of the two informants. Thus, the above zoning was applied to the analysis of the utterances produced by both speakers. For the description of the tongue, the terminology suggested by Catford (1977) was adopted.
FIGURE 1-1: EPG ARTIFICIAL PALATE OF SPEAKER MN
Figure 1-2: EPG Artificial Palate of Speaker TM
1.4.2.2 Data Collection and Reduction

Two native speakers of standard Japanese, one male (MN) and one female (TM) took part in the experiment. They are in their mid-thirties and mid-forties respectively. General articulatory and coarticulatory characteristics were studied for the V1CV2 sequences where the vowels /i, a, u/ are in all possible combinations. The speech items were mostly nonsense disyllabic (two-mora) words. All the target words were embedded in a frame sentence /moo (word) bakarida/ ['moo (word) 'bakar^ida'] (‘There is/are only (word) now’). They were written on a series of cards where one sentence containing one target word was written in the Japanese kana characters. After more than thirty minutes practice the cards were randomised. The subject then repeated each sentence six times at normal speed, with the default accentual pattern on the target word (i.e. a low-high pattern). Five repetitions were used for the analysis. The EPG and acoustic recordings were done in the phonetics laboratory of the School of Oriental and African Studies, University of London.

Various methods of EPG data reduction have been proposed (Engstrand, 1989; Farnetani, 1990; Hardcastle, 1984; Hardcastle et al., 1991; Fontdevila et al. 1994). There are two crucial analytical points that must be involved in the data reduction process. One is the linguopalatal contact pattern at some specified stage during the production of the target segment. The other is the developing pattern of articulatory movements. In this study the raw EPG data signal was converted into two modes of representation. First, the spatial properties were measured at the point of maximum lingual constriction or narrowing (MAX). Contact configurations averaged over five repetitions were represented by an EPG prototypical palatogram. Second, the articulatory trajectories were computed from the raw EPG data: time was plotted against varying degrees of region contact (front, central, and back) (see the zoning described in (6) above). The general articulatory and coarticulatory characteristics were quantified. Speakers’ control strategies were inferred from those spatiotemporal EPG correlates of the given articulation of the phonetic segment.

1.4.3 Statistical Inference

Statistical analysis allows us to summarise complex data and to draw inferences on tendencies which are obvious across the speakers and the target articulations. We used three statistical analyses for the measured values of various spatial and temporal parameters: (i) an analysis of variance (ANOVA); (ii) Student-t (t-test); and (iii) regression analysis. The data for these analyses was pooled across speakers, target consonants or vowel contexts. Separate analyses were also carried out for each speaker in order to identify significant individual differences.

An ANOVA was used to compare the differences in the means between more than three groups of the EPG contact data. A series of two-way ANOVA were first applied to the data set made up of two variables such as the consonant type and vowel contexts. Then, a
one-way ANOVA was carried out separately for those two variables. In each case F represents the ratio of the variance of each sample to the variance between the groups of the given data: large values of F throw doubt on the validity of the null hypothesis (Woods et al. 1986). A t-test was used to compare the differences between the means of the two particular pairs of data: for instance, the differences between the effects of the particular vowel pairs. The statistic t indicates the ratio of observed differences of sample mean to common standard deviation between the given two groups (Woods et al. 1986). To examine the interdependence (or correlation) between the given two parameters, the Pearson product-moment coefficient of linear correlation ($r^2$) was obtained: for instance, the correlation between the duration of articulatory closure and the degree of contact in the front region. The value of $r$ represents the ratio of the covariance of the two given variables to their multiplied standard deviation; a high value of $r$ implies a significant correlation between the two; $r^2$ is interpreted as the proportion of the variability in either variable which was explained by the other (Woods et al. 1986). The significant level of all the statistical tests in this study was set at $p<0.01$.

1.5 Organisation of the Present Study

This chapter set out the five goals of this dissertation. We discussed the two essential notions: the parametric view of speech; and the concept of phonetic individuality. We described the general methodology for conducting the experiments using EPG and the basic procedure of the data analysis.

In the following chapters we will further investigate the sound system of Japanese. We will present the results of a series of EPG experiments and will explore the articulatory and coarticulatory characteristics of the lingual consonants and some phonetic processes.

Chapter 2 sets out the introductory groundwork for the subsequent chapters by reviewing the phonological and articulatory characteristics of the Japanese sound system. This chapter attempts to describe Japanese phonology with a special emphasis on the systematic articulatory controls underlying it.

Chapters 3 and 4 are combined to examine the articulation of the coronal stops [t, d, n, ñ, r]. We present a brief critical review of previous studies in chapter 3 and the hypotheses that will be examined in the EPG experiment are formulated. In chapter 4 we describe the experiment examining the patterns of linguopalatal contact during the production of the five coronal stops above. Of specific interest is the variability of the degree of interdependency between the two components of the tongue: this is examined for other target segments. We examine the three correlations: (i) closure duration and gestural duration; (ii) closure duration and the degree of contact in the front region; and (iii) the degree of contact in the front region and in the back region. These correlation analyses will clarify the posited strategies used by a
Chapter 5 discusses the articulation of the fricatives [s, ç, ç] and the affricates [ts, tç]. In addition to the general articulatory and coarticulatory characteristics, we examine the variability of the groove width of the fricatives and the fricative element of the affricates. Furthermore, the correlation between the groove width and the constriction place is tested. This serves as the data for the discussion about the distinctive properties of the fricative articulation.

Chapter 6 examines the articulation of palatalisation: [n, ç, r³, k³]. We will develop a working hypothesis of palatalisation as ‘a specialised use of a raising gesture of the tongue dorsum’. We will present a detailed analysis of gestural coordination and timing between the primary gesture and the (secondary) dorsum gesture during the production of the four palatal(ised) consonants. Our discussion will extend to the issue of articulatory complexity and that of gestural antagonism.

Chapter 7 investigates the selected phonetic and phonological processes based on the results of our EPG experiments reported in the previous chapters. The three phenomena chosen for the analysis are: (i) /r/ variability in child and adult phonology; (ii) CyV variability and de-palatalisation; and (iii) vowel devoicing. The articulatory nature of these processes is discussed in the light of theoretical assumptions made by Articulatory Phonology (Browman & Goldstein, 1986, 1989, 1990a,b, 1992). We present preliminary observations for the articulatory analysis of vowel devoicing.

Chapter 8 concludes the study with some implications for further research.
Chapter 2

Phonological and Articulatory Characteristics of
The Japanese Sound System

2.1 Introduction
This chapter discusses the sound system of present-day Japanese, the phonetic basis of which is the Tokyo accent. The purpose of this chapter is two-fold. One is to describe segmental phonetic and phonological facts that serve as the background for further discussions; suprasegmental aspects are limited to the general system of pitch accent. The other is to approach the problem of linguistic structure and articulatory organisation by reviewing previous descriptive and experimental studies.

We first begin our discussion in terms of traditional articulatory phonetics. The phonemicisation and systematic phonetic transcriptions used in this study are set up based on previous referential works such as Akamatsu (1997), Kawakami (1977), Miller (1967), Sakuma (1929), Shibatani (1990) and Vance (1987), and my informal observations. Shifting the focus from characterising speech as a series of static postures to characterising it as a dynamic articulatory process, we attempt to incorporate physiological, acoustic, and perceptual findings of previous experimental research into the characterisation of segmental aspects of spoken Japanese. We shall argue how phonetic and phonological structures can be explained parametrically in terms of the systematic articulatory control underlying them.

The organisation of this chapter is as follows. After considering the phonological inventory, syllable structure, and basic accentual patterns in section 2.2.1, the distribution of allophones is illustrated in section 2.2.2. Variable realisations before high vowels are summarised in section 2.3, in which a particular focus is placed on palatalisation. We argue that the 'correlation of palatalisation' (Trubetzkoy, 1958 /1962) is one of the crucial factors in Japanese phonology, giving an historical sketch of its previous linguistic modelling. The CyV variability, the phenomenon related to palatalisation, is also explained. This section serves as a preliminary discussion to Chapter 6, where we will attempt to uncover the articulatory mechanisms of palatalisation. In section 2.4 phonetic segments in loanword pronunciation are discussed. In section 2.5, vowel articulation is described in detail and an articulatory perspective for diphthong production is developed. This is exemplified by comparing 'smoothing' (Wells, 1982) in British English and glide formation in Japanese loanwords. In section 2.6 the production of moraic nasals and obstruents are considered. Lastly, section 2.7 discusses what evidence exists for describing the rhythm of Japanese in terms of moras.
2.2 The Architecture of Japanese Phonology

2.2.1 Phonemes, Syllable Structure, and Accent

The Japanese inventory consists of twenty segmental phonemes: five vowel phonemes and fifteen consonantal phonemes. The phonological inventory in Table 2-1 above may be the most orthodox one that is derived from the fundamental idea that phonemes are contrastive.

The basic syllable structures are CV and CyV (C/j/V). Traditional prosodic phonology has proposed a constituent structure of segmental string, mora, and syllable. The representation of possible Japanese syllable structure is shown in (1) and the examples are given in (2).

\[
\begin{align*}
\sigma &= \text{syllable} \\
\mu_1, \mu_2, \mu_3 &= \text{mora} \\
(C_1)(j/j) V_1 (V_2) \left\{ (N) \right\} (C_2)
\end{align*}
\]

1) Three other phonemes, /Q/, /J/ and /R/, have been proposed by Japanese phonologists and sociolinguists such as in Joo (1977), Kubozono (1992), and Shibata (1958). The phoneme /Q/ is used as the representative of such voiceless obstruents as constitute the first element of geminate consonants: for instance, [kat,ta] ‘buy (past)’ is phonemised as /kaQta/; and [kaçça] ‘pulley’ as /kaQsja/. The segment /J/ represents /i/ as the second element of a two-vowel sequence and /R/ represents the second element of a long vowel: for instance, /kaJ/ ['kai] ‘shellfish’ and /toRkjoR/ /to,ok^oo] ‘Tokyo’ respectively. These three phonemes, as well as /n/, are so-called moraic phonemes. The advantage of this proposal lies mainly in the description of accent assignment: the accent nucleus tends not to fall on moraic phonemes phonologically. However, there are several reasons that we do not accept such a convention. The segments /Q/ and /R/ have no phonetic substance. And the high front vowel is phonemised either by /i/ or by /j/ depending on the position within a syllable: only /ei, ai, oi, ui/ are considered as diphthongs. Because the moraic status of those segments is predictable from their position within the syllable, there is no reason to squeeze the prosodic property into a segment. Therefore, we shall represent the examples above as /katta/, /kai/ and /tookjoo/.

2) The syllable structure in so-called moraic theory (e. g. Hayes, 1989) is assumed: syllables are decomposed into onset consonant(s) and weight units or mora: the onset is supposed to be weightless. The advantage of this model is that it minimises the differences between the languages. If the hierarchical onset-rhyme model is applied, the syllable structure of Japanese is a mirror image of English.
A syllable, as can be seen in (2a-(i)), is minimally composed of a single vowel. The examples in (2a) constitute light syllables and those in (2b-d) are heavy syllables. Super-heavy syllables in (2e-f) that comprise three mora-units are uncommon. There are only two types of closed syllables: the final consonant must be either a nasal /n/ or a voiceless obstruent C₂. Because C₂ is regarded as the initial part of a geminate consonant, the second part is always present as the onset consonant in the following syllable. Three segments, namely the second element of a two-vowel sequence (V₂ in both diphthongs and long vowels), and both syllable-final consonants /n/ and C₂, have an independent status at the mora level: they are functionally the same as a sequence of (C)V₁. Given the word /oto/ ([o, to] 'sound'), we analyse it as either a two-syllable word or a two-mora word. On the other hand, the word /on/ ([o, n] 'sound') is analysed as a word of one syllable but two moras. The representation in (1) reflects those characteristics by aligning a segment under the mora unit. We will assume that the Japanese 'mora' is a weight and positional unit of the syllable until we discuss the concept of mora later in this chapter.

The syllable structure in (1) is also used to derive the phonological form of western loanwords. Given that the underlying form of the loanword pronunciation is developed from the original phonetic form, through default operations such as the vowel quality adjustment and the consonant replacement, then we could illustrate the derivational process of the English word 'strong' as in (3a,b) below.

The insertion of a vowel breaks up a consonant cluster. The velar nasal /ŋ/ is decomposed into two elements, a uvular nasal and a velar stop, with a default vowel /u/ attached. The phonological form of loanwords often violates phonotactic constraints and creates new segmental combinations (section 2.4). Yet, the syllable structure in (1) is persistently maintained.

Standard Japanese is a pitch accent language, in which every word has a fixed pitch.
pattern. There are two kinds of pitch movements, step-up (low-high, LH) and step-down (high-low, HL). The crucial factor controlling where a lexical accent is assigned is the location of the pitch step-down, or HL tone, which is called the accent nucleus. It has been a common practice to identify words with an accent nucleus as 'accented' words; those only with a pitch step-up are 'unaccented' words. The default second mora (i.e. (C)V syllable) from the left is specified as step-up, unless the first mora has nucleus step-down. Only one accent nucleus is permitted within a word: once the pitch is dropped by the nucleus step-down, the step-up never occurs until another word or accentual phrase is set out. For marking the accent patterns the traditional symbols '↑' and '↓' are not used in this study. Following Wells (2000), a default step-up is indicated by '↑' and a nucleus step-down is indicated by '↓'. These two diacritics are placed before the relevant mora. Table 2-2 summarises basic accent patterns, the variation of which is the number of moras plus one. Examples for each pattern are given in Table 2-3.

### Table 2-2: Basic Accent Patterns of Japanese

<table>
<thead>
<tr>
<th>1-mora word</th>
<th>2-mora word</th>
<th>3-mora word</th>
<th>4-mora word</th>
<th>• • •</th>
</tr>
</thead>
<tbody>
<tr>
<td>unaccented</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0↑</td>
<td>0↑</td>
<td>0↑</td>
<td>0↑</td>
<td></td>
</tr>
<tr>
<td>0↑ 0↑</td>
<td>0↑ 0↑</td>
<td>0↑ 0↑</td>
<td>0↑ 0↑</td>
<td>•</td>
</tr>
<tr>
<td>0↑ 0↑↑</td>
<td>0↑ 0↑</td>
<td>0↑ 0↑</td>
<td>0↑ 0↑</td>
<td>•</td>
</tr>
<tr>
<td>acented</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0↑</td>
<td>0↑</td>
<td>0↑</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0↑ 0↑</td>
<td>0↑ 0↑</td>
<td>0↑ 0↑</td>
<td>•</td>
</tr>
<tr>
<td></td>
<td>0↑ 0↑↑</td>
<td>0↑ 0↑</td>
<td>0↑ 0↑</td>
<td>•</td>
</tr>
</tbody>
</table>

N.B. '0' indicates mora. The traditional diacritics, '↑' and '↓', are used to facilitate the comparison with the practice in this study.

### Table 2-3: Examples of Basic Accent Patterns

<table>
<thead>
<tr>
<th>1-mora word</th>
<th>2-mora word</th>
<th>3-mora word</th>
<th>4-mora word</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ha.ga/</td>
<td>/ka.ce.ga/</td>
<td>/sa.kan.a.ga/</td>
<td>/ni.wa.to.ri.ga/</td>
</tr>
<tr>
<td>LH</td>
<td>LHH</td>
<td>LHHH</td>
<td>LHHHH</td>
</tr>
<tr>
<td>[ha,ga]</td>
<td>[ka,cega]</td>
<td>[sa,kanaqa]</td>
<td>[pi,quato,ri,na]</td>
</tr>
<tr>
<td>'leaf'</td>
<td>'wind'</td>
<td>'fish'</td>
<td>'hen'</td>
</tr>
<tr>
<td>/ra.ga/</td>
<td>/so.ra.ga/</td>
<td>/ka.ra.su.ga/</td>
<td>/ka.ma.ki.ri.ga/</td>
</tr>
<tr>
<td>HL</td>
<td>HLL</td>
<td>HLLL</td>
<td>HLLL</td>
</tr>
<tr>
<td>['ha,ga]</td>
<td>['sor,ga]</td>
<td>['kar,asu,na]</td>
<td>['kama,ki,ri,na]</td>
</tr>
<tr>
<td>'teeth'</td>
<td>'sky'</td>
<td>'crow'</td>
<td>'mantis'</td>
</tr>
<tr>
<td>/ja.ma.ga/</td>
<td>/so,ba,ja.ga/</td>
<td>/ku.da.mo.no.ga/</td>
<td></td>
</tr>
<tr>
<td>LHL</td>
<td>LHH</td>
<td>LHHH</td>
<td>LHHH</td>
</tr>
<tr>
<td>[ja,ma,ga]</td>
<td>[so,ba,ja]</td>
<td>[ku,da,mono,ga]</td>
<td></td>
</tr>
<tr>
<td>'mountain'</td>
<td>'noodle shop'</td>
<td>'fruits'</td>
<td></td>
</tr>
<tr>
<td>/o.to.ko.ga/</td>
<td>/o.to.o.to.ga/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LHHL</td>
<td>LHHH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>['o.to,ko,na]</td>
<td>['o.to.o.to,na]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>'man'</td>
<td>'brother'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/o.to.o.to.ga/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[o.to,o.to,na]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N.B. A subject marker /ga/ follows each word and a mora boundary is indicated by a period. Features of phonetic realisation will be accounted for later in this section.
2.2.2 Phonological Analysis of the Syllabary

We develop and exemplify the distribution of allophones by making use of the written syllabary. The advantage of this approach is that it concisely catalogues all the significant sounds and their distributions in Japanese. There are two points to note. The syllabary is a chart of letters symbolising moras, rather than syllables. It comprises all the kana letters used in Japanese, arranged systematically on the basis of their pronunciation. To use it as a sound chart, it needs to be rectified in some respects. Another reason for using the syllabary as the primary data is because of our view of Japanese phonological organisation as a whole. We hypothesise that the phonological system underlying the syllabary forms the core part of the Japanese phonology. This view presupposes that there are phonological operations - however they might be conceptualised - relating to the different strata of the lexicon.

Based on the phonological inventory in Table 2-1, the phonological analysis of the Japanese syllabary is given in Table 2-4 below.

<table>
<thead>
<tr>
<th>/a/</th>
<th>/i/</th>
<th>/u/</th>
<th>/e/</th>
<th>/o/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/k/</td>
<td>a</td>
<td>あ</td>
<td>i</td>
<td>い</td>
</tr>
<tr>
<td>/s/</td>
<td>ka</td>
<td>か</td>
<td>k'i</td>
<td>き</td>
</tr>
<tr>
<td>/t/</td>
<td>sa</td>
<td>さ</td>
<td>ci</td>
<td>し</td>
</tr>
<tr>
<td>/n/</td>
<td>na</td>
<td>な</td>
<td>ni</td>
<td>に</td>
</tr>
<tr>
<td>/h/</td>
<td>ha</td>
<td>は</td>
<td>ci</td>
<td>ひ</td>
</tr>
<tr>
<td>/m/</td>
<td>ma</td>
<td>ま</td>
<td>mi'</td>
<td>み</td>
</tr>
<tr>
<td>/j/</td>
<td>ja</td>
<td>い</td>
<td>ju</td>
<td>ゆ</td>
</tr>
<tr>
<td>/w/</td>
<td>wa</td>
<td>わ</td>
<td>wi</td>
<td>わ</td>
</tr>
<tr>
<td>/g/</td>
<td>ga</td>
<td>が</td>
<td>g'i</td>
<td>ぎ</td>
</tr>
<tr>
<td>/z/</td>
<td>dzai</td>
<td>だ</td>
<td>gi</td>
<td>ぎ</td>
</tr>
<tr>
<td>/d/</td>
<td>da</td>
<td>だ</td>
<td>z'i</td>
<td>ざ</td>
</tr>
<tr>
<td>/b/</td>
<td>ba</td>
<td>ば</td>
<td>bi'</td>
<td>び</td>
</tr>
<tr>
<td>/p/</td>
<td>pa</td>
<td>ぱ</td>
<td>pi'</td>
<td>ぴ</td>
</tr>
<tr>
<td>/k'w/</td>
<td>k'ae</td>
<td>きゃ</td>
<td>k'u</td>
<td>きゅ</td>
</tr>
<tr>
<td>/s'j/</td>
<td>s'ae</td>
<td>しゃ</td>
<td>s'u</td>
<td>しゅ</td>
</tr>
<tr>
<td>/t'j/</td>
<td>t'ae</td>
<td>ちゃ</td>
<td>t'u</td>
<td>ちゅ</td>
</tr>
<tr>
<td>/n'j/</td>
<td>n'ae</td>
<td>にゃ</td>
<td>n'u</td>
<td>にゅ</td>
</tr>
<tr>
<td>/h'j/</td>
<td>h'ae</td>
<td>ひゃ</td>
<td>h'u</td>
<td>ひゅ</td>
</tr>
<tr>
<td>/m'j/</td>
<td>m'ae</td>
<td>みゃ</td>
<td>m'u</td>
<td>みゅ</td>
</tr>
<tr>
<td>/r'j/</td>
<td>r'ae</td>
<td>りゃ</td>
<td>r'u</td>
<td>りゅ</td>
</tr>
<tr>
<td>/g'j/</td>
<td>g'ae</td>
<td>ぎゃ</td>
<td>g'u</td>
<td>ぎゅ</td>
</tr>
<tr>
<td>/z'j/</td>
<td>dz'ae</td>
<td>じゃ</td>
<td>dz'u</td>
<td>じゅ</td>
</tr>
<tr>
<td>/b'j/</td>
<td>b'ae</td>
<td>びゃ</td>
<td>b'u</td>
<td>びゅ</td>
</tr>
<tr>
<td>/p'j/</td>
<td>p'ae</td>
<td>ぴゃ</td>
<td>p'u</td>
<td>ぴゅ</td>
</tr>
</tbody>
</table>

Table 2-4: Phonetic and Phonological Interpretation of the Japanese Syllabary
Although the proper written arrangement of the syllabary is vertical, it is shown here horizontally, retaining the usual order of the consonants. The two axes show phonemic representations: the vowel phonemes are on the horizontal axis and the consonant phonemes are on the vertical axis. There are two kinds of entries: one is the systematic phonetic transcription based on the IPA principles, and the other is the hiragana letter shown on the right of a phonetic symbol. The allophones in brackets tend to occur in intervocalic positions.

The phonological syllabary above is divided into three major groups in terms of their entries: CV syllables, CyV (CjV) syllables, and two syllable-final consonants /n/ and C. Three points seem to be helpful in attempting to sketch out cases of skewed distribution of allophones: (i) /j/ and /w/, (ii) CyV syllables, and (iii) /d/.

First, distribution of /j/ and /w/ is limited: there are no phonemic combinations such as */ji, je/ and */wi, we, wo/. These are the result of diachronic sound change. It is documented that historically both /j/ and /w/ participated in five phonetically realised syllables [ja, i, ju, je, jo] and [wa, wi, u, we, wo] respectively. The former incorporates [i] and the latter incorporates [u] in the five vowels (Komatsu, 1981). This suggests that homorganic combinations are avoided. Such distributions are influenced and altered by the phonological development of /h/. It has been widely accepted that the consonant /h/ in present-day Japanese has developed from /p/ (before A.D.710: Old Japanese (OJ)) through /$/ (Nara era: A.D.710-: Old Japanese) to /h/ (Edo era: 1603-1867: Early Modern Japanese (eModJ)) (e.g. Komatsu, 1981; Miller, 1967). The important change in pronunciation of /$/ took place in the middle of the Heian era (A.D. 794-1184: Late Old Japanese (LOJ)), which is called ‘Ha-gyō tenko on’ [' h a ÿ o  te,g'ko on] (Komatsu, 1981* 29ff.).

As observed in the allophonic variation of /t/ in American English, it is a consonant in intervocalic position which may be voiced and weakened, as in better [ber$] and water [wor$]. It is likely that a similar principle may operate for LOJ /$/ [φ] in intervocalic position. In /ka$fi/ ‘shellfish’, for instance, there are four phonetic segments and each has its own ideal articulatory target. They are successfully coarticulated almost all the time, achieving the four targets to generate the proper phonetic form [ka$fi]. At the onset or during the articulation of [φ], the tongue medio-dorsum may be raised towards the palatal region for preparing the upcoming [i]. If such a gesture misses the right timing and the voicing continues

---

3) According to Komatsu (1981) and Miller (1967), it is widely accepted that OJ (and LOJ) had an eight-vowel system /i, e, i, é, ë, a, o, u/. Assuming that /u/ in OJ is a close back ‘rounded’ vowel and is phonetically realised as [u], we can say that /w/ in OJ and LOJ is a voiced ‘labial’ velar approximant. There is considerable validity to this interpretation. Because of the existence of the central vowels, /u/ may occupy the peripheral position (i.e. close-back) in the phonetic vowel space, in order to maximise perceptual contrast. Also, the phonological system of OJ has a voiceless bilabial fricative as a phoneme /$/ and ‘it phonetically makes up a voiced-voiceless contrast with [w] (my translation)’ (Komatsu, 1981* 50). It seems reasonable to suppose that the phonetic realisation of /w/ in OJ and LOJ may have been accompanied by a higher degree of lip rounding and lip protrusion.
throughout the [θ], then the consonant results in a sound similar to [β] perceptually. Furthermore, if the constriction of [β] becomes opener, [w] will be realised. This misalignment of articulatory timing might have happened in the LOJ period: /ka$ɑ/ [ka$ɑ] ‘river, leather’ has shifted to [kawa]; [ka$ɪ] ‘shellfish’→[kawi]; /ka$ʊ/ [ka$u] ‘buy’→[kau]; but not in /φɑtɔ/ [φɑtɔ] ‘pigeon’→*[wɑtɔ] ([hato]) (examples are taken from Komatsu (1981) and Miller (1967). We call this phenomenon /φ/-allophony. This sound change, taken with the distribution of /j/ and /w/, may be illustrated as in (4):

(4) **THE DEVELOPMENT AND INTERPLAY OF /h/, /w/, AND /j/ ALLOPHONIC VARIATION**

<table>
<thead>
<tr>
<th>/h/</th>
<th>/w/</th>
<th>/j/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/pa, pi, pu, pe, po/</td>
<td>/wa, wi, u, we, wo/</td>
<td>/ja, i, ju, je, jo/</td>
</tr>
<tr>
<td><a href="OJ">pa, pi, pu, pe, po</a></td>
<td><a href="OJ">wa, wi, u, we, wo</a></td>
<td><a href="OJ">ja, i, ju, je, jo</a></td>
</tr>
<tr>
<td>/θɑ, θi, θu, θe, θo/ = ► /wa, wi, u, we, wo/</td>
<td>/a, i, u, e, o/</td>
<td></td>
</tr>
<tr>
<td><a href="OJ">θɑ, θi, θu, θe, θo</a></td>
<td>(LOJ /θ/-ALLOPHONY)</td>
<td></td>
</tr>
<tr>
<td>/hɑ, hi, hu, he, ho/</td>
<td>/wa, -, -, -, -/</td>
<td>/ja, -, ju, -, jo/</td>
</tr>
<tr>
<td><a href="EModJ~">hɑ, ɡi, ɡu, ɡe, ɡo</a></td>
<td>[uɑ, -, -, -, -]</td>
<td>[ja, -, ju, -, jo]</td>
</tr>
</tbody>
</table>

The new pronunciation habit, /θ/-allophony, conflicts with a series of /w/ syllables and appears to intensify the intercategorical connection between the vowels and the two approximants /j, w/. During the LOJ period, the CV combinations of /wi/ and /we/ merged into /i/ and /e/ respectively: /ka$ɪ/ ‘shellfish’, for instance, shifted from [ka$ɪ] to [kawi] by /θ/-allophony, and further to [kai] by the /wi/>/i/ merger. Words such as /jama wi/ ‘living in the mountain’ and /wehɪ/ ‘intoxication’ in OJ become /jamai/ and /ehi/ in present-day Japanese. Thus, it is supposed that in the intervocalic position the string of /w/ plus any vowel, except /a/, merges with a corresponding vowel, and loses consonantality. Orthographically, as can be seen in Table 3, there exist two kana-characters, ‘œ’ for [uɑ] and ‘œ’ for [uɑ], but they are rarely used in present-day Japanese. Also, the majority of people definitely use [o] for /wo/, whose kana representation is ‘œ’.

Turning to the second point, we consider the skewed distribution of palatal(ised) consonants. They are found not only in CV syllables, in which a vowel is [i], but also in CyV syllables, the vowels of which are limited to non-front /a, o, u/. Since the palatal(ised) consonants in this phonetic environment are phonologically contrastive, they must be

---

4) The velars and labials share a common acoustic feature, a low F2 (Ohala & Lorents, 1977). The consonant [θ], when voiced, may be heard as [w], because the formant transitions of these consonants resemble each other. In contrast, the labials (e.g. [b]) and velars (e.g. [w]) differ in the duration of the transition (Miller & Baer, 1983). Raising the dorsum towards the velum may be considered as the ‘re-organisation of gestural parameter’ (Browman & Goldstein, 1991) motivated by the change in the status of syllabic.

5) I presuppose a five-vowel system as in present-day Japanese. However, when we speak of /u/ in OJ and LOJ, we shall assume that it is phonetically realised as [u] (See note 2 above).
specified explicitly: the palatal-glide analysis has commonly been applied. Historically the sequence of a consonant plus /j/ (i.e. CyV) originated from the influence of (Middle) Chinese during the Heian era (Late Old Japanese: 794-) (Komatsu, 1981; Miller, 1967). The contrast between palatalised (i.e. CyV) and non-palatalised (i.e. CV) was established at the end of the Middle Japanese period: the Kamakura era (1185/1192-1330) to the Muromachi era (1331/1392-1602). It is generally agreed that all the syllables before the LOJ period, namely OJ, were CV open syllables: there is no need to make a distinction between ‘syllable’ and ‘mora’ in OJ. The phonologisation of palatalised consonants introduces the new syllable type and causes the change of the internal structure of the syllable. Orthographically CyV syllables are made up of the two kana letters. One is the same as that of syllables with a non-palatalised consonant, while the other is one of the letters in the /j/ series. In arehouse(/kja/ [kjang]), for instance, the first and the second characters are the same as ones used for /ki/ and /ja/, but the latter is reduced in size. CyV syllables are exceptional in the sense that they are written in two letters but are counted as one mora.

For the group of Cy/u/ syllables in Table 2-4, the vowel is systematically represented as a high back rounded [u], rather than a typical unrounded [u]. This is because the active involvement of the lips is noticeable. Take a minimal pair of /kasi/ ‘confectionaries’ and /kasju/ ‘singer’ for example. One could tell the difference in auditory quality of the second syllable without difficulty. In pronouncing [kaçi], an alveolopalatal fricative has sharp and clear fricative quality, whereas the sound has grooved and dark quality in [kaçu]. This darkness (or less bright phonetic quality) is due to both a rounding and protruding gesture of the lips superimposed on the second syllable /sju/. It seems very unusual that the second syllable of /ka.sju/ is pronounced with the same lip gesture as that of /kasi/. In fact, such an articulatory gesture is hardly reducible to one of the properties in a particular segment: the representation of the Cy/u/ syllables may become more precise if we use the symbol of labiality for the preceding consonant such as in [kaç^u]. Such labial activities in question are found not only in the Cy/u/ syllables with [e], but also in those with other consonants such as /kju/ [k^u] and /nju/ [n^u]. This point, though strangely ignored in common phonetic descriptions, can easily be confirmed by observation. Note in passing that, although the devoicing of the high vowels /i, u/ usually occurs in that position, the difference in phonetic quality, or lip activity, prevents the two words from becoming homophonous.

The third point of the skewed distribution concerns a combination of /d/ and a vowel. Synchronically it is undetermined whether the pronunciations [dzi] and [dzu] in standard Japanese are represented by /di/ and /du/ respectively, or by /zi/ and /zu/ respectively. In intervocalic position the two affricates often become fricatives [zi] and [zu], which are the expected realisation of /zi/ and /zu/. That is, the distinction between [dzi] and [zi] and that between [dzu] and [zu], are neutralised. Exceptionally some regional dialects such as in
Kochi and Kagoshima differentiate four pronunciations (Shibatani, 1991). We reserve the combinations of /di/ [d'j] and /du/ [du] for transcribing loanwords such as ['d'inaa] ‘dinner’ and [d'ii, itto jii, a(')serw'uu] ‘do-it-yourself’, otherwise assuming the phonemicisations /zi/ and /zu/ for [(d)zi] and [(d)zu] respectively.

Several observations have shown that there are some characteristic distributions of the allophones: (i) common syllable types are CV and CyV; (ii) CyV syllables co-occur only with non-front vowels; (iii) homorganic sequences of the segments are avoided as in */wu/ and */ji/; (iv) /w/ can constitute a CV syllable only with /a/, and /j/ only with non-front vowels; (v) /di/ [d'j] and /du/ [du] are potential sounds in loanword pronunciation. It must be noted that the syllabary does not illustrate all the possible ‘syllables’ in Japanese, since it excludes closed syllables with final consonants /n/ and C2 (e.g. /hon/ ‘book’, /kit.te/ ‘stamp’) and those with a vowel sequence (e.g. /ai/ ‘love’, /koo/ ‘a tortoise shell’). Furthermore, new segmental combinations that are found in the phonetic form of loanwords do not appear in the syllabary, either. Therefore, the syllabary is regarded as a formal set of possible C(y)V-syllables and syllable final consonants, for the native, Sino-Japanese, and onomatopoeic lexicon and some aspects of loanwords.

2.3 Allophonic Variation before the High Vowels
The consonantal segments and the distribution of their allophones are accounted for by formulating sets of rewrite rules that convert a systematic phonemic representation into a systematic phonetic representation.

2.3.1 Palatalisation, Affrication, and Labialisation
The ‘isation’ terms such as palatalisation, affrication, and labialisation, have been used in two different ways. This is where the ambiguity of –isation terms comes from (Lass, 1984; Wells, 1974). The terms, on the one hand, are applied both in describing static (or postural) characteristics of the articulators and in describing dynamic processes involving the interplay of articulator actions. On the other hand, there are two weighting factors in the use of the ‘-isation’ terms, namely more phonetic or more phonological. In describing secondary articulations the focus is placed on the events in the vocal tract. In contrast, the terms refer to diachronic processes in historical phonology but indicate the part played by mental operations in transformational generative phonology. In this section, we use the terms, palatalisation, affrication, and labialisation, as illustrating allophonic processes: a given underlying representation of a segment, or phoneme, is transformed to a particular systematic phonetic representation (of an allophone) under a particular phonological environment. The processes described by the ‘-isation’ terms are assumed to take place at the higher pre-motor level. We describe the allophonic variations before the high vowels, since other cases are self-explanatory.
The rewrite rules in (5-15) below represent the variant realisations of the underlying segments that make up the entries of the phonological syllabary in Table 2-4.

(5a) /p^/-[p] / elsewhere
(5b) /p^/-[p] / elsewhere
(6a) /b^/-[b] / elsewhere
(6b) /b^/-[b] / elsewhere

(7a) /m^/-[m] / elsewhere
(7b) /m^/-[m] / elsewhere
(8a) /r^/-[r] / elsewhere
(8b) /r^/-[r] / elsewhere

(9a) /t^/-[tc] / elsewhere
(9b) /t^/-[tc] / elsewhere
(10a) /h^/-[ç] / elsewhere
(10b) /h^/-[ç] / elsewhere
(10c) /h^/-[ç] / elsewhere

(11a) /s^/-[ç] / elsewhere
(11b) /s^/-[ç] / elsewhere
(12a) /z^/-[(d)z] / elsewhere
(12b) /z^/-[(d)z] / elsewhere

(13a) /n^/-[n] / elsewhere
(13b) /n^/-[n] / elsewhere

(14a) /k^/-[k] / elsewhere
(14b) /k^/-[k] / elsewhere
(15a) /g^/-[g] ([ŋ])/ elsewhere
(15b) /g^/-[g] ([ŋ])/ elsewhere

There are thus three allophonic processes: the most extensive is palatalisation that is conditioned by the high front vowel /i/. The other two are affrication and labialisation: the former is found in /t/ before /u/ as in (9b) and the latter is found in /h/ before /u/ as in (10b).

We shall focus on the patterns of palatalisation.6)

Palatal(ised) consonants appear before the high front vowel /i/ and before non-front vowels /a, o, u/. The latter cases, as can be seen in the phonological syllabary in Table 3, are

---

6) The velars before /i/ are transcribed as [k] and [g] as in (14a) and (15a). These transcriptions are usually considered for those for the velars in CyV syllables: for instance, the realisations of /kja/, /kjo/, and /kju/, are systematically transcribed as [k'æ], [k'ø], and [k'u]. It is possible, using the IPA conventions, to draw a distinction between a 'fronted (advanced)' velar [k] and a 'palatalised' velar [k:], as suggested by Keating & Lahiri (1993). Since both velar do involve fronting of the closure, the distinction is phonological: the term 'fronted' refers to contextual fronting before /i/, while the term 'palatalised' refers to 'fronting' that creates a surface phonemic contrast. We do not follow this usage and indicate both velar in the palatalising environments by the symbol [^]: our choice is made on the basis of the events that we discuss, rather than abstract phonological contents. It is assumed that the velar in the given context involve the movement of the tongue medio-dorsum, as well as fronting of the closure.

7) A velar nasal [n] in (15a,b) is an allophone of /g/ in intervocalic position. This use of [n] has been considered as one of the characteristics of Standard Japanese: [ka,na'mi] 'mirror', [ka,go] 'basket', [k'age] 'shadow', rather than [ka,ga'mi], [ka,go], [k'age]. However, this pronunciation habit is gradually declining as the form with [g] takes the place of [n]: a velar plosive occurs both in initial and in intervocalic positions. This is a sound change in progress and may frequently be found in the speech of the younger people. In addition, in their speech, the plosive [g] in the intervocalic position may be further reduced to a fricative [y].
phonologically contrastive, being analysed as a consonant plus /j/ (i.e. CyV syllables). For instance, we may find the minimal pairs with /s/ and /sj/ as in (16a) and those with /k/ and /kj/ as in (16b) below.

(16a) /siku/ [çi,km] ‘spread’ - ----
/soku/ ['sjku] ‘in a hurry’ - ----
/saku/ [sa,km] ‘bloom’ - /sjaku/ [çæ,km] ‘serve’
/soku/ ['sokm] ‘at once’ - /sjoku/ [çw,km] ‘job’
/suku/ [su,km] ‘comb’ - /sjuku/ ['çukm] ‘celebration’

(16b) /kiku/ [kç,km] ‘listen to’ - ----
/kekku/ ['kækm] ‘eagerness’ - ----
/kaku/ ['kæk], ‘write’ - /kjaku/ [kçæ,km] ‘customer’
/koo/ ['koo] ‘item(s)’ - /kjoo/ ['kç∫,oo] ‘today’

The combinations */Cj+/+i/ and */Cj+/+e/ are not permissible in the native and Sino-Japanese lexicon. The consonantal phonemes, other than the two semivowels (/j/ and /w/), are susceptible to palatalisation.

The palatalisation phenomenon, as summarised in the rewrite rules in (5-15) above, is not a single unitary articulatory process. It involves not only the superimposition of a stricture of open approximation onto the consonant, but also the modification of the target consonant into a palatal including alveolo-palatales, postalveolars and palatals. Keating (1993)*6 introduces a useful distinction, primary and secondary palatalisation. ‘Primary palatalisation’ suggests that ‘the primary articulation [of a given consonant] is affected and that there is no separate secondary articulation’; ‘secondary palatalisation’ indicates that ‘a secondary [palatal] articulation is added to a primary one [of a given consonant]’. The two modes of palatalisation may help to define different kinds of articulatory processes in a simple way, since they are evidently distinctive compared to other articulators at the systematic phonetic level.

The underlying segments /t, s, z, h, n/ in Japanese are the targets of the primary palatalisation as in the rewrite rules (9-13). On the other hand, the secondary palatalisation, in which the main articulator is unchanged, includes the segments /p, b, k, g, m, r/. Bilabial nasal and plosives are opaque to palatalisation and velars are effectively fronted with their constriction location of velic closure. The consonant /r/ ([r]) can be regarded as an instance of secondary palatalisation. Although in practice the articulation of the sound varies idiosyncratically, it is expected that the articulatory point on the tongue effectively becomes more laminal.

From his survey of 120 instances of palatalisation in various languages, Bhat (1973)*67 mentions that primary palatalisation ‘affects only a limited portion of the consonant system’, while secondary palatalisation, if it takes place at all, ‘appears to affect almost all the
consonants occurring in the language’. This suggests that the two modes of palatalisation together create the secondary oppositions in the design property of a language, which is reflected in the inventory structure or phoneme combinations. Trubetzkoy (1958 /1962) characterises such a particular use of the phonetic feature as the ‘correlation of palatalisation’.

2.3.2 Linguistic Modelling of Japanese Palatalisation: a historical sketch

Secondary phonetic features are used to signal different lexical units. Trubetzkoy (1958 /1969; 129f.) applies the notion of the ‘privative’ and ‘proportional’ opposition to secondary phonetic characteristics, identifying ‘correlations of consonantal timbre.’ A correlation is organised when the consonants in a series of localisations (places and manners of articulation) are split into two series marked or unmarked by a phonetic characteristic. The ‘correlation of palatalisation’ is the opposition between neutral and i- (or j-) coloured consonants. Trubetzkoy includes Japanese as one of the languages whose consonant system demonstrates the two-way contrast with or without palatality. Other such languages are Gaelic, Polish, Russian, and Ukrainian (see Appendix 1). We will sketch out previous attempts to account for the correlation of palatalisation in Japanese, and will point out that, to approach and develop the proper characterisation of palatalisation, we must separate the issue of the articulatory nature of the phenomenon on the one hand, from that of the linguistic nature on the other.

There have been various attempts to reflect the correlation of palatalisation in the phonological description of Japanese. Because of its prosodic nature, palatalisation is a typical example of phonemic indeterminacy, or the multiplicity of possible phonemicisations. Logically there are three possible methods of assigning palatality to the segment*^: (i) the palatalised consonant analysis (i.e. /CjV/); (ii) the palatalised vowel analysis (i.e. /CVj/); and (iii) the palatal-glide analysis (i.e. /CjV/). The first method, which parallels plain consonants with palatalised ones, is commonly applied for the consonant system of Russian (e.g. Jones & Ward, 1969). Although the first and the second methods were once applied (e.g. Hattori, 1949; Uemura, 1972; see Tables 4 and 5), the palatal-glide analysis as in our previous discussion has commonly been adopted for describing the Japanese consonant system. All three representations share the general strategy that the palatalisation phenomenon can be reduced to one segment or a set of segments. Furthermore, all the models suffer from the same linguistic problem, the automatic or programmed nature of palatalisation.

A common view in linguistic modelling is that the contrastive secondary articulation is generally regarded as a phonological matter (i.e. a programmed coarticulation), whereas the non-contrastive one is a phonetic matter (i.e. an automatic coarticulation). In the segmental phonology of Japanese, it is generally agreed that, while the palatalisation in CyV syllables is

8) It must be noted that Firth’s prosodic analysis (Firth, 1948) can be applied to palatalisation (e.g. Henderson, 1966).
phonologically specified, that in C/i/ syllables is mostly predictable. This belief is implied both in the palatalised-vowel analysis by Hattori (1949) and in the palatalised-consonant analysis by Uemura (1972).

Table 2-5 shows the palatalised-vowel analysis by Hattori (1949), in which the three extra ‘palatalised’ vowels /ä, ö, ü/ are added to the plain five vowels in Japanese, yielding an eight-vowel system. Hattori might have attempted to show that Japanese palatalisation is a contextual accommodation before both /i/ and /ä, ö, ü/: these vowels automatically assign the palatality feature to the preceding consonants (the relevant portion is in the dotted square). However, since inclusion of the extra palatalised vowel in the inventory is assumed, it would not make sense to argue whether palatalisation is a contextual effect or not.

In contrast, the palatalised-consonant analysis in Table 2-6 explicitly specifies the palatality by paralleling the palatalised and the non-palatalised consonantal segments.

### Table 2-5: The Palatalised-Vowel Representation of the Japanese Syllabary: Hattori (1949, 171)

<table>
<thead>
<tr>
<th>a</th>
<th>o</th>
<th>u</th>
<th>i</th>
<th>ä</th>
<th>ö</th>
<th>ü</th>
<th>â</th>
</tr>
</thead>
<tbody>
<tr>
<td>ke</td>
<td>ka</td>
<td>ko</td>
<td>ku</td>
<td>ki</td>
<td>kâ</td>
<td>kō</td>
<td>kū</td>
</tr>
<tr>
<td>ge</td>
<td>ga</td>
<td>go</td>
<td>gu</td>
<td>gi</td>
<td>gâ</td>
<td>gō</td>
<td>gū</td>
</tr>
<tr>
<td>te</td>
<td>ta</td>
<td>to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>da</td>
<td>do</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pe</td>
<td>pa</td>
<td>po</td>
<td>pu</td>
<td>pi</td>
<td>pâ</td>
<td>pō</td>
<td>pū</td>
</tr>
<tr>
<td>be</td>
<td>ba</td>
<td>bo</td>
<td>bu</td>
<td>bi</td>
<td>bâ</td>
<td>bō</td>
<td>bū</td>
</tr>
<tr>
<td>ne</td>
<td>na</td>
<td>no</td>
<td>nu</td>
<td>ni</td>
<td>nâ</td>
<td>nō</td>
<td>nū</td>
</tr>
<tr>
<td>me</td>
<td>ma</td>
<td>mo</td>
<td>mu</td>
<td>mi</td>
<td>mā</td>
<td>mō</td>
<td>mü</td>
</tr>
<tr>
<td>re</td>
<td>ra</td>
<td>ro</td>
<td>ru</td>
<td>ri</td>
<td>râ</td>
<td>rō</td>
<td>rū</td>
</tr>
<tr>
<td>he</td>
<td>ha</td>
<td>ho</td>
<td>hu</td>
<td>hi</td>
<td>hâ</td>
<td>hō</td>
<td>hū</td>
</tr>
<tr>
<td>se</td>
<td>sa</td>
<td>so</td>
<td>su</td>
<td>si</td>
<td>sâ</td>
<td>sō</td>
<td>sū</td>
</tr>
<tr>
<td>ze</td>
<td>za</td>
<td>zo</td>
<td>zu</td>
<td>zi</td>
<td>zâ</td>
<td>zō</td>
<td>zū</td>
</tr>
<tr>
<td>cu</td>
<td>ci</td>
<td>cā</td>
<td>cō</td>
<td>cū</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2-6: The Palatalised-Consonant Representation of Japanese Phonological Inventory

<table>
<thead>
<tr>
<th>LABIALS</th>
<th>ALVEOLARS</th>
<th>PALATALS</th>
<th>VELARS</th>
<th>GLOTTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>pj</td>
<td>t</td>
<td>k</td>
<td>kj</td>
</tr>
<tr>
<td>b</td>
<td>bj</td>
<td>d</td>
<td>g</td>
<td>gj</td>
</tr>
<tr>
<td>c</td>
<td>cj</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>zj</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hw</td>
<td>s</td>
<td>sj</td>
<td>hj</td>
<td>h</td>
</tr>
<tr>
<td>m</td>
<td>mj</td>
<td>n</td>
<td>nj</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>rj</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9) The symbols /ă/ and /ç/ indicate [uə] and [ts] respectively. A posited phoneme /ʔ/ reflects the fact that a glottal stop frequently occurs at the beginning and the end of an utterance.
This representation makes it clear that the correlation of palatalisation overlays all the localisations in the consonant system. Because of this specification, the phonemic statement of the distribution of the palatalised consonants deals with the distinction between contrastive and non-contrastive palatalisation.

In phonemic phonology it is presupposed that phonemes are the direct reflection of sounds in an utterance. It follows from this that palatalisation is automatically implemented if the phonemicisation is appropriately dependent on the phonetic observation. Both Hattori (1949) and Uemura (1972) accept that the non-contrastive palatalisation before /i/ is an accommodative coarticulation, therefore there is an automatic smoothing effect. The same is true for the contrastive case, since it is the direct manifestation of a palatalised vowel in Hattori's analysis or that of a palatalised consonant in Uemura's analysis.

In the framework of transformational generative phonology the problem of palatalisation is dealt with as the specification of the feature [sharp] that opposes palatalised segments to non-palatalised. McCawley (1968) claims that all the consonants are specified as having either an underlying [+sharp] or [-sharp] feature, the distribution of which varies according to the Japanese lexical strata (see Table 6 below). In other words, the palatalised-consonant analysis is adopted and the phonological rules to predict sharpness depend on the morpheme features, [+Native], [+Sino-Japanese], [+Onomatopoeia], and [+Foreign]. In addition, the number of the vowels is distinguished for each lexical stratum.

As summarised in Table 2-7 below, the plain and sharp consonants are distinctive in the entire vowel environment for the [+Foreign] vocabulary\(^{10}\). Thus, the underlying system of five vowels is sufficient for the prediction. In the [+Native] lexicon, the sharpness is distinctive only before /a/ as in the quasi-minimal pair, /sjaberu/ [çæ'beriu] ‘chatter’ and /sakeru/ [sa'keru] ‘avoid’ (McCawley's examples and my transcription), and is predictable from the feature [+diffuse] and [+grave] in the other vowel contexts: a sharp consonant before /i/ and a plain consonant before /e, o, u/. From this, McCawley asserts that the underlying

<table>
<thead>
<tr>
<th>(C' denotes sharp consonant, C plain consonant, * occurrence, * non-occurrence)</th>
<th>Native</th>
<th>Sino-Japanese and Onomatopoeia</th>
<th>Foreign</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>C'u</td>
<td>*</td>
<td></td>
<td>×</td>
</tr>
<tr>
<td>Co</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>C'o</td>
<td>*</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Ca</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>C'a</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Ce</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>C'e</td>
<td>*</td>
<td>*</td>
<td>×</td>
</tr>
<tr>
<td>Ci</td>
<td>*</td>
<td></td>
<td>×</td>
</tr>
<tr>
<td>C'i</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

---

\(^{10}\) This, as we will see below, is an over-generalisation of the phonetic forms of loanwords.
distinction between Ca and Cya is a difference in the vowel rather than that in the consonant, suggesting the six-vowel system /i, e, ã, a, o, u/ for the [+Native] lexicon: the sharpness is specified only for the distinctive case by adding the ‘palatalised’ vowel /â/ whose feature is represented as [+diffuse, +compact], namely [+high, +low] in Chomsky & Halle’s system. This vowel system is carried over to the [+Sino-Japanese] and [+Onomatopoeia] strata.

The phonological rules in (17) were devised to cover the distributional characteristics that are unpredictable from the abstract palatalised-consonant analysis.

\[(17a)\] [+consonantal, + Native]→[-sharp]

\[(17b)\] [+consonantal] →[α sharp] / -consonantal

\[(17c)\] [+compact]→[-diffuse]

The presence of the palatalised vowel /â/ guarantees the correct operation of rule (17a), and the feature [+diffuse] in that vowel will be changed by rule (17c) in the course of phonological derivation. The rule in (17b) essentially says that a consonant before /i/ is palatalised in the [-Foreign] lexicon. The process is exemplified by the sequence of /ni/ in /niku/ [ni]ku] ‘meat’ ([+Native]) and that of /nja/ in /haninja/ [hanja] ‘woman devil’ ([+Sino-Japanese]) below.

\[(18a)\] /n/ \[+consonantal\] \[+nasal\] \[+sharp\] \[+obstruent\]

\[(18b)\] /i/ \[+consonantal\] \[+compact\] \[+diffuse\] \[+obstruent\]

\[(18c)\] /â/ \[+consonantal\] \[+nasal\] \[+sharp\] \[+obstruent\] \[+compact\] \[+diffuse\]

In both cases, the surface form is an alveolopalatal nasal [n] that is a case of primary palatalisation. There is no underlying sharpness present in the consonant but it is supplied by the rule.

It is reasonable to suppose that McCawley (1968) attempts to incorporate the secondary opposition into the fundamental part of the phonological description. This is because a set of allophonic rules is a poor reflection of palatalisation as a phonological class of distinctive oppositions in Japanese. However, the six-vowel analysis, in which the /â/ simply functions as a diacritic, is hardly acceptable. Another question arises as to the validity of the distinction between sharpness and diffuseness. Both features suggest acoustic resonance characteristics but are selectively specified for the type of segments, vowel or consonant, even though the two features have a common articulatory correlate ‘frontness’. Furthermore, different articulatory

---

\[11\] McCawley (1968, 70) acknowledges the strange feature specification for /â/: ‘...I will take the position that strangeness is no cause for worry just as long as the grammar is set up in such a way as to prevent that combination of features from appearing in the output of the grammar’.
manoeuvres are required to realise the acoustic effects of the feature [+sharp], since the articulatory processes in palatalisation, as discussed above, are different with respect to the nature of the target consonant. It is inadequate to indicate all the processes in question by simply assigning the feature [+sharp].

Overall, it has become clear that all the modelling explicitly or implicitly assumes that the actual phonetic implementations of palatal(ised) consonants are the same, even though they have distinct phonological origins. Thus, the argument about the relationship between linguistic structures and their potential motor control turns out to be definitional rather than empirical. If we assume that, while the vocal tract in speech production is constrained bio-mechanically, it is organised phonologically, the problem becomes more approachable. This view leads us to another way of understanding the relationship between language-specific allophonic rules in the phonological component and seemingly automatic coarticulation in the traditional speech production model that is assumed in the approach reviewed above. It is possible to conceptually incorporate the notion of coarticulation with linguistically specified assimilation, characterising it as phonologised coarticulation or soft coarticulation (Fujimura, 1990). Such a perspective will be further developed in Chapter 6.

2.3.3 CyV Variability

An interesting variation concerning the syllables that consist of an alveolopalatal fricative and high vowels is found in present-day standard Japanese. We call this phenomenon 'CyV variation'. The CyV sequences of [ç'u] and [(d)z'u] alternate with the CV sequences of [çi] and [(d)zi] respectively (examples from Nakajô (1989), Sakurai (1966) and my informal observations of native speaker's speech: phonetic transcriptions are mine). The standard phonetic forms are written on the left-hand side.

(19) /sjukudai/ [ç'u,kudai]-[çi,kudai] ‘homework’
/sjuzjutu/ [ç'u,utzuj]-[çiz'u,utzuj]-[çizitsuj] ‘medical operation’
/sjurui/ [ç'u,urui]-[çirui] ‘kind (n)’
/sisjutu/ [çi,ç'utzuj]-[çi,çitsuj] ‘expenditure’
/sujuh/ [ç'u,ûhu]-[çifu] ‘house wife’
/seesjuku/ [se,çc'ûku]-[se,ççjku] ‘quietness’
/gaisjutu/ [ga,ic'utzuj]-[ga,icjtsuj] ‘going out’
/gesjuku/ [ge,çc'ûku]-[ge,ççku] ‘lodgings’

(20) /zjumjoo/ [dz'u,m'o0]-[dzi,m'o0] ‘life span’
/zjukuren/ [dz'u,kuren]-[dzi,kuren] ‘expert’
/sinzjuku/ [çi,çz'ûku]-[çi,çziçku]12) ‘Shinjuku (place name)’
/bizjutu/ [b'iz'u,utzuj]-[b'izitsuj] ‘art’
/hançjuku/ [ha,çz'ûku]-[ha,çziçku] ‘soft-boiled (egg)’

12) An allophone of /n/ before [a] is indicated by the symbol [z]. This symbol implies only the tongue shape: during the production of [z], there occurs no audible turbulent noise.
It is supposed that the CyV variability may be a pronunciation habit inherited from
the original Tokyo accent (Sakurai, 1966; 42). Whereas those pronunciations on the
right-hand side in (19) and (20) are not always recognised as Standard Japanese, CyV
variability is lexicalised phonetically and orthographically in the following four words, which
can be represented by different phonemicisations (examples from Akinaga (1981), Nakajô
(1989), and NHK (1998): phonetic transcriptions are mine).

(21) /zjukko/ ['dz^ukko] or ['dzikko] (=/zikko/) ‘ten of something’
/sake/ /3/ ['sake] or ['cæke] (=/sjake/) ‘salmon’
/rjokjaku/ /4/ [r^o,k^ækuld] or [r^o,kakuld] (=/tjokkuld/) ‘traveller’
/kjuuri/ ['k^uuffr] or ['kfruiNi] (=/kiuri/) ‘cucumber’

The phonetic form of /zjukko/, for instance, fluctuates between ['dz^ukko] and ['dzikko]
through CyV variability, and can be written in kana characters either じっこ or じっこ
respectively. Notice here that the small kana letter ‘�’ in the former version is absent in the
latter. It is true that the same orthographic convention can, in principle, be applied to the
examples in (19) and (20) above. However, they very rarely involve the orthographic changes,
even though they are so pronounced with the CyV variability. It is supposed that the phonetic
form with the CyV variability is essentially lexicalised at the allophonic level.

The CyV variability raises a question concerning palatalisation in Japanese. A
common phonetic explanation is that the succession of syllables with a palatal(ised)
consonant causes difficulty in maintaining a standard pronunciation (Sakurai, 1966; Nakajô,
1989). This presupposes the ease, or effort, of articulation and seemingly appeals to
native-speaker’s intuition." Yet, such an explanation leaves the central problem untouched.
Since the resultant phonetic form still contains a palatal(ised) consonant, the ‘difficulty’, if
any, remains. In other words, a speaker does not intend to avoid producing a syllable with a

---

13/ This case may not be considered as the same phenomenon as CyV variability in (19) and (20).
14/ On the other hand, this explanation relies heavily on the changes of the orthographic representation.
In traditional Japanese linguistics and phonetics, there is a distinction, in a pre-theoretical sense,
between choku-on (literally straight sound, or nonpalatalised syllables), which are written by one kana
letter, and yô-on (literally non-straight sound, or palatalised moras), which are written by two kana
letters, one of which is reduced in size. The former suggests CV syllables while the latter is CyV ones.
The CyV variability, therefore, has been considered as the phenomenon that converts yô-on into
choku-on. This interpretation implicitly suggests that the phenomenon is de-palatalisation, which turns
out to be wrong.

As we noted above, CyV syllables were added to the syllabary, or the sound system of
Japanese, during its historical development and display a peculiar distribution. Furthermore, their
orthographic representation and phonetic realisation have special status in the syllabary. This is
certainly the most popular of all delusions about CyV syllables. Consequently, the palatalisation
phenomenon has been attributed mainly to CyV sequences, whereas the characteristic found in CV
sequences where a vowel is /i/, has received little attention. Sakuma (1929; 223) explicitly points out
that ‘the palatalisation phenomenon of /i/ has been strangely neglected by phoneticians and linguists in
Japan’ (my translation).
palatal(ised) consonant. Another explanation appeals to vowel devoicing as a conditioning factor (Nakahō, 1989). This, however, cannot account for the variation in the voiced fricative environment as in (20).

There are two points that are clear in the CyV variability. One is that the target consonant is principally limited to alveolopalatal fricatives [ç] and [(d)z]. The other is that the consonant in question remains palatalised before and after the process of CyV variability. As an allophonic process, two palatalising environments suggests that there may be two distinct kinds of palatalisation, while the phonetic effects of /i/ and /j/ upon the target consonant (/s/ or /z/ in this case) are assumed to be the same, namely generating [ç] or [(d)z]. This paradoxical situation leads us to the question of whether there is any difference between what is characterised as the same palatal segment, and what is characterised as different palatalising environments.

The problem becomes more accessible if we focus on the systematic articulatory control underlying the palatalisation. Two speculations are possible for the nature of CyV variability. First, the phenomenon, given that the effects of the two palatalising environments are somehow different, is a switching between the [j]-element and [i]-element of articulatory gesture for the production of [ç] (or [(d)z]). The change of the vowel from [u] to [i], then, is seen as a by-product of such a gestural reorganisation. Second, given that the articulatory gesture for [ç] (or [(d)z]) is the same regardless of the two distinct palatalising environments, the fluctuation is simply the exchange of the vowels. The core of the question is the articulatory activities during the production of [ç] ([(d)z]) before /i/ and before non-front vowels /a, o, u/, particularly /u/. We will later expand this argument on the basis of the results of EPG experiments.

2.4 Phonetic Segments in Loanword Pronunciation

The Japanese lexicon is traditionally stratified into four strata (e.g. McCawley, 1968): Yamato (the native vocabulary), Sino-Japanese (the technical and learned vocabulary), Onomatopoeia (the mimetic vocabulary), and Foreign (the new technical vocabulary, or loanwords). Itô & Mester (1995a) make interesting observations about the following sets of synonymous quartets.

(22) YAMATO SINO-JAPANESE ONOMATOPOEIA FOREIGN GLOSS
/kagajaku/ /koo/ /kira kira/ /sjain/ ‘shine’
/inu/ /ken/ /wan wan/ /doggu/ ‘dog’

The phonetic form of loanwords is one of decisive factors that cause the fluctuation, or expansion, of the standard pronunciation habits. The emergent patterns in the foreign stratum are obviously different, in some respects, from the patterns in the other three strata. Because
different principles are used for accepting growing pronunciations as established, the
inventory size and its entries differ in various ways. Assuming the concept of ‘core-periphery
structure’ (Itô & Mester, 1995a,b) for the lexical organisation, we characterise the emergence
and types of the phonetic segments in the foreign stratum, relating them to the other lexical
strata. We do not attempt to develop a constraint-based model such as in Itô & Mester,
(1995a,b), the theoretical basis of which is Optimality Theory (Prince & Smolensky, 1993).
Rather, we relate our traditional rule-based description with the qualitative distinction
between the core and periphery in the organisation of the lexicon.

It is proposed by Itô & Mester (1995a,b) that the four lexical strata constitute a
core-periphery structure, the core of which includes Yamato, Sino-Japanese, and
Onomatopoetic lexicons. The Foreign lexicon, the periphery, can be divided into two classes,
assimilated and unassimilated loanwords. Such an organisation is illustrated below\(^\text{15}\).

\[(23)\]

\[
\text{Yamato, Sino-Japanese} \quad \text{Onomatopoeia}
\]

\[
\text{assimilated foreign} \quad \text{unassimilated foreign}
\]

\{ \text{PERIPHERY} \}

\{ \text{CORE} \}

From our observations of the phonological syllabary and allophonic rules, the
following phonological characteristics were found.\(^\text{15}\)

(24a) No tautosyllabic consonant clusters but /Cj/ are allowed in the onset position.
(24b) /t/, /n/, /s/, and /h/ must become [tç], [nç], [ç], and [ç] respectively before the high
front vowel /i/ (primary palatalisation): the secondary palatalisation occurs for
labials, velars, and tap.
(24c) /t/ and /h/ must become [ts] and [ç] before the high back vowel /u/ (affrication and
labialisation).
(24d) There are no phonemic sequences /di/ and /du/. Since voiced fricatives [z(i)] and
[z(üi)] are free variation with voiced affricates [dz(i)] and [dz(üi)] respectively, we
use /zi/ for [(d)zi] and /zu/ for [(d)züi].
(24e) CyV syllables co-occur only with the non-front vowels /a, o, u/.\(^\text{16}\)
(24f) The distribution of /j/ is limited only before the non-front vowels /a, o, u/.
(24g) Only one CV combination is allowed for /w/; namely /wa/.

The elements in the loanword pronunciation (i.e. (un)assimilated foreign) do not entirely

---

\(^{15}\) This diagram is a simplified description of the core-periphery organisation proposed by Itô &
Mester (1995a). They use four labels: from the inner to the outer regions, native vocabulary,
established loans, foreign vocabulary, and unassimilated foreign vocabulary. The ‘established loans’
and the ‘foreign vocabulary’ correspond to Sino-Japanese and assimilated foreign in (23) respectively.
follow these eight phonological restrictions, and demonstrate the segmental combinations and allophones that are never found in the core strata. In addition, the stability of such uncommon patterns varies socially and idiosyncratically with the degree of adaptation and adjustment (i.e. lexicalisation): the phonetic form of a given loanword reflects the assimilation status of foreign words. If we assume that the regulations in (24) are applied to the items in all the strata, the tendency of apparent violations in the loanword pronunciation is considered as 'peripheral weakening' (Itô & Mester, 1995b) of the segmental processes and restrictions: phonological behaviours delimit each stratum; in our case, a broad distinction is made between unassimilated foreign and the other four vocabularies.

The emergent phonetic segments that are nonexistent in the core strata are summarised in Table 2-8. The frequency of the syllable types varies: some occur very rarely, while others are easily observed in everyday conversation.

| TABLE 2-8: EMERGENT SEGMENTS IN LOANWORD PRONUNCIATION |
|-------------------|-----------------------------------------------|
| **STOPs AND CLUSTERS** | **EXAMPLES** | **VARIABILITY** |
| [ti] | [ti'iN] ‘teen’, [ti,iimu- to,iimu] ‘itiθeŋguw (*tiit'inggu)’ | [ti]-[t] |
| | ‘team teaching’, ["pi'aa'tii(*pi'aa'tii)] ‘party’ | |
| [tu] | [turururu] ‘Tours (Fr), [tsuua] ‘tour’, [tsuururu] ‘tool’ | [tu]-[tsu] |
| [di] | [di'siτ'ku] ‘disk’, [rad'iqw'u] ‘radish’, [di,iuq'kaunto] | [di]-[(d)zi] |
| | ‘discount’, [ku'rezito (*kù'red'ito)] ‘credit’ | |
| [du] | [ciνuduwi] ‘Hindu’, [du, ui ίto ju, a'seruðu] | |
| | ‘do-it-yourself’ | |
| [tju] | [kosuτjuumumu-ke'suτjuumumu] ‘costume’ | [tu]-[t] |
| [dji] | [pu, ro'djuusaa] ‘producer’, [djuuaruu] ‘dual’ | |

| **FRICATIVES AND AFFRICATES** |
|-------------------|-----------------------------------------------|
| [φe] | [φe'iι] ‘ferry’, [φe'suτ-φe'suτ] ‘face’ | |
| [φo] | [φooku] ‘fork’, [φoomaruu] ‘formal’ | |
| [φiu] | [φiuuzuz] ‘fusion’, [φiuuzuz- 'ujuuzuz] ‘fused’ | |
| [si] | [si,dii-oi] ‘CD’, [si'guru 'ngurui] ‘single’ | [si]-[si] |
| [(d)zi] | [dzippu-dzippu ‘koodo] ‘zip code’ | [(d)zi]-[(d)zi] |
| [se] | [cer'iι] ‘sherry’, [sepaado-sepaado] ‘shepherd’ | [se]-[se] |
| [tsi] | [totsii] ‘Totsie’ | |
| [tse] | [φi,rentse] ‘Firenze (It)’ | |
| [tsa] | [p'iitta] or [p'iza] ‘pizza’ | |
| [tso] | [ka,n'tsoone] ‘canzone (It)’ | |
| [tee] | [t'e'reiι] ‘cherry’, [t'eeN] ‘chain’, [teek,kw 'in] ‘check in’ | |
| [(d)ze] | [dzeravi] ‘jealousy’, [daiz'esui] ‘digest’ | [(d)ze] | [(d)ze] |

| **APPROXIMANTS** |
|-------------------|-----------------------------------------------|
| [je] | [j,e,ru- j,e,ru] ‘Yale’, [j,e,roo] ‘yellow’ | [je]-[j] |
| [uui] | [dauiN] ‘Darwin’, [uijku- uijku] ‘week’ | [ui]-[ui] |
| [uie] | [υebeυe-υebeυe] ‘wave’, [υeφaasui] ‘wafers’ | [u]-[u] |
| [wo] | [wotcji-wotcji] ‘watch’, [kui'wootaα-kui'wootaα] ‘quarter’ | [w]-[w] |
The violations, or types of peripheral weakening, are divided into three main groups. First, the syllables [tʰi], [si], [zi], and [tu], reveal the weakening of palatalisation and affrication that must be applied in the core strata. It is worth noting that the pronunciation of those syllables fluctuates between 'assimilated' [tçi, çi, zi, tsm] and 'unassimilated' [th, si, zi, tu]. Second, three consonant clusters are created: [tʃə, dʒə, ɸjə]. The syllable [tʃə] is interchangeable with [tçu], the 'assimilated' form, and [ɸjə] is extremely rare. Finally, there occur atypical allocations of consonants. The syllables with [ɸ] are especially noteworthy in terms of underlying segments. The bilabial fricative is eliminated in the inventory of underlying segments on the grounds that the segment is considered as an allophone of /h/ before /u/ (see Table 1). The words ['ʃeri], ['ʃaijaa], and ['ʃooku] in Table 2-3 constitute (quasi-)minimal pairs with ['heri] 'helicopter', ['haijaa] 'heir', and [hooku] 'hawk' ([ʃiito] 'feet' with [çiito] 'heat'). This leads us to the invention of an alternative representation such as /hw/ for the contrastive cases above (Table 2-6 and Uemura, 1972).

In previous studies the phonetic segments above have been a subject of phonological interest. Bloch (1950) tentatively proposes the distinction between the 'conservative dialect', which treats loanwords as fully assimilated to other words in the core lexicon, and the 'innovating dialect', which uses special sound types and combinations for loanwords. Assuming that these two types of the dialects constitute a continuum (i.e. that of the degree of adaptation), idiosyncratic differences in the (loanword) pronunciation are considered as reflecting the number of phonemes in the (individual speaker's) phonological inventory. Bloch considers the variability in loanword pronunciation as a part of phonetic knowledge of a speaker and it corresponds to the inventory size. In contrast, McCawley (1968) suggests that the distinctive segment inventories and feature representations for each of the four distinct lexical strata. Itô & Mester (1995a,b) develop McCawley's idea and assume that the four lexical strata constitute a continuum, namely a core-periphery structure. The authors attempt to show that each lexical stratum is a domain of phonological constraints: the lexical stratification is characterised in terms of how phonological constraints overlap and interact. The constraints are derived not only from segmental processes specific to a particular stratum, but also from morpho-phonological behaviours of the lexical entries. The observations made by Itô & Mester (1995a,b) are given in (25) below.

16) The pronunciation of loanwords has provoked a great deal of controversy. There are various proposals for incorporating emergent segments to the phonological inventory of Japanese. The most recent suggestion is made by Okada (1999). He adds rʰ to the inventory because r/th/ does not undergo palatalisation and the expected realisation [ti] is commonly heard these days (Okada, 1993; 76; my translation): [tɕi] is derived from r̩/ plus /i/ by palatalisation. Although it appears to be true that [ti] is often used, the rʰ/ is infrequent even in the loanword pronunciation. In addition, other emergent segments in Table 4 are ignored in spite of his premise that phonemes are contrastive. Accordingly, the proposal leaves out many important aspects for improving the phonological inventory of Japanese.
(25a) *P: no single (i.e. nongeminate) [p]; Yamato, Sino-Japanese
e.g. /kappa/ 'river imp', /nipponi/; compare [Foreign] /peepaa/ 'paper'

(25b) NT: post-nasal obstruent must be voiced; Yamato, Onomatopoeia
e.g. /tonbo/ [to,mbo] ‘dragonfly’, /sjombori/ [so,mbor,i]; compare [Sino-
‘Santa Claus’

(25c) *DD: no voiced obstruent geminates; Yamato, Sino-Japanese, Onomatopoeia
e.g. /katte/ ‘buying’, /sippai/; failure, /nikkori/ ‘smiling’; compare [Foreign]
/doggu/ ‘dog’

(25d) *#R: no stem initial /r/; Yamato, Onomatopoeia
e.g. [Sino-Japanese] /riki/ ['r-i-k-i'] ‘power’, [Foreign] /ranpu/ ['rampu'] ‘ramp’

(25e) TI: Coronal consonants must be palatalised before /i/; Yamato, Sino-Japanese,
Onomatopoeia. This also applies in assimilated foreign words.
e.g. /tikara/ [tç,kara] ‘power’, /tiku/ [tçikũ] ‘bamboo’, /tiratira/ [tçiratira]
‘fluttering’; compare [Foreign] /tiimu/ [tçiimiu] ‘team’

(25f) CE: no sequence [alveolopalatal C]+[e]; Yamato, Sino-Japanese, Onomatopoeia
e.g. [Foreign] /cer[i]i/ ‘sherry’

(25g) KYE: no sequence [palatalised C]+[e]; all the four strata

(25h) YE: no sequence [palatal glide]+[e]; all the four strata

The peripheral weakening in unassimilated loanwords is considered as the violation
of the constraints in (25e-h): for instance, the constraint *KYE in (25g) is violated, or
weakened, in the foreign stratum and this generates various combinations of a palatalised
consonant and [e] such as [çe] and [tçe]. Itô & Mester (1995b, 186) assume that ‘the
underlying inventory of a certain language (segments, clusters, syllable types, etc.) is
determined indirectly: input it self are not directly regulated, anything at all can in
principle serve as an input’. This implies that a set of constraints determines language-specific
contents of the phonological inventory. This view is attractive but premature. It is not at all
clear how many such constraints are necessary for describing the phonological inventory
of Japanese. Evidently there remain a number of the emergent syllables in Table 2-8 that are not
accounted for by the violation, or weakening, of the constraints in (25). Therefore, it is more
realistic to assume that rather than an unconfident set of phonological constraints, the
conventional allophonic rules and phonotactic constraints are weakened in the peripheral
lexicon, namely loanwords.

2.5 The Production of Vowels and Diphthongs
2.5.1 Vowels
Standard Japanese and most regional accents\(^{17}\) are characterised by a small simple system
with five vowels /i, e, a, o, u/. The quality of those five vowels is commonly described as /i,\n/e~e/, /a~a/, /o~o/, and /u/ with no diphthongisation. It has been observed that they are

\(^{17}\) The systems range from a three-vowel system (/i, u, a/) in the Yonaguni accent of Okinawa to an
eight-vowel system in the Nagoya accent (the standard five vowels and /ũ, ō, æ/) (Shibatani, 1990; 160).
unevenly distributed in the phonetic vowel space. A common vowel quadrilateral for Japanese is skewed towards the high front vowel [i] (e.g. Vance, 1987). This is due to the typical characteristics of the point vowels [i, a, u]. Kiritani et al. (1977) obtained the movements of the tongue in the production of the five vowels by using the x-ray microbeam system. The tongue shapes specific to each vowel articulation were modelled by a set of articulatory parameters (high, low, front, and back) and a given set of the three-formant frequencies (F1, F2, and F3) that were derived from about 1000 time samples of pellet-coordinates. KKK (1978), also using X-ray techniques, recorded the production of [a, i, u, e, o], with a pause inserted between the vowels, spoken by one native speaker. These reports allow us to study the configurational characteristics of the segmental articulation.

Figure 2-1 below shows the highest points of the tongue for the five vowels that are approximated by my visual inspection of the X-ray tracings of the two studies above. The five points are plotted on a ‘polar co-ordinate vowel diagram’ proposed by Catford (1977). We shall summarise the observations by Kiritani et al. (1977b) and KKK (1978) and shall refine their remarks on the tongue and lip configurations within the current framework.

The tongue configuration is related to the elevation of the tongue body.

In the production of the two front vowels [i] and [e] the whole body of the tongue moves upward and forward. Consequently, the middle and lower pharynx are enlarged. The tongue pre- and medio-dorsum is raised towards the palatal region, which is higher in [i] and is slightly retracted in [e]. The tongue tip is usually pushed against the back of the lower incisor in [i]. This gesture is weaker in [e]. The [i] vowel, as will be shown later, demonstrates a large amount of tongue-palate contact. For the open vowel [a] the medio-dorsum is slightly lifted up towards the intersection of the post-palatal and velar regions. The posterior part of the tongue root tends to be pulled back, making the pharynx width narrow. For the vowel [o] the post-dorsum moves up towards the uvula and, at the same time, the anterior part of the tongue root constricts the upper region of the pharynx. For this approximation the anterior portion of the
tongue is effectively moved backwards. In the production of the vowel [u:] the medio- and post-dorsum is elevated towards the back part of the palatal region, or the front part of the velar region. The width between the tongue root and the pharynx wall is large but is less than that of [i]. In addition, the [u:] vowel is classified as a close vowel but the actual tongue height is lower than [i]. The overall profile of the tongue is very similar to that of the neutral configuration as in [a]. (my translation; KKK, 1978; 97-114)

The lip configuration, in relation to the jaw opening, is specific to the five vowels. The vertical dimension of the ‘inter-labial space’ (Laver, 1980, 1994) increases in the order [i, e, a], while the horizontal dimension is kept almost the same.

The close vowel [i] shows a modest horizontal expansion: the vowel is not always accompanied by the horizontal spreading. The lip rounding gesture is actively involved in the articulation of the vowel [o]: the inter-labial space is horizontally constricted. The lips protrude with the upper and lower lips directing downwards and upwards (rather than outwards) respectively. The vowel [u:] has almost the same horizontal coordinate as in [o], but differs in the vertical size: the close vowel [u:] shows the narrowest inter-labial space of the five vowels. This is characterised as the vertical contraction. The features of the lip gesture described above generally correspond to the jaw opening, the degree of which increases in order of [u:]>[i]>[e, o]>[a]. (my translation; KKK, 1978; 97-114)

It is worth noting that, although the two close vowels, [i, u:], are commonly described as unrounded, they essentially differ in the horizontal aspect of the lip gesture: the [u:] vowel does not involve lip spreading. The lack of such a horizontal expansion can be clarified when compared to the Korean [u:], in which the lips are considerably spread in the horizontal dimension. Because the language has a phonological contrast between [u] and [u:] as in [un] ‘exercise’ and [un] ‘silver’, the two close-back vowels are differentiated by close rounding (i.e. horizontal contraction) and by spreading (i.e. horizontal expansion) respectively, in order to maximise the perceptual contrast. The vertical contraction of Japanese [u:] appears to be the habitual muscular setting of the lips. The same specification of the lip position, namely vertical contraction, in principle, applies to that of the non-syllable counterpart, [u:] (/w/), the lip gesture of which is not so close as that of the English [w].

The Cy/u/ syllables, as mentioned in the previous sections, involve the rounding (and often protruding) gesture of the lips. This lip gesture is characterised by the horizontal contraction and it is attributed to both the preceding consonant and the vowel /u/: the feature of the entire syllable. This is closely connected to the reason for the systematic exchange between CyV and CV syllables (see section 2.3.2). Later we will further discuss this rounded configuration in the Cy/u/ sequences and will suggest that the lip gesture in question is phonologically conditioned.

Ladefoged & Maddieson (1996; 295) have suggested that there are two parameters of the lip configuration for vowel articulation, namely vertical lip compression and protrusion. It is noted that Japanese u can be regarded as having compressed lips rather than being simply unrounded.
The observations of the configurational characteristics have shown that the articulatory vowel space is typically stylised by the shaping activities of the articulators such as the tongue (dorsum), the lips, and the jaw opening; and that the space occupied by the Japanese vowel system is skewed towards the close front vowel [i]. This fact leads us to the hypothesis that, although each vowel gesture may vary across various phonetic contexts, nevertheless, the total amount of the space used is limited. This hypothesis is examined by Keating & Huffman (1984).

In their acoustic study of vowel variation, Keating & Huffman (1984) measured formant frequencies of vowel tokens from word lists and from prose passages, spoken by seven native speakers of Japanese. It can be seen in Figure 2-2 below that that there are four degrees of F1 values, [i], [u], [e, o], and [a]; and that there are four degrees of F2 values, [i], [e], [u, a], and [o]. The mean values of F1 and F2 suggest the acoustic pattern where the vowel [e] is comparable with [o] in height; and [u] is comparable with [a] in backness; and [i] is distinguished from the other four vowels in the two features, namely the highest and most fronted of the five. Areas specific to vowels [u, æ, a] and [a, o] are typically empty and are not filled with contextual variations of the vowels*.

The vowels close to the phonologically empty areas exhibit more variation: the variability increases in the progression [i, e]<[o]<[u]<[a]. In the production of the vowel [u] the major variation is found in F2, so that the contextual allophones spread inwards in the acoustic vowel space. It is supposed that the degree of the medio-dorsum involvement for [u] mainly varies across the contexts. On the other hand, the vowel [a] mainly varies in F1 and

---

*19 We may doubt this assertion by Keating & Huffman (1984). When the open vowel is preceded by the palatal(ised) consonant, it is supposed that the higher positioning of the tongue dorsum is carried over to the vowel articulation. The consequence of this is to raise the F2 frequency. In the current study the vowel /a/ in question is systematically transcribed as [æ] (see Table 2-4).
the allophones distribute upwards. This suggests that the tongue root position, rather than the
tongue dorsum, is primarily affected. Keating & Huffman (1984) concluded that the
allophonic variation in the Japanese vowels involves a tendency to fill in the central region of
the space; and that the more peripheral areas are not influenced, or occupied, even by the
contextual variants of the vowels.

Phonologically the Japanese five vowels are specified by the height and the backness
features [high], [low], and [back]. The diagram in (26) shows the phonological specification
of the five vowels.

\[
\begin{array}{c|c|c|c}
& \text{+ high} & \text{+ back} & \text{- low} \\
\text{+ high} & i & u & \\
\text{- high} & e & o & \\
\end{array}
\]

The feature [round] (or [labial]) is not distinctive in the underlying vowel segments. Thus, it
is left to the low-level phonetic rules to specify the realisation of the lip configuration specific
to the five vowels.

Before turning to the Japanese diphthongs, we will note two points concerning the
articulatory and acoustic properties of [a] and [ui] observed above. In the phonological form
of loanword words the open vowel /a/ in Japanese subsumes various [a]-like vowels that are
phonologically contrastive in the source language. This is illustrated in the examples in (27)
where the pronunciation of loanwords and that of their corresponding source words are listed.
The reference pronunciation of the source words is Standard British English.

(27a) [æ] → /a/: manager [ˈmændʒər] → /mʌnˈdʒɪər/, manner [ˈmænə] → /mnən/; cap
(27b) [ʌ] → /a/: cup [kʌp] → /kæp/, colour [ˈkʌlə] → /kərə/, mother [ˈmʌðə] →
/ˈmʌðə/
(27c) [ə] → /a/: start [stɑːt] → /sʌtə/, bath [bɑːθ] → /ˈbɑːθ/, tomato [təˈmɑːtəʊ] →
/ˈtəmɑːtəʊ/
(27d) [ɒ] → /a/: swan [ˈswɒn] → /ˈswɑːn/, cauliflower [ˈkəʊliˌflɔːə] → /ˈkɑːriˌflɔːə/, raw
[ˈkɒlə] → /ˈkɑːrə/

The English vowels [æ, ʌ, ə, ɒ], as well as [a, ɔ], are converted into Japanese /a/. It is
interesting to note that the sequence of a velar stop plus [æ] in (27a) is phonemicised as /kæ/ instead of /ka/. This is because the [æ] is not contrastive in Japanese: the frontal quality of the
velar (i.e. the fronting of velic closure) is replicated as the quality of the consonant: there is a
minimal pair, \textit{cap} and \textit{cup} in (27a,b)\textsuperscript{20}. The vowel [d] is an open vowel but presumably involves the constriction or narrowing of the pharynx. Consequently, there are cases where the [d] vowel is transformed into /o/: sponsor \textsuperscript{[ˈspɔnsə]}→/suˈponsa/, stop \textsuperscript{[stɒp]}→/suˈtoppu/, hot-dog \textsuperscript{[ˈhɒt dɒg]}→/hot,ˈtoʊˈdɒggu/, hospital \textsuperscript{[ˈhɒspɪtl]}→/hosuˈpitaru/. The influence of the spelling and that of the original pronunciation are not negligible, either.

For the vowel [uː] the second point of the contextual variability consists of its major variation in F2. Also, the allophones tend to distribute over a more frontal area in the acoustic vowel space. This frontal quality of [uː], as suggested above, corresponds to the magnitude and timing of the medio-dorsum raising gesture. It has been observed that the [uː] vowel tends to acquire a more frontal quality when preceded by [s, (d)z, ts] such as [stʊ̯,ztʊ̯me] ‘sparrow’, [tsʊ̯̯,ztʊ̯̯mɨ] ‘drum’, and [tёizʊ̯] ‘map’ (e.g. Sakuma, 1929): the quality may be transcribed by using a centralised allophone [ʊ̯]. This pronunciation habit can be regarded as one of the features in standard Japanese, as well as a regional feature that characterises particular accents such as that in the Tohoku area. In addition to the centralised [ʊ̯], the close-front vowel [i] is considerably retracted, or centralised, in that regional accent, and is transcribed as [ɨ]: the raising gesture of the tongue pre- and medio-dorsum is weaker in degree and the horizontal dimension of the lip gesture is only weakly expanded. This makes the quality of the vowels [ʊ̯], to some extent, similar to that of [i]. Thus, the words such as [ˈtёizi] ‘governor’, [ˈtёizʊ̯] ‘map’, [tsʊ̯,zi] ‘street corner’, become homophonous: [tёizɪ], [tɛi̯zɪ], [tsʊ̯̯zɪ]→[tsʊ̯̯zʊ̯]. Notice here that the qualitative merger between [i] and [ʊ̯] causes the shift of consonants from [tʃi] to [(tsi~)tsiː]. This phenomenon is essentially the same as the CyV variability in standard Japanese: the variation involves the exchange of syllable types between CyV and CV (see section 2.3.3): [ʦuˌkudai]→[ɕiːkudai] ‘homework’, [dzum′oo]→[dʑim′oo] ‘life span’.

2.5.2 Diphthongs

Diphthongs, as opposed to monophthongs, constitute a categorically distinct class, which is commonly characterised as a gliding sound from one vowel to the other, giving a continually changing auditory quality (e.g. Catford, 1977; Jones, 1960; Gimson & Cruttenden, 1994). Whereas a language like English has been described as having a diphthong system in the phonology, Japanese has not. As we have seen, it is characterised by a small simple system

\textsuperscript{20} There is an apparent counter-example: gas [ɡæs]→/gəsu/ (*/ɡjasu/).

\textsuperscript{21} An alveolopalatal fricative [ɕ] must be included in the conditioning consonants: in Akita and Aomori accents the words [stʊ̯̯,ɡi] ‘sushi’ and [ɕiːɕi] ‘lion’ are homophonous [ɕiːɕi]. Those words are also homophonous in Fukushima and Sendai accent, but differ in vowel quality, [stʊ̯̯stʊ̯]. Such an ‘enhanced’ centralised quality of [uː] in the fricative context, as we will show, is partly due to the strong coupling effects between the tip/blade and the dorsum components of the tongue in the production of fricatives and a fricative portion of affricates.
with five vowels /i, e, a, o, u/, where the back vowel /u/ is typically realised as unrounded [u] with the horizontal contraction of the lip gesture (i.e. without the vertical expansion, or lip spreading). This yields twenty-five combinations of two-vowel clusters that include five consisting of two identical vowels (long vowels).

The judgment of trained phoneticians in fact disagrees as to the description of these non-identical vowels. Some claim that there are no diphthongs in Japanese, while others admit diphthongs but say that their types are limited. In the pronunciation of the word *tai* (‘sea beam’), for instance, the vowel in question, although it sounds very similar to *tie* in English, is considered a succession of [a] and [i] (i.e. two different monophthongs) by Akamatsu (1994) and Bloch (1950), but a diphthong [ai] (or [aɪ]) by Edwards (1903), Hattori (1984), Kawakami (1977) and Sakuma (1929). Strangely, however, not all the two-vowel clusters are diphthongs. Edwards (1903) admits [aɪ, oɪ, tʊi/ʊi, ɛi, ʊə, təɪ, æə, aʊ, ʊə]: Kawakami (1977) lists [ai, oi, uti, ae, ao, oe]: Hattori (1984) allows only /i/ [i] to be the second position of Japanese ‘diphthongs’. The core of the problem lies in the characterisation of the dynamic property of diphthongs.

In the following discussion we shall attempt to reconcile those views. We use the term diphthong for the non-identical vowel sequences in Japanese. Our discussion is limited to the comparison between closing diphthongs in Japanese and those in English, because this diphthong type shows the most common correspondence between the two languages. We demonstrate that, if we properly characterise the nature of phonetic parameters, the phonetic categorisation of diphthongs, both within and across languages, can be interpreted in a straightforward manner. This idea will be further exemplified by comparing the phenomena involving a closing diphthong: smoothing in British English; and glide formation in the Japanese pronunciation of loanwords.

The most important finding of earlier phonetic and acoustic investigations (of American English) is that not all the English diphthongs share exactly the same phonetic properties (e.g. Pike, 1947; Lehiste & Peterson, 1961; Gay, 1968; Holbrook & Fairbanks, 1962). Lehiste & Peterson (1961), using a spectrographic technique, examined various nuclear vowels in CNC monosyllable words spoken by American English speakers. It was reported that complex nuclei in a syllable were grouped into two categories: one is /ei, oʊ/ as in *fate* and *lope*, where there is a single target position (steady-state) followed by a transitional glide; and the other is /ai, au, ʊ/ as in *fight, lout, and voice*, where there are two target positions. They concluded that /ei, oʊ/ ‘should not properly be classified as diphthongs (276)’. This agrees with the claim made by Pike (1947) that, while /et/ and /oʊ/ act as phonetically complex single units (single phoneme), /at, ɔt, au/ function as sequences of two units (two phonemes). These conclusions are attainable when the influence of different speaking rates is considered. Gay (1968) examined the effects of different speaking rates (slow, moderate, fast)
on the formant movements of diphthongs /ei, ai, oi, au, o/, and found that, while offset steady-states become considerably shorter, only /au/ and /ou/ contain relatively prominent steady-states at onset for all conditions. Steady-states are least prominent in /ei, o/.

Instead of regarding the fact as a license for eliminating the diphthong, we assume that such acoustic properties reflect the articulatory gestures underlying different complex vowels. We view diphthongs as consisting of two elements: articulatory gesture corresponding to the first element (V₁); and that to the second element (V₂). This allows us to assume that diphthongs can be identified in terms of various degrees of gestural overlap between a V₁ gesture and a V₂ gesture. Catford (1977; 215f.) proposes two extreme types of diphthong, a sequential type such as [ai], [ɔi] and a gliding type such as [ei], [ɔu]. The two types differ in how far vowel gestures travel in the phonetic vowel space. It is possible to align the diphthongs on a continuum and to indicate how they relate to each other.

\[
(28) \quad [ei] \quad [ou] \quad [ou] \quad ([au] \quad [ai]) \quad [ɔi]
\]

There is articulatory evidence to support two extreme types of gestural overlap in (28). Kent & Moll (1972) obtained a cinefluorographic film of lingual movements during the production of the diphthongs /ai/ and /oi/ in American English. Using a two-point parameterisation of the tongue body (anterior and posterior), the patterns of tongue displacement are detected. During the production of /ai/, the posterior point (the mideo-dorsum) is relatively stable, while the anterior point (the pre-dorsum) is considerably raised for the final configuration of the second element. In the course of producing /oi/, the posterior point of the tongue moves downwards, with the anterior point moving upwards. It is concluded that the general patterns of tongue point displacement are sufficient to distinguish two diphthongs. The observation that the initial and terminal vocal tract configurations are not invariant attributes of diphthong production, agrees with one conclusion shared by previous acoustic studies: the first and the second elements of diphthongs are not the same as two monophthongs, as is implied in the way in which diphthongs are usually transcribed (Gay, 1968; Lehiste & Peterson, 1961; Holbrooks & Fairbanks, 1962; Wise, 1965).

Comparing Japanese [ai] with English and French, Akamatsu (1994; 51) asserts that 'a succession of two different vowels in Japanese are pronounced as a succession of two different monophthongs in such a way that they are pronounced in neither a decrescendo nor crescendo manner (i.e. with an equal articulatory and auditory weight) and the articulatory and auditory passage from one to the other is abrupt and clear-cut'. This implies that there must be a definite boundary between monophthongs and diphthongs. Successive vowels can be pronounced one after the other in a diphthongal manner, whether the given language has a
diphthong system in the phonology or not. This is simply because speech production is
dynamic in nature and phonetic segments are coarticulated with each other. There are
examples where a disyllabic sequence of English [u:i] is pronounced monosyllabically (i.e. in
a diphthongal manner). This is called ‘compression’ (Wells, 1990): chewing ['tju:ij->tjuin],
doing ['du:ij->duin], fewest ['fju:ist->fjuist], fluid ['flu:id->fluid], queuing ['kju:ij->
kuin], ruin ['ru:in->ruin]. Furthermore, the ‘succession of two different monophthongs’ in
Japanese tends to coalesce into a long vowel with intermediate quality. We call this
‘monophthongisation’, which is a stylistic modification in Standard Japanese but a usual
pronunciation habit in some regional dialects such as those in the Tohoku and Kanto areas:
['ci'roi->'ci'ru] ‘wide’, [ka,ru(ka)->ka,ka] ‘buy (+question particle ‘-ka’)’. These two
phenomena, compression in British English and monophthonging in Japanese, suggest that
the two consecutive vocoids tend to behave similarly to each other in terms of the changes in
articulatory structures involved, even though there are underlying phonological diversities.

It follows from the gesture-based perspective in (28) that the psychophysical
parameters, underlying a phonetic category, form a continuum ranging from a typical
exemplar of a diphthong to that of two consecutive vowels. This assumption allows us to say
that the discrepancy in the phonetic judgment on Japanese diphthongs comes from the
question of to which end of the continuum the complex vowel belongs. This suggests that the
variability of the psychophysical parameters simply represent phonetic individuality of the
complex vowel production among languages, rather than the basis of the dichotomous choice
between a diphthong and a two-monophthong cluster. We may develop the model in (28) and
propose the following idealised continuum of two extreme overlap types. It is assumed that
the individual ranges of gliding and sequencing gestures are determined language-specifically
and are covered with a general scale of gestural overlap at a higher level.

The diagram in (29) captures the differences between English and Japanese
diphthongs as the degree of overlapping gestures, rather than the arbitrary choice between the
phonetic categories.

The Japanese diphthongs occupy a more ‘sequential’ area of the overall continuum,
suggesting that they may be less influenced by the different speaking rates than English
diphthongs. The data reported by Dolan & Mimori (1986) supports this view.
Dolan & Mimori (1986) examined variation in the production of Japanese /ai, oi, ui, au/ with two accent patterns, and that of American English /ai, ei, oi, au, o/ with no contrastive stress, at three different speaking rates (slow, normal, fast). They measured F2 frequency of the starting and the terminal vowels, and found that: mean F2 values for the terminal vowels in Japanese reveal remarkable constancy across rate changes; mean F2 ranges between the onset and the offset vowels indicate a similar tendency; the effect of the accentual pattern on the variability proved non-significant. Overall, Japanese diphthongs exhibit far less variation across the different speaking rates than English diphthongs. Although Dolan & Mimori make no attempt to establish cross-language similarities, their data for English and Japanese helps us to develop the model proposed in (29).

Figure 2-3 below is a reproduction of the mean F2 range values reported in Dolan & Mimori (1986; 135): the diphthongs in each language are aligned in increasing order of the F2 range. The two languages show distinctive characteristics in terms of the relative stability of the F2 movement values across different speaking rates. It can be seen that each language indicates both gliding and sequential types of gestural overlap for the diphthong production. The Japanese F2 ranges, except /au/, are distributed over a higher area, which overlaps the F2 ranges of English /ai/ and /oi/. The unusual characteristics of Japanese /au/ may be related to the fact that Japanese /u/ is typically unrounded and is articulated at a more central area in the phonetic vowel space. These results suggest that the articulatory gestures for the second element of Japanese diphthongs reveal constantly achieved targets.

The observation that the second element of Japanese diphthongs displays a constantly achieved target needs further clarification. Sakuma (1929; 106) mentioned that ‘in pronouncing [ai] the tongue-position does not reach [i] but actually somewhere around [e] (my translation)’. In other words, the tongue body raising gesture undershoots the target (position) in the production of the diphthong [aɪ]. This seemingly contradicts the observation of relative stability of the second element in Japanese diphthongs. As we will see later in the results of electropalatographic experiments, the target undershoot occurs in the ‘spatial’, rather than ‘temporal’, domain of the lingual articulation: there is no substantial divergence.

![Figure 2-3 Mean F2 Frequency Range of Five American English Diphthongs and Four Japanese Diphthongs at the Three Different Speaking Rates (based on Table in Dolan & Mimori, 1986; 135)]

- Fast
- Normal
- Slow

X-axis = diphthongs in the two languages
Y-axis = difference between F2 for V1 and that for V2 in Hz.
between the amount of tongue-palate contact for [i] and that for [j] in all the contexts. However, the two gestures significantly differ in the duration of their tongue-palate contact and in the coarticulatory influence from the surrounding vowels: the [i]-gesture is longer in duration and resistant to coarticulatory effects. The tongue dorsum raising gesture for [i] is reduced in its spatial extent but not in its temporal extent. This result implies that there is a temporal requirement that underlies a constantly achieved (acoustic) target in the second element of Japanese diphthongs: such a durational feature is related to the moraic status of \( V_2 \).

Fujisaki & Horiguchi (1979) conducted production and perceptual experiments using all the possible combinations of the five Japanese vowels. It was found that the ‘production’ onset of the transition to the second vowel varies considerably with the combinations of the two vowels; the ‘perception’ onset of the second vowel, in contrast, is concentrated within a narrow range regardless of the particular vowel combination. Fujisaki & Horiguchi (1979; 159) suggested that ‘the apparent diversity at the level of speech production may be the consequence of the speaker’s effort to maintain the uniformity of perceived syllabic durations regardless of vowel combination’.

Several observations in the last few paragraphs have shown how the concept of articulatory continuum captures both similarities and differences between the diphthong articulations in the two languages. We have distinguished English diphthongs from Japanese ones in terms of the degree of gestural overlap, gliding and sequencing. It is worth noting that these two types of overlap are partly constrained by production requirements for a particular diphthong. The articulation tends to become sequential as the relative distance between the \( V_1 \) and \( V_2 \) gestures increases. By changing speaking rates the \( V_2 \) gesture in English diphthongs undershoots both spatial and temporal targets. In contrast, the \( V_2 \) gesture in Japanese diphthongs tends to maintain the temporal target, but not the spatial target. It is plausible to assume that such language-specific articulatory patterns reflect the internal structure of a prosodic unit. Given assumptions about the relationship between prosodic constituency and articulatory timing, we could propose the representations of English and Japanese closing diphthongs as in (30) below.

(30a) \[
\begin{array}{c}
\sigma \\
\mu_1 \\
\mu_2 \\
C \\
t \\
a \\
V \\
V \\
C \\
V \\
V \\
C \\
V \\
im \\
\end{array}
\]

(30b) \[
\begin{array}{c}
\sigma \\
\mu_1 \\
\mu_2 \\
C \\
t \\
a \\
V \\
V \\
C \\
V \\
V \\
C \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
V \\
mora (\(\mu\)). Notice in passing that the diphthongs in each language make up one syllable. Differences in the alignment of a timing unit in (30) reflect the gliding nature of English diphthongs on the one hand and the sequential nature of the Japanese diphthongs on the other: the two gestures for \(V_1\) and \(V_2\) are relatively cohesive in English (30a), while they are relatively disjointed in Japanese (30b). Even though we consider the production requirements, it is reasonable to suppose that the general type of gestural overlap in a language is largely dependent upon language-specific aspects of timing control. We may go on from this to develop the following assumption about the general model of the diphthong articulation: two vowel elements forming a single moraic constituent, in principle, are more cohesive, or more coproduced in articulatory timing, than those forming two separate moraic constituents.

2.5.3   Smoothing and Glide Formation: Two Sides of the Same Coin?

There are two phonetic processes that exemplify the above characterisation of the diphthong production in English and in Japanese: smoothing in British English and glide formation in Japanese loanword pronunciation. We shall argue that the two processes are language-specifically chosen to realise one general constraint, namely 'to avoid three-vowel cluster'; and that the choice depends on the phonetic status of diphthongs in each language.

Let us begin by reviewing smoothing in British English. Smoothing, which is a usual pronunciation habit in mainstream RP, is the process of monophthonging a diphthong in a prevocalic environment. The resultant monophthong retains the quality of the starting-point of the underlying diphthong (Wells, 1982* 238ff.). Although this process applies to /i:/ and /u:/, we are not concerned here with them. (All examples from Gimson, 1980, Gimson & Cruttenden, 1994 and Wells, 1982.)

(31a) [ei] — [e i]: chaos ['keio-^' kems], conveyor [kon'veiə→ kon'veiə], layer ['leio-^leio], player ['pleio→pleio], saying ['seiio→seiio]
(31b) [ai] — [a i]: buyer ['baiə→ baia], drier ['draio→ draio], empire ['empaiə→ empaiə], fire ['faiə→ faiə], higher/hire ['haiə→ haiə], sapphire ['saiəfaiə→ saiəfaiə], science ['seioəns→ saiəns], tire ['taiə→ taiə], trial ['traio→ traio], trying ['trainJo→ traio]
(31c) [oi] — [o i]: annoying ['o'naiiŋ→ o'naio], buoyant ['buiaont→ buiaont], employer ['im'plaiə→ im'plaiə], royal ['rəioi→ rəioi]
(31d) [au] — [a u]: mower ['moaio→ moaio], going ['gouin→ gaiiŋ]
(31e) [au] — [a u]: coward ['kauəd→ kəuəd], flower/floor ['fiəluə→ fiəluə], our ['auə→ aua], Howard ['hauəd→ hauəd], ploughing ['plouiaŋ→ plouiaŋ], safflower ['saiəfaiə→ saiəfaiə], shower ['ʃauər→ ʃauər], sour ['sauə→ sauə], tower ['təuə→ təuə]
(31f) they are ['ðeiə→ ðeiə], way out ['weiəut], my aunt ['maioiənt], boy and girl ['boiəgaiəl], go away ['gaiəwei], go off ['gaiəf], so early ['saiəli], how odd ['hauəd], buy a house ['bəiəhaus], now and then ['naioiəden]

There are two distinct types, smoothing by deletion and by coalescence. When the underlying closing diphthong ends in /i/ (31a-c), the second element is deleted and the first
one is lengthened. When the underlying closing diphthong ends in /ə/ (31d-e), the two elements merge with each other and result in an intermediate vowel [ə:] and [əː]. Among the five closing diphthongs in (31a-e), smoothing of /ə/ is not common (Wells, 1982; 238). Smoothing also applies across word boundaries as in (31f).

Wells (1982; 239) mentions that the smoothed forms of /aːə/ [əːə] and /auə/ [əːə], in particular, are further reduced to a long monophthong, where the third element, [ə], disappears (Monophthonging). Together with syllabicity loss of /ə/, we can list at least five possible realisations that are derived from the interaction of smoothing and monophthonging.

(32) /empaio/ (33) /jauo/
   (32a) ['empalə] 3 (33a) ['jaʊə] 2
   (32b) ['empalə] 2 (33b) ['jaʊə] 1 syllabicity loss of /ə/
   (32c) ['empaː] 3 (33c) ['jaː] 2 smoothing
   (32d) ['empaə] 2 (33d) ['jaː] 1 smoothing + syllabicity loss of /ə/
   (32e) ['empaː] 2 (33e) ['jaː] 1 monophthonging

Jones (1960, 106) points out that monophthonging (which he calls levelling) may not apply when the /ə/ is followed by a dark /l/: trial ['traɪæl→'traɪəl→'traɪəl→*traɪəl], but possibly the trial ended ['ðə traɪl endid].

Most of us would accept that the three processes, smoothing, syllabicity loss of /ə/, and monophthonging, are a simplification process of the complex vowel. In addition, we distinguish the function of smoothing and monophthonging on the one hand, from that of syllabicity loss of /ə/ on the other. Since the words in (31-33) will maintain the original number of syllables after smoothing and monophthonging are applied, the processes essentially produce the qualitative change of a target sequence. Smoothing, thus reflects the weakness of the last element of diphthongs (Gimson, 1980, 141). This view can be expanded to include the case of smoothing with syllabicity loss of /ə/ and monophthonging: the second element of diphthongs, either underlying or derived, is weaker than the first element. Syllabicity loss of /ə/, in contrast, produces a quantitative change that reduces the original number of syllables, and perhaps is independently motivated in connected speech.

We now turn to the Japanese pronunciation of the loanwords corresponding to the words in (31). The data of loanwords, which is illustrated as a set of an original English word and the phonetic transcription of standard Japanese pronunciation, is given in (34a-d) below.

(34a) [eta]→[e(e)jaa]: conveyor [ko,m'bejaa~kombejaa], player [pu'u'reejaa], layer [reijaa~rejaa]
(34b) [aia]→[aija(a)]: drier [do,raijaa], empire [em'paija], fire [faijaa], higher/hire [hiraijaa], sapphire [sa'faijaa], science [saiensiu], trial [to'raijaru→to'raijaru], wire [uqa,ija~uqa,ija]
(34c) [oia]→[ojiaa]: empplayer [empu'u'roijaa]
(34d) [au9]^[aiqa(a)]: flower/flour [fu'rauqaa], our [a'qaa], Howard [ha'qaadə], safflower [sa,fu'rauqaa], shower [ca'aqaa], sour [souqaa], tower [touqaa]

76
Some words admit more than one pronunciation, while others admit only one. Such variability is approximated on the basis of different orthographic representations used: trial, for instance, is written as either ‘トライアル’ or ‘トライヤル’.

The conversion of the original English words systematically involves a palatal approximant /j/ (\[j\]) as in (34a-c), where the original closing diphthong ends in [i], and a velar approximant /w/ (\[w\]) as in (34d), where the original closing diphthong ends in [u]. Given that the underlying form of the loanword pronunciation is developed from the original phonetic form, through default operations such as the vowel quality adjustment and the consonant replacement, then we could illustrate the derivational process in (35a,b) below.

Whereas a palatal approximant [j] in (35a) is inserted (i.e. glide insertion), a velar approximant [u] in (35b) is derived from the underlying back vowel (i.e. gliding of a back vowel). We use the term glide formation to cover both operations.

Further examples of the glide insertion occur in the following loanwords.

Further examples of the glide insertion occur in the following loanwords.

(35a) [ˈfaɪə] (35b) [ˈjauə] input

//faɪə// //ˈcaʊə// gliding by default operations

ˈfaɪja ˈcaʊəa glide-formation and vowel lengthening

ˈfaɪja ˈcaʊəa output

Our/hour [ˈaɪə], papaya [ˈpaɪə], power [ˈpaʊə], royal/loyal [ˈroɪə], wire [ˈwə]

Whereas a palatal approximant [j] in (35a) is inserted (i.e. glide insertion), a velar approximant [u] in (35b) is derived from the underlying back vowel (i.e. gliding of a back vowel). We use the term glide formation to cover both operations.

Further examples of the glide insertion occur in the following loanwords.
Some of the words in (36a) contain a sequence of /aɪə/ in their original English pronunciation, so they are susceptible to smoothing. In the examples in (36b) and (36c), the fluctuation of the phonetic form with or without /j/ occurs in a sequence of a non-low (and front) vowel and low vowel such as /aɪ/ and /eɪ/. Although some words are described as having only one pronunciation, there is always a possibility for pronunciation with an epenthetic [j]. It seems that the /aɪ/ sequence is more often susceptible to the glide insertion. In addition, the glide epenthesis is often found in the pronunciation of the non-foreign lexicon. (Phonetic transcriptions are based on my informal observation of native speakers’ speech and epenthetic [?] and ["] are bracketed.)

(37) /ianrjokou/ [i(̂)ə.nr'ikoo] ‘overnight trip for employees’
/omiai/ [o.m'iai] ‘formal meeting for marriage’
/siriai/ [si.'riai] ‘acquaintance’
/iai/ [i.ai] ‘a technique of a sword’
/kaiage/ [kai.'iage] ‘purchase’
/takuan/ [ta."kuan] ‘pickled radish’

It appears that such a variation occurs lexically and idiosyncratically.

The phonetic processes observed in the last section are very systematic if each language is examined independently. The patterning of the smoothing and the other two phenomena in British English is summarised in (38a-c).

(38a) smoothing (by deletion and by coalescence): V1V2=closing diphthongs

\[
\begin{array}{c}
\text{[\ldots CV}_1\text{V}_2\ldots ] \\
\text{V}_3(C)\ldots
\end{array}
\rightarrow
\begin{array}{c}
\text{[\ldots CV}_1\text{V}_1\ldots ] \\
\text{V}_3(C)\ldots
\end{array}
\]

e.g. fire ['faɪə→'faɪə], tower ['taʊə→'taʊə]

(38b) syllabicity loss of /a/

\[
\begin{array}{c}
\text{[\ldots CV}_1\text{V}_2\ldots ]
\end{array}
\rightarrow
\begin{array}{c}
\text{[\ldots CV}_1\text{V}_2\ldots ]
\end{array}
\]

\[\text{(}\text{V}_1\text{V}_2\text{)}\]

e.g. fire ['faɪə→'faɪə], tower ['taʊə→'taʊə], fire ['faɪə→'faɪə], tower ['taʊə→'taʊə]

(38c) monophthonging

\[
\begin{array}{c}
\text{[\ldots CV}_1\text{V}_1[\text{[}\ldots ]\ldots ]
\end{array}
\rightarrow
\begin{array}{c}
\text{[\ldots CV}_1\text{V}_1\ldots ]
\end{array}
\]

e.g. fire ['faɪə→'faɪə], tower ['taʊə→'taʊə']
The general form of the glide formation in Japanese pronunciation of loanwords is illustrated in (39a,b) below.

(39a) /j/ insertion: V₂=/i/, V₃=/a/

\[
\begin{array}{c}
V_3(C) \\
\end{array} \quad \rightarrow \quad \begin{array}{c}
V_2(C) \\
\end{array}
\]

e.g. fire ['faiə]→['faija]a

(39b) gliding of /u/: V₁=/a/, V₂=/u/, V₃=/a/

\[
\begin{array}{c}
V_3(C) \\
\end{array} \quad \rightarrow \quad \begin{array}{c}
V_2(C) \\
\end{array}
\]

e.g. tower ['taua]→['taəja]a

These five phenomena, (38a-c) and (39a,b) can be formalised as language-specific rules that may or may not apply to a particular diphthong in a particular environment. Yet, such a formalisation at best reflects the superficial regularities in the examples, simply emphasising the autonomous characteristic of each process in each language.

If, however, we assume that the ‘target string’ is a sequence of three vowel segments, V₁V₂V₃, where the V₁V₂ is a closing diphthong and the V₃ is another vowel, then one general tendency emerges from the five processes across the two languages (cf. Gimson, 1980: 140). The general form of the processes in British English and that of Japanese loanwords are shown in (40) and (41) respectively.

(40) \[
\begin{array}{c}
CV_1V_2V_3(C) \\
\end{array} \quad \rightarrow \quad \begin{array}{c}
CV_1V_3(C) \\
\end{array}
\]
\[
\begin{array}{c}
\nearrow j \\
\end{array}
\]

(41) \[
\begin{array}{c}
CV_1V_2V_3(C) \\
\end{array} \quad \rightarrow \quad \begin{array}{c}
CV_1V_2 /j/V_3(V_3) \quad \rightarrow \quad \begin{array}{c}
CV_1V_3(V_3) \\
\end{array}
\end{array}
\]
\[
\begin{array}{c}
\nearrow w \\
\end{array}
\]

In (40), the target sequence of V₁V₂V₃ is simplified, in phonetic quality, to the first element V₁, by smoothing with syllabic loss of /a/ and by monophthonging. It is reasonable to suppose that both processes work towards one aim to avoid the surface realisation of the three-vowel sequence qualitatively and quantitatively. This observation is not incompatible with the examples of glide formation in Japanese. In (41), glide formation, which selectively applies to a V₁V₂V₃ cluster, produces the same effect as seen in (40) in a different way, that of breaking up the three-vowel sequence in the underlying form.

From the articulatory point of view, it is not harmful to regard the three-vowel cluster as composed of three distinct vowel gestures in the vocal tract. When we look at lingual activities for a V₁V₂V₃ sequence, an articulatory trajectory starts in the open zone,
moving upward and then downward in the phonetic vowel space. This travel effectively generates continuously changing acoustic effects and needs to be made, with precise timing control of the articulator, for implementing each vowel segment in the string.

We are now in a position to say that the three processes, smoothing with syllabicity loss of /æ/, monophthonging and glide formation, are the modification, or reorganisation, of articulatory structures. This is language-specifically chosen to avoid creating a $V_1V_2V_3$ sequence at the surface phonetic realisation. One possible explanation for this language-specific choice can be derived from language-specific production strategies for the diphthong articulation. As discussed in the previous section, we have distinguished English diphthongs from Japanese ones in terms of the degree of gestural overlap, gliding and sequencing. We have pointed out, given the prosodic constituency proposed by traditional prosodic phonology, that these two types of overlap reflect the internal dominance relationship. In English the two gestures for $V_1$ and $V_2$ are relatively cohesive since they are aligned under a single timing unit (mora). In Japanese, in contrast, the two gestures are relatively disjointed since they are associated to two timing units. Consequently, vowel gestures in English are more coproduced, or blended, with each other than those in Japanese. This articulatory nature of the diphthong production underlies the language-specific choice of a phonetic process to avoid creating the three-vowel cluster at the surface phonetic level: the general preference for coalescence or that for epenthesis of the tongue dorsum gesture.

Further support for this idea comes from the English pronunciation of Japanese loanwords. It is reasonable to expect that in the English phonology, the phonological conversion involves a process of making a diphthong, or changing the original approximants [j, u] into the corresponding English vowels. This is an 'inverted' process of the Japanese.

Further examples:

(i) 

(ii) 

These pronunciations are commonly considered as phonetic deviations from adult pronunciation. Yet, it is more plausible to discuss them in terms of the timing control strategy unique to children. This might explain the more frequent occurrence of the forms in child's speech than adult's.

There are further examples.

(iii) 

We can see some connection between an alveolar tap [r] and the insertion of [j] or a voiced alveolar stop [d]. A common argument may appeal to invent the tap deletion rule, since it creates a sequence of a high vowel and a non-high vowel, the environment that motivates the glide formation. This line of explanation misses the underlying relationship to the phonetic form involving [d]. However, there is no need to go into details about such interesting alternations in the present discussion.

It can be argued, given the assumption of a special status of the three-vowel cluster, that a single mechanism of spatio-temporal modification of articulatory gestures better characterises the phenomena, rather than positing allophonic or low-level phonetic rules (see Browman & Goldstein, 1989, 1991, 1995).
pronunciation of loanwords, in which the three-vowel cluster is avoided by the glide formation. The Japanese words in (42) below contain either the approximant [j, u] in the intervocalic position or the palatalised consonant. All examples are taken from Wells (2000). (The pronunciation of British English is given unless there is a relevant difference in American English.)

(42a) Aiwa [a, i, ə] - [ˈaiwa], Kawasaki [ka, ə, səkəi] - [ka, wə], Kurosawa [ku, rəsəu] - [ku, əsəu], Okinawa [o, kənəu] - [əki, wə],

(42b) Kyoto [kə, ətə] - [ki, ətə], Kyushu [kə, ətə] - [kə, ətə], Nagoya ['nagọja] - ['nobə], Okayama [o, kə, jəma] - [o, kə, jəma], Ryukyu [rə, u, əkə] - [ri, u, əkə], Sanyo [sa, əjə] - [sa, əjə], Shoyu [ə, əjə] - [ə, əjə], Sukiyaki [su, i, əkə] - [su, əkə], Teriyaki [te, rə, i, əkə] - [te, rə, i, əkə], Tokyo [to, əkə] - [to, əkə], Toyota [to, ətə] - [to, ətə]

It appears that the approximants (including the [j]-like gesture in the palatalised consonants) in the original words are treated in two ways: the velar approximant [u] is retained as [w] or is replaced by [u]: the palatal approximant [j] is maintained or is replaced by [i] (or [i]). When the replacement is applied, it effectively creates a vowel sequence or a diphthong. Although the conversion process apparently involves phonotactic constraints in the English phonology, the words such as Kawasaki and Toyota clearly show the opposite process of the glide formation in the Japanese loanword phonology.

In the last two sections (2.5.2 and 2.5.3), the diphthong articulation has been characterised from the viewpoint of articulatory organisation. We have shown that the gesture-based model proposed in (29) makes it possible to approach the issue of cross-language variation in the diphthong production and gives a good account of the relationship between English and Japanese diphthongs, as well as the relationship between psychophysical parameters and cross-language variation. It is reasonable to conclude that: (i) English and Japanese diphthongs are similar to each other in that they show coalescence, or blending, between the V1 gesture and the V2 gesture (namely smoothing and monophthongisation); (ii) diphthongs in the two languages differ from each other in their temporal requirements particularly for the V2 gesture. This latter characteristic reflects a language-specific treatment of the three-vowel cluster and moraic status of the second element of a Japanese diphthong.

2.6 The Articulation of Two Moraic Consonants
2.6.1 Uvular Nasal and its Allophones
The underlying segment /n/ occurs only in syllable-final position. At the end of a word or an utterance the segment /n/ is realised as a uvular nasal. In contrast, it assimilates the place feature of the following segment in word-medial position or connected speech. The
systematic alternations are summarised in (43) by eight patterns in terms of the following segment type. Examples are given in (44).

(43a) /n/ → [n] /...______#
(43b) /n/ → [m] /...______[p, b, m]
(43c) /n/ → [n] /...______[t, d, dz, n, ts, r]
(43d) /n/ → [n] /...______[g, ŋ, dz, d]
(43e) /n/ → [ŋ] /...______[kg]
(43f) /n/ → [ʊ̃] or [z] /...______[s]
(43g) /n/ → [j] or [z] /...______[c, e]
(43h) /n/ → [v] /...______any vowel or [h, j, u]

(44a) /n/→[N]: ['kαN] ‘can (n.)’, ['mON] ‘gate’
(44b) /n/→[m]: [ka,mpaN] ‘dry biscuit’, [ka,mhaN] ‘signboard’, [ka,mmaN] ‘slow’
(44e) /n/→[N]: [ka,ŋk’ınN] ‘currency exchange’, [ka,ŋjinN] ‘poetry reading’
(44f) /n/→[ʊ] or [z]: [ka,ʊsaN–ka,₂san] ‘conversion’
(44g) /n/→[j] or [z]: [ka,j’in–ka,ʒ’inN] ‘admire’, [se,esajç] ‘product cost’

These allophonic representations direct our attention to the contextual change, or implementation, of the point of articulation that are estimated from rather deliberate speech. It must be noted that there is no audible friction in the allophones [z] in (43f and 44f) and [ʒ] in (43g and 44g): they simply illustrate the configurational characteristics of the tongue with the velopharyngeal port opening.

The segment /n/ and its allophones are considered moraic in that they have an independent phonological status at the mora level and are, in psychological terms, judged as having the same duration as in a single CV or V unit by a native-speaker of Japanese. This characteristic becomes less obvious in the alphabetical writing of Japanese words24. The representation, usually used in introductory textbooks, makes it difficult for foreign learners to produce a word with the proper segmentation. The spelling ‘kani’, for instance, can be pronounced either as [ka,ŋi] ‘crab’ or as [ka,ŋi (~ka,ti)] ‘simple’. The two words differ in the number of moras: two and three moras respectively. Similar examples are: ani [’aŋi](2) ‘elder brother’ and [’aŋi](3) ‘easy’; shinan [ci,naN](3) ‘extreme difficulty’ and [ci,naN (~ci,laN)](4) ‘new idea’; kanyuu [ka,nyuui](3) ‘to join’ and [ka,njouu (~ka,ŋjouu)](4) ‘to invite’; kinyuu [kî,nyuui](3) ‘to fill in’ and [kî,ŋjuu (~kî,ŋjouu)](4) ‘finance’; kinyoo [kî,ŋou](3) ‘ogre’ and [kî,ŋouo (kî,ŋjoo)](4) ‘Friday’ (the number of moras are bracketed). These pairs of pronunciations differ from each other in that the letter ‘n’ is either segmented as an onset consonant or as a coda consonant. In the latter case, namely the moraic ‘n’, the

24 I thank John Wells for indicating the interest of this point to me.
nasal resonance is distinctively longer than the non-moraic ‘n’. It is this durational characteristic that foreign learners of Japanese make a great effort to overcome.

Phonologically a single mora unit of the syllable dominates the uvular nasal in a coda position: the moraic property is assigned to the segment by phonological position within the syllable. The autosegmental framework provides a principled account of the derivation of the eight variants at the surface phonetic representation in terms of place-feature spreading from the following consonant to /N/ in the melodic tier. The diagram in (45) below represents such melody-internal spreading for the example of the word /saNpo/ ‘stroll’ (Itô & Mester, 1996).

![Diagram](image)

The first syllable ends in a uvular nasal, which is converted to a bilabial nasal by a spreading of the place feature [labial] of the /p/ in the following syllable. The nasalised vowel in (43h and 44h) is accounted for by a spreading of the tongue body feature of the following vowel. For all the allophones of /N/ to be generated by the phonological feature spreading above, it is necessary to assume timing controls specific to each case. In addition, while feature spreading implies a complete (or categorical) regressive assimilation of the constriction place between /N/ and the subsequent segment, the phonetic qualities of nasal(ised) allophones are impressionistically different from those of proper nasals and nasalised vowels (e.g. Kawakami, 1977). This leads us to the question of how the relevant articulators are coordinated in the production of the uvular nasal stop.

The uvular is the backmost point of articulation on the palate. For the production of [N], the main constriction is formed by lowering the velum and by pulling ‘the back of the tongue somewhat upwards and backwards’ (Akamatsu, 1997; 133) from the neutral position. These coordinative movements in the vocal tract suggest that the main articulators, namely the tongue and the velum, are both ‘active’ articulators. Thus, a complete closure that characterises a uvular nasal differs in its mode from other stop consonants such as [t], [d], [r], [k] and [g]. We would like to discuss the velum as an active articulator in the production of [N], relating this view to the variable realisations of the consonant in different phonetic contexts, and to the ‘moraic’ status of /N/ at the phonological level.

There is articulatory evidence to support the claim that the velum is active and
specialised for making a constriction with a certain point on the back of the tongue. Ushijima & Sawashima (1972), using the fiberscope method, observed the degree of the velar height during the production of /te,Nte,Nte,Nte,Nte,N/ (a /teN/ sequence). The velum activity in this utterance was compared with a sequence /nenenenene/ (a /ne/ sequence). It was found that in a /teN/ sequence a peak of velum elevation occurred for [t] and this is followed by a ‘marked descent’ for [n]. The uvular nasal in that particular context, as the authors acknowledge, is supposed to be realised as [n]. Nevertheless, velum lowering for the /N/ allophone is far greater than that for [n] in a /ne/ sequence. This suggests a stronger degree of acoustic coupling of the nasal cavity to the vocal tract in the production of the uvular nasal allophone(s).

Another finding is that velar lowering for [N] starts before the release of [t], and velar raising for the subsequent [t] already sets off at the beginning of the preceding [N]. This ascending and descending movement exhibits a symmetric pattern, suggesting that the speed of velar elevation is almost the same as that of velar lowering. There was a difference in anticipatory velar lowering between [n] and [N]. Comparing /seNen/ and /seenen/, the authors reported that the lowering movement was initiated at a point near the /s/-/e/ boundary in /seNen/; in contrast, in /seenen/, such a movement was not observed after the onset of the first vowel. These findings imply that raising and lowering of the velum are consciously controlled with reference to the type of nasal segments; and that different degrees of the velum height generate different modes of the coupling between the nasal and oral cavities. This results in distinctive auditory effects for [N] and its allophones.

In subsequent electromyographic (EMG) and X-ray microbeam studies (Ushijima & Hirose, 1974; Kiritani et al. 1980), the authors focused on the relationship between velar movement and its motor command. Ushijima & Hirose (1974) found that during the production of [N] the EMG activity of the levator veli palatini (LVP), which is considered the principal muscle of velopharyngeal closure, was reduced to the level of the resting state. Their data indicates that the descending slope of the EMG curve for [N] in /se,Nen/ is steeper and deeper than that for [n] in /te,enee/; and that the LVP activity remains suppressed after [n] but not after [N]. The acoustic duration of [N] in that context is clearly longer than that of [n]. These results indicate that the anticipatory effects are stronger in [N] than the carryover effects at the EMG level. This agrees with the results of Ushijima & Sawashima (1972) above.

Kiritani et al. (1980), comparing the syllable-initial [m] and syllable-final [N], reported that, although there is no obvious difference in the level of the EPG suppression (i.e. nearly complete for both segments), the LVP activity differed in the duration of the suppressed period (i.e. [N]>[m]). The overall results led both Ushijima & Hirose (1974) and Kiritani et al. (1980) to conclude that the greater velum lowering gesture for [N], that is
observed in Ushijima & Sawashima (1972), is partly due to mechano-inertial limitations on the articulators as a physical system (Ushijima & Hirose, 1974); and that the difference in velum height correlated with the duration of the suppression period of the LVP activity (Kiritani et al., 1980). In other words, the velum movement to produce the distinctive nasal resonance for [\text{n}] is derived partly from inevitable biomechanical effects and partly from the longer duration of the velopharyngeal port opening. These conclusions rely entirely on the fact that the suppression of the LVP activity is almost the same in degree but differs in duration: the suppression period is longer in the syllable-final nasal than that in the syllable-initial nasals.

However, the longer suppression duration of the LVP activity in [\text{n}] can also be accounted for by the fact that the segment is moraic at the phonological level. Given this, it is uncertain whether a simple linear correlation between the degree of velum height and the suppression period of the LVP activity can be assumed. Furthermore, the correlation does not explain the nature of the rapid (and earlier) descending and ascending EMG signal in [\text{n}], the contour pattern that corresponds to a suppression and reactivation cycle of the LVP action. There is another possibility, which the studies above fail to take into account, for attaining the greater degree of velum lowering during the production of [\text{n}].

Recall that a complete closure for [\text{n}] is formed by the two active gestures, namely lowering the velum and raising the tongue dorsum upwards and backwards. This allows us to make a speculation that ‘the significant lowering of the velum’ (Ushijima & Sawashima, 1972) is attained not only by the activity of the LVP muscle but also by that of the palatoglossus muscle. In his x-ray motion picture study Takada (1982) measured the various organs of the vocal tract during the production of /VN/, /VN/ and /VNCV/ sequences, where one of the five vowels were preceded and were in symmetrical combination, produced by one native-speaker of Japanese. There are two major findings relevant for our discussion. First, the x-ray tracings indicate that there is one general tendency for the movement of the tongue dorsum during the production of [\text{n}] in /VN/ and /VN/ sequences. In the context of /a, o/ the highest point of the tongue medio- and post-dorsum moves forwards and upwards, whereas in the context of /i, e/ the tongue point shifts backwards. These horizontal and vertical displacements were not found when the contextual vowel was /u/. That is, the position of the tongue body is consciously adjusted towards near-[ui] position. Takada (1982; 198) mentions that ‘such an adjustment facilitates an articulatory closure with the uvular point on the palate in /VN/ sequences (my translation)’, and continues that ‘in /VN/ sequences [where the /n/ is conventionally transcribed as a nasalised vowel] the positioning of the tongue body contributes to the typical resonance quality of a uvular allophone [\text{v}] which is clearly different from that of a proper nasalised vowel’.

The palatoglossus is the main antagonist muscle to the LVP in lowering the velum,
when the tongue is fixed\(^{25}\) (Hardcastle, 1976, 99). It may be assumed that the palatoglossus is involved in shifting the tongue body towards the near-[\textit{ui}] position. At the same time, the muscle in question may possibly contribute to lowering the velum. Given these assumptions, the palatoglossus activation draws up the tongue body closer to the uvula that is drawn down by the suppression of the LVP and possibly by the concurrent activity of the palatoglossus\(^{26}\).

The acoustic duration of [\textit{n}], given the legitimate articulatory-acoustic relation, may support our hypothesis that the longer suppression period of the levator palatini has a link to the moraic status of [\textit{n}] at the phonological level. The data reported by Takada (1982) indicates that both the vowel and the nasal duration are almost equal in the sequences of /\textit{VN}/ and /\textit{VNV}/. In contrast, the duration of [\textit{n}] in /\textit{VNCV}/ sequences becomes longer when the following consonant is voiced, but becomes shorter when the following consonant is voiceless. Sato (1993) measured syllable-final nasals in Japanese, English, and Korean, and reported that, although durational differences in syllable-final nasals are due to the voicing feature of the following consonant, nevertheless, the duration of syllable-final nasals is substantially longer than that of syllable- initial nasals: mean durations of syllable-initial and syllable-final nasals are 42.58msec and 101.72msec (ratio 1:2.39) for 8 Japanese speakers; 68.05msec and 92.34msec (ratio 1:1.36) for 8 (American) English speakers; and 56.15msec and 100.68msec (ratio 1:1.79) for 8 Korean speakers\(^{27}\). Overall, the velar gesture for [\textit{n}] and its allophones appears to be sustained longer in duration, even though the voicing feature of the subsequent consonant influences timing of velar closing.

In summary, we have discussed articulatory controls of [\textit{n}] production with the assumption that the segment is a stop articulation. Two active articulators, the tongue dorsum and the velum, form an occlusion. It seems reasonable to suppose that the distinctive auditory quality of [\textit{n}] and its contextual variants are derived by specific patterns and timing of oral-nasal coarticulation. The longer acoustic duration may correspond to the motor command of the LVP suppression and the palatoglossus activation. These activities at the EMG signal level have a potential relationship to the linguistic concept of mora at the abstract phonological level.

\(^{25}\) Conversely the palatoglossus, together with other lingual muscles, assists in raising the back part of the tongue for velar articulation, when the velum is fixed (Hardcastle, 1976, 124).

\(^{26}\) It is, in principle, possible, as Hardcastle (1976) mentions, that the palatoglossus muscle is active in nasal articulation. The EMG activity of the muscle, however, has not always been observed: Fritzell (1969) and Lubker \textit{et al.} (1970) (both cited in Bell-Berti (1973)) found the occurrence of the activity, whereas Bell-Berti (1973) failed to find such evidence.

During the vowel production the velum is elevated higher in /\textit{i}, \textit{u}/ than in /\textit{e}, \textit{o}/; /\textit{a}/ shows the lowest position (Ushijima & Sawashima, 1972). This is explained by a mechanical linkage between the levator and palatoglossus muscles (Hardcastle, 1976; 99). However, the positioning of the tongue body for [\textit{n}] around the vowel [\textit{ui}] may not be equated to the gesture properly targeted to the high back vowel [\textit{u}].

\(^{27}\) It has been recognised that the vowels have a similar tendency.
2.6.2 The Articulation of Moraic Obstruents

There is a phonological contrast between a singleton and a geminate consonant\(^{28}\) in Japanese. A geminate is analysed as a sequence of two identical consonantal segments, the first element of which constitutes the final consonant of the preceding syllable; and the second element of which is composed of the initial consonant of the following syllable. Thus, the geminates create a closed syllable in a similar way to /N/. We find the following minimal pairs (46a-g).

- (46a) [su'pai] ‘spy’ [sup'pai] ‘sour’
- (46b) [ka,too] ‘fructose’ [kat,too] ‘frustration’
- (46c) [ka,ko] ‘manufacturing’ [ka,koo] ‘figure’
- (46d) [ka,so] ‘imagination’ [kas,so] ‘gliding’
- (46e) [ka,co] ‘singing’ [kaç,ço] ‘flying’
- (46f) [it.tsuu] ‘stomach ache’ [it.tsuu] ‘one copy (of a document) adj.’
- (46g) [ka,tç'u] ‘into the flame’ [kat,č'u] ‘armour’

Geminate consonants in Japanese are basically limited to voiceless obstruents\(^{29}\), and are sequences of ‘homorganic closed transition’ (Catford, 1977; 220) without any modification of phonation and stricture types. Among voiceless obstruents the gemination of [ϕ], [ç], and [h] is not common in native vocabulary but is found in onomatopoeia and foreign words: [ϕuϕuϕu] ‘chuckle’, [çççç] ‘grin’, and [ahhahha] ‘laugh loudly’ are onomatopoeic expressions for describing the manner of laughing; [su'ţaţuţu] ‘staff’, [e,ɛɛ'ɛɛ] ‘P. Ehrlich’, and ['bahh] ‘J.S. Bach’ are loanwords.

It has been common practice to describe such characteristic properties by using the symbol /Q/ for the first element of a geminate consonant. This approach, however, introduces unnecessary complications to the analysis. The ‘phoneme’ /Q/ is an archiphoneme that is specified only by two features, [consonantal] and [voiceless], and possesses moraic status of its own. Since the segment lacks the place feature, it is regressively assimilated from the following consonant by rule. There are reasons for rejecting such a ghost consonant. The phonetic contents are entirely predictable from the following consonant. Moraic status in Japanese is prosodically determined by unifying a particular consonant with a vowel and also by specifying a particular location of a segment within the syllable. It seems more

\(^{28}\) Catford (1977; 210) proposes a distinction between a long consonant and a geminate sequence of two segments. From the articulatory stand, a long consonant involves a prolongation of the articulatory posture. A geminate consonant, in contrast, involves the bi-segmental nature of the sequence by the presence of re-establishment of initiator power within the continuity of articulation. From the phonological stand, a long consonant is heterosyllabic or heteromorphemic, while a geminate is tautosyllabic or tautomorphemic. By the term ‘geminate’ we mean a phonetically long consonant, regardless of the presence or absence of a linguistic boundary.

\(^{29}\) In emphatic pronunciation of some words, a voiced obstruent is geminated: [su'barâ'tî] ‘wonderful’—[stûb,bara'cîl]: [su'i'gôi] ‘surprising’—[stûg'gôi]. Also, a voiced geminate can be found in loanword pronunciation such as ['beddo] ‘bed’ and ['baggu] ‘bag’ (yet the voicing in those words tends to be lost, producing a voiceless geminate in actual pronunciations).
parsimonious that the moraic property is considered exclusively as prosodic, rather than as exceptionally inherent in the segment. Thus, we describe an underlying geminate consonant by doubling a singleton consonant.

Phonologically the prosodic association commonly accounts for the singleton-geminate contrast. Whereas the singleton consonant former simply occupies the onset position, the geminate is doubly linked to the mora unit and to the onset position. The underlying bi-segmental geminate is converted into the surface mono-segmental geminate by root fusion (Itô & Mester, 1996). The representation of /gak + koo/ ‘school’ is shown below.

In particular, differences in closure (or narrowing) duration between single consonants and their geminate counterparts may become significantly larger. It is interesting, in the first place, to ask how native speakers distinguish a geminate from a singleton consonant; more specifically whether the phonetic cue for the distinction is the duration of consonantal closure or narrowing. In the second place, complementary to the first question, we may ask how a geminate consonant is coordinated with the surrounding segments. These questions help us to approach the nature of language-specific timing control, namely mora-timing.

The geminate consonants in Japanese have an acoustic duration three times as long as their single counterparts (ratio 2.6–3.0:1.0; Han, 1962): the timing control of geminate consonants is not a simple doubling of the given singleton segment. Fujisaki et al. (1975) artificially lengthened the duration of [t] and [s] in the words [ita] and [isse] to localise the perceptual boundary between singleton and geminate consonants.

Varying the duration of the complete closure for [t] and that of friction noise for [s], they determined that the perceptual boundaries were 169 ms for [ita] ‘existed’ and [itta] ‘went’; and 166 ms for [isse] ‘place name’ and [isse] ‘a unit of area’. Han (1992) measured closure durations for short and long voiceless stops spoken by American-English speakers who are ‘fluent’ in Japanese. She

---

30) Fujisaki et al. (1975) examined the single-geminate (or short-long) contrast in vowels and NC sequences (namely, /n/ + C), as well as geminate obstruents. As the duration of steady-state of [o] in [o,i] and nasal resonance of [m] in [a,ma] changed, the perceptual boundaries were settled on 156 ms for [o,i] ‘nephew’ and [o,oi] ‘cover’; and 141 ms for [a,ma] ‘nun’ and [a,mma] ‘massage’. These results suggest that the duration is a distinctive factor in segmental and prosodic information in spoken Japanese.
found that fluent Americans produce the short-long contrast in Japanese by simply doubling the closure duration of a short stop. The timing control is transferred from that of English geminates such as in the phrases *get Tom* and *top part*, in which there is a grammatical boundary. Han (1992) suspects that this timing transfer, or disparity in ratios for the short-long contrast, may be related to the presence or absence of a foreign accent in non-native speaker’s speech.

It thus seems clear that the geminates, when considered as a segmental phenomenon, are characterised by distinctive duration, relative to their non-geminate counterparts. This durational characteristic, or timing strategy, may be language-specifically determined. Lahiri & Hankamer (1988) measured the acoustic closure duration of long and short /p, t, k/ in Turkish and Bengali. The data indicates that, while the geminates are substantially longer than the non-geminate counterparts, the ratios between the two consonant types appear to be language-specific. The mean closure durations for short and long consonants in Turkish are 60ms and 176ms (ratio, 1.0: 2.93) but those in Bengali are 129.8ms and 251.4ms (ratio, 1.0: 1.94). Geminates stops in Turkish appear to show a trend similar to those in Japanese. In a subsequent perceptual study, Hankamer *et al.* (1989) reported that, in spite of such an obvious cross-language difference in durational ratios, the single overriding cue distinguishing geminate from non-geminate counterparts is the duration of the complete closure.

As a temporal phenomenon, the second point to be discussed, the Japanese geminate stops reveal unusual adjustment behaviour with the surrounding vowels. Maddieson (1985) reviews acoustic studies of vowel duration in CV and CVC syllables and that in CV.CV (i.e. singleton C) and CVC.CV (i.e. geminate C) syllables. It was found that the vowel duration tends to be shorter in closed syllables and is much shorter in syllables with geminate consonants. This effect is called ‘closed syllable vowel shortening (CSVS)’. In contrast, Japanese, as Maddieson acknowledges, is one of apparent counterexamples. There is little difference between the duration of a vowel before a single and that of a geminate consonant (Homma, 1981: Smith, 1992). Actually, a similar shortening of \( V_1 \) duration does occur in \( C_1V_1C_2V_2 \) sequences in Japanese. However, the directionality of the shortening effects is different. Homma (1981) found that, while the voicing feature of both \( C_1 \) and \( C_2 \) affected the duration of \( V_1 \), the preceding consonant (i.e. \( C_1 \)) has a stronger effect on \( V_1 \) duration.

To summarise, we have observed the segmental and temporal aspects of geminate consonants. It was found that the durational property is crucial in both production and perception of geminate consonants; and that the phonological representation in (47) reasonably reflects the phonological length of a geminate. These points are also true for the characterisation of \( /N/+C \) sequences that we discussed in the last section. However, the fact that the temporal adjustment, CSVS, is less explicit in Japanese suggests the necessity of a language-specific strategy for temporal adjustment. It is worth noting that in the idea of CSVS,
the 'syllable duration' is a given unit: it is not a total sum of the durations of independently determined segments. Because CSVS deals with the durational relationship between the segments within the syllable frame, the absence of CSVS can be interpreted as the interference, or suppression, of syllable-level timing control by mora-level timing control (Kubozono, 1992, 1999).

At the beginning of this chapter (section 2.1 (1)), we assumed, following the prosodic constituency proposed by traditional prosodic phonology, that the mora was a unit of weight and position within the syllable. We have shown that this view of the mora helps to explain some prosodic characteristics of the diphthong articulation, uvular nasal allophones, and geminate obstruents. Yet, such analytic profits do not entirely support the necessity of the language-specific prosodic unit for describing the Japanese sound system. There remains a question concerning the frequently encountered observation that Japanese is a mora-timed language, in which the mora is a basic unit of temporal regulation.

2.7 Aspects of a Mora-timed Rhythm

'Rhythm, in speech as in other human activities, arises out of the periodic recurrence of some sort of movement, producing an expectation that the regularity of succession will continue' (Abercrombie, 1967, 96). There is a natural link between the perception and production of connected speech. This raises the question of which properties in spoken language are perceived as 'rhythmic'; and how they are specified in the spatiotemporal domain. This section examines the rhythm of spoken Japanese, namely mora-timing, in terms of hypothesised rhythm categories, perceptual isochrony and co-ordination of articulatory movements. Although evidence for classifying Japanese as a mora-timed language comes from various linguistic domains, we discuss the mora as a temporal structure in speech production. We shall consider what evidence exists for describing the rhythm of Japanese in terms of moras.

2.7.1 The Mora as a Unit of Duration and Temporal Compensation

There is a widely accepted assumption that a language can be categorised as having either a syllable-timed rhythm or a stress-timed rhythm (e.g. Abercrombie, 1967; Dauer, 1983). It is assumed that the syllables recur at regular intervals in syllable-timed languages and that the stresses occur at roughly equal intervals in stress-timed languages. Spoken Japanese is characterised as having a mora-timed rhythm that can be regarded as a subcategory of a syllable-timed rhythm (e.g. McCawley, 1968; Kubozono, 1999). A single mora unit is composed of CV, CyV, V, syllable-final nasals and voiceless obstruents (the first element of a

---

31 Kubozono (1999) lists various phonological and morphological phenomena such as speech errors and word-formation, that can be explained adequately if the mora is assumed.
geminate consonant): /ta/, /tja/, and /a/ are counted as one mora; /ta.i/ and /ta.N/ as two moras, and /ta.t.ta/ as three moras. Even though mora units have various types of segmental combinations, they are all considered approximately equal in length.

Recall from the previous sections that: the second element of a diphthong undershot spatially but not temporally; the consonant /n/, its allophones and the geminate consonants are significantly longer than the other plain consonants; and that those moraic consonants appear to be one main cause of the so-called foreign accent. In addition, improper timing can easily be observed in the production of a long vowel in: for instance, [bIO,oiN] 'hospital' may sound like *[biɔiN,biɔiN]. These examples reveal that the duration plays an important role in spoken Japanese. This characteristic may become clear if we attempt to describe mora with reference to sonority. The contours (48) are drawn based on a sonority hierarchy commonly assumed (e.g. Ladefoged, 1993).

\[
\begin{array}{c}
\text{Open-mid vowels} \\
\text{Close vowels} \\
\text{Approximants} \\
\text{Nasals} \\
\text{Fricatives} \\
\text{Stops}
\end{array}
\]

It is clear that a sonority peak defines a syllable but not a mora. The sonority contour in (48b) and (48a) exhibits two peaks, suggesting that the two words consist of the same number of syllables. However, the word in (48b) is identified as having four units: [ro.n.do.N]. Counting peaks of sonority does not uncover number of the moras in a word or an utterance. Thus, the mora is considered an abstract unit of duration, defined in terms of perceived length, and the unit is psychologically real to a native speaker of Japanese (Bloch, 1950; Han, 1994; Hockett, 1955; McCawley, 1968).

Isochronous rhythm perceived by native speakers of Japanese must have constant acoustic manifestations. This idea has been examined by measuring the duration of various mora units. In a pioneering experimental investigation, Han (1962) examined a durational characteristic in a short-long (or single-geminate) distinction for vowels and consonants. She reported that the phonemic duration of a long segment was not simple doubling of a short counterpart: the average ratio was 1.0 to 2.6 and often 1.0 to 3.0 for both vowels and consonants. This allows Han to reject the idea that the length of a segmental phoneme is the basic unit of duration. Instead, we can give a natural account for a durational characteristic, if a unit of duration associated with a set of segments is assumed. Han (1962; 73) proposed a strong version of the mora hypothesis: that the actual length of each mora was approximately the same; and that there was a strong tendency for the components of mora to balance each
other within a syllable, in order to obtain equal duration with neighbouring units.

How does a speaker systematise the segment elasticity to equalise the duration of a mora unit? This question logically arises because the inherent segmental durations of vowels and consonants vary considerably in the data adduced by Han (1962). If the test word consists of either relatively short segments such as [u] and [r], or relatively long segments such as [a] and [s], then a speaker must control the segmental disparity in duration in order to generate a sequence of isochronous moras. Port et al. (1980) examined two-mora utterances such as [basa], [busu], [bara] and [burui]. The vowel durations were found to be longer when a sequence contained the inherently shorter consonant in the intervocalic position (i.e. [burui]>[busu]); and the inherently longer consonant became shorter when a sequence contained the inherently longer vowels (i.e. [busu]>[basa]). In spite of such inherent and contextual variability of each phonetic segment, what was found to be ‘isochronous’ is the overall duration of the test words. This demonstrates that temporal compensation functions at the word level to maintain the overall word length (i.e. two-mora in this case), as well as at the syllable (or mora) level that was reported by Han (1962). Homma (1981) also reports similar results for the word level durational adjustment.

The presence of the word-level temporal compensation can be considered as one characteristic of a mora-timed rhythm. Comparing the temporal compensation in English with that in Japanese, Port et al. (1980; 246) mention these three differences: (i) the tendency for a constant sum (i.e. the temporal adjustment at the syllable level) in English is primarily restricted to stressed syllables, while in Japanese the constant sum of segment duration continues throughout all syllables; (ii) the temporal compensation in English is applied only to segments in the nucleus and coda of a stressed syllable, while that in Japanese is bi-directional; and (iii) the durational adjustment in English is related to a segmental contrast between voiced and voiceless obstruents, while that in Japanese has no direct contrastive function, but only works to maintain a prosodic pattern when the number of moras in each word is different. In addition, it appears that the stress pattern in English, rather than the duration, is an important cue for speakers/listeners to recognise the syllables of a primary (sentence) stress; and that duration, in contrast, is more deliberately used as a pragmatic cue to turn-taking in conversation and as a phonetic cue to the syntactic structure (e.g. Cutler & Isard, 1980). Rather than using a conventional unit of isochrony, the above contrasting properties of the two languages might better be characterised with reference to the overall rhythmic structure: a rhythm of alternation for English and that of succession for Japanese (Allen, 1975); a duration-compensating language for English and a duration-controlling language for Japanese (Hoequist, 1983).

The question now becomes not whether the mora is a constant durational unit in speech, but how a durational constancy is maintained in the proposed rhythm type. Port et al.
(1987) further examined the mechanisms of the temporal compensation. They found that increasing the number of moras adds nearly the same amount to the total word duration. This partly supports a strong version of the mora hypothesis, although the durational increase, as Port et al. acknowledge, may or may not be adjusted with respect to the segment intervals. Such a proportional increase of the total word duration was further scrutinised by changing speaking tempo and various segments. It was found that these two variables had little influence on the proportional increase of the word duration. The results therefore disconfirm the simple notion that each mora becomes the same duration as every other mora, but confirm a general notion that mora timing in Japanese is extremely regular. By this it is meant that a constancy of duration is maintained with reference to the number of moras in a word, not to the components of a mora unit. Hence, mora or mora timing is viewed as one implementation system that generates the rhythm type, rather than a ‘visible’ isochronous unit that composes the sprung rhythm itself.

An attack on the above view of the mora comes from Beckman (1982). Support was not found either for the predictions entailed by a strong version of the mora hypothesis or for the effect of temporal compensation. Beckman’s data indicates that there is no substantial divergence in duration between a CV syllable with a devoiced vowel (i.e. a fricative nucleus) and one with a voiced vowel; the duration of geminate consonants does not reflect their moraic structure; and that the durations of mora units are not approximately the same and there is no general tendency of temporal compensation to equalise the unit duration with neighbouring segments. Based on these results, Beckman (1982) concluded that the ‘staccato rhythm’ of Japanese was largely due to knowledge of the moraic kana writing system. Although the results presented by Beckman (1982) are contrary to those of Han (1962) and Port et al. (1980, 1987), her conclusion is actually unsupported. It does not make sense to assume that reading knowledge is prerequisite for the acquisition and production of the ‘staccato’ rhythm in Japanese: a child can speak with that type of rhythm before learning kana letters.

An intriguing consideration is the relation between acoustic duration and articulatory duration in the analysis of the mora hypothesis. The above studies explicitly demonstrate that the durations of the acoustic segments vary considerably with phonetic contexts. This leads Port et al. (1987, 1584) to a proposal that the mora exists only in larger units like words, rather than a single CV unit. At the same time, Port et al. (ibid.) suggest that the mora ‘exists at an underlying articulatory level that is not directly revealed in the “acoustic edges” we measure on spectrogram’. However, the treatment of moraic consonants raises the issue, in articulatory analysis of kinematic correlates, of rhythm. Vatikiotis-Bateson & Kelso (1993) obtained the movement data of the lower lip and jaw during the production of reiterant
speech\(^2\) by Japanese, French and English speakers. It was found that the three languages are qualitatively the same, but they differ in the general patterning of the stiffness parameter: the covariance between peak velocity and displacement characterises stress in English and French but characterises accent and mora in Japanese. The peak-to-peak movement in Japanese was found to be cyclic but was asymmetrical in that opening gestures tend to be longer than closing gestures. Exceptions to continuous cyclic opening-closing gestures are moraic consonants: for instance, /tt/ in motte, /nw/ in obaasan wa and /nt/ in sentaku. In the corpus of Vatikiotis-Bateson & Kelso (1993) the closing gesture by the lips was maintained so that a break in the continuous opening-closing alternations appeared. It is, however, unsatisfactory to interpret the behaviour of moraic consonants as exceptional. This is because, as we mentioned before, it is the existence of moraic consonants that phonologically characterises Japanese as a mora-timed language, not the relatively simple syllable structure, i.e. a consonant plus a vowel.

There are preliminary results to show the articulatory regularities of the production of moraic consonants. Figure 2-4 shows the typical spectrograms of the nonsense words /banta/, /banda/ and /banra/ produced by speaker MN. Figure 2-5 presents the EPG patterns of those three words. The test words were embedded in the frame sentence and repeated six times at normal speaking rate.

It can be seen from the spectrograms in Figure 2-4 that the three words do not differ substantially in the overall duration (indicated by the bracket (a)): it was measured from the implosion of the word-initial [b] to the implosion of the next [b] in the carrier sentence: although the sequence /banta/ is the longest of the three, the sequences /banda/ and /banra/ are nearly the same. In contrast, the significant difference can be observed for the duration of the nasal element (indicated by the bracket (b)). The nasal duration increases in the order [t]<[d]<[r]. This is a function of oral-nasal coarticulation. These results generally agree with previous studies such as Port et al. (1987) and Sato (1993). It is confirmed that the duration of the moraic nasal varies with the type of the following consonant. In contrast, the EPG patterns in Figure 2-5 exhibit fairly similar durations of articulatory closure for the three target sequences. It is supposed that the duration of articulatory closure is shorter in /(b)ata/, /(b)ada/ and /(b)ara/, the sequences without the moraic nasal /n/ (see chapter 4, section 4.3.4): all the durations are less than 100.00msec. Note in passing that the alveolar tap evidently shows complete occlusion at the (dent)alveolar region.

These preliminary results have two implications. First, acoustic irregularities do not always correspond to articulatory irregularities. What is found to be systematic is the duration of articulatory closure for the sequence /-NC-/. Second, this durational constancy can be

\(^2\)Every syllable or mora of a normal sentence is replaced by /ba/ or /ma/.
FIGURE 2-4: SPECTROGRAMS OF /banTa/, /banda/ AND /banra/ PRODUCED BY SPEAKER MN.
The three nonsense words were spoken with the frame sentence 'moo ... bakarida'.
The sampling rate is 10kHz. The bracket (a) approximates the measured region for total duration of a
word; and the bracket (b) estimates the measured region of the duration of the nasal element.

(a) /banTa/: total duration, 437.00 msec; nasal duration, 96.10 msec.
(b) /banda/: total duration, 408.70 msec; nasal duration, 102.30 msec.
(c) /banra/: total duration, 407.80 msec; nasal duration, 119.90 msec.
Figure 2-5: EPG Contact Patterns of /ba, Nta/, /banda/ and /banra/ produced by speaker MN. The durations of complete occlusion are: 141.76 msec. for /banra/; 118.77 msec. for /banda/; and 126.43 msec. for /banra/.
accounted for by the concept of ‘articulatory syllable’ proposed by Stetson (1951: 33): ‘the consonants...have a function in releasing and arresting the chest pulse which constitutes the main movement of the syllable’. Although Stetson’s chest-pulse theory was disproved by Ladefoged et al. (1958), this does not negate the characterisation of the syllable as a movement. Given that the ‘arresting’ and ‘releasing’ movements together form a syllable, the expected differences between the sequences with /n/ and those without it can be considered as a lengthening, or shortening, of the arresting movement. The characteristic lengthening in the /-NC-/ sequences could be quantified as the relative duration of the corresponding single consonant.

The mora hypothesis that each mora unit takes exactly the same duration, proposed by Han (1962), is easily disproved by measurement as the elasticity of phonetic segments differs from each other. A weak version of the mora hypothesis, that mora timing is regulated towards isochrony, is too weak to explain the results of Port et al. (1980, 1987) and Homma (1981). An alternative, sketched by Port et al. (1980), is that there is a set of linguistic rules that assign time schedules for balancing the segment elasticity, the directionality and the domain of temporal compensation. Although speech timing appears to be controlled at multiple levels, such rules for timing checking are not plausible, since the control variables are so many that the attempts to specify all of them cannot yield an appropriate model (Port et al., 1987). As mentioned at the beginning of this section, if the essence of rhythm lies in ‘the periodic recurrence of some sort of movement’ and the anticipation of its ‘regular succession’ (Abercrombie, 1967), it is reasonable to suppose that the timing of motor events in the vocal tract is systematic and its control is associated with the relevant areas of the brain. Fry (1964; 218) mentions:

It is undoubtedly this close link between motor organisation and syllabification that accounts for the major role played in the perception of syllables by a listener’s own speech habits...If he is listening to a foreign language, his count [of the number of syllables] may differ materially from that made by native speakers; in English and Japanese, for example, speech habits are so different in all matters connected with rhythm that there are wide discrepancies in syllable judgments by speakers of the two languages. It is hard for a Japanese listener to hear the English word *sticks* as having less than four syllables, and for an English listener hear more than two in the Japanese word *sinbun*. The perception of syllable structure, like that of stress which is so closely allied with it, is based mainly on kinaesthetic memory, on a store of patterns derived principally from proprioceptive information, and it is clear that there must be a level of brain activity where these patterns are held in some well-defined form and where they are readily accessible for the control of the manifold co-ordinated muscle movements of speech.

Currently, we can only assume that there are brain mechanisms responsible for speech timing. This is implicitly or explicitly assumed in the previous studies reviewed above. They attempted to use durational measurements to specify the prosodic unit, mora, in which
both vowels and consonants were taken into account. The temporal adjustment is not a simple
phenomenon where vowels and consonants come together, but is a logical reflection of
coordinative actions between the consonantal and vocalic gestures in the vocal tract. This
implies that potential differences in the durational control can be sought in the systematic
articulatory control underlying them. It is therefore necessary to ask how articulatory gestures
for vowels and consonants are coordinated; and whether there are any language-specific
tendencies for coordinative movements of the articulators with reference to the given category
of rhythm.

2.7.2 The Mora as a Pattern of Articulatory Coordination

The distinction between vowels and consonants can be made in terms of their general
movement characteristics. In his comparative study of the VCV articulations, Öhman (1965)
suggests that trans-consonantal vowels are produced as a smooth diphthongal gesture with an
independent consonantal gesture superimposed. Fowler (1983) also discusses nonlinearity in
the articulatory system and suggests that vowels are produced in a (nearly) continuous cycle
and exhibit relatively slow changes in the global shape of the vocal tract: this implies that
consonants, in contrast, are produced as relatively fast and isolated deformations of the vocal
tract. This is compatible with the description by Sproat & Fujimura (1993): ‘consonantal
gestures are those that produce an extreme obstruction in the mid-sagittal plane, while vocalic
gestures are those that do not produce an extreme obstruction’. In more metaphorical terms,
Browman (1991) mentions that ‘articulatorily, vowels act as a “ground” to consonantal
“figures”’. Thus, the vowel-consonant contrast can be characterised parametrically by
different dynamic properties effected by the articulators in the vocal tract.

A systematic treatment for the dynamic differences between vowels and consonants
has been set forth presenting the approach called Articulatory Phonology (Browman &
set of vocal tract variables have two aspects. On the one hand, gestures are events in the vocal
tract referring to the formation and release of the given articulator(s); and on the other hand,
they are abstract units of action describing phonological entities. There is a broad distinction
for the oral constriction gestures, vocalic and consonantal. Both gestures use the same sets of
articulators but differ in constriction degree and in time constant required: a consonantal
gesture has a higher degree of constriction and a shorter time constant. Because a gesture is a
unit that is assumed to have its own intrinsic temporal extent, the timing relations between
vowels and consonants are considered as natural consequences of coordinative patterns
between the gestures, rather than as outputs of the rules describing time factors for the
individual segments or a set of segments.

Viewed in this way, the conventional rhythm categories can be treated as
language-specific patterns of temporal interaction between vocalic and consonantal gestures. Assuming the independency of the vowel producing mechanism and the consonant producing mechanism, representing them as two separate tiers, timing relations between vocalic and consonantal gestures are modelled below (Browman & Goldstein, 1990; Browman, 1991).

(49a) V ——— V (49b) V ——— V (49c) V ——— V
    C                   C                   C

Both interactions in (49a) and (49b) exemplify V-V timing, in which two vowels are sequentially and directly timed and the consonant can be associated with either of the vowels. The situation in (49c), in contrast, illustrates V-C timing, in which two vowels are indirectly timed only through the intervening consonant. Underlying the two models is the idea that languages are expected to be different in whether vowels are indirectly or directly timed with each other; and that different patterns of articulatory coordination would have different consequences for the acoustic duration. It is supposed that languages with V-V timing would include languages that have been characterised as syllable-timed; and that languages with V-C timing should include mora-timed languages (Browman, 1991; Smith, 1992). We shall discuss the results reported in previous articulatory studies in the light of the two models of vowel-consonant timing and shall examine the assumption that Japanese is a V-C timing language.

An expert phonetician notices language-specific characteristics of articulatory timing in an explorative way. In his observations of the consonants [p, t, k, b, d, g, m, n, r, s, z, ç, h, ts, dz, tç, dz, tq] (my transcriptions) in Japanese, Daniels (1958; 58) mentions:

These consonants are said to be ‘prefixed to’ the vowels rather than to ‘precede’ them because the relation between the consonant and the vowel is generally closer than that between the two elements of a consonant-vowel combination in English...[I]n Japanese, it is necessary to put the speech organs into the position for the vowel...before producing the consonant...There is therefore more or less difference in almost all cases between the ways in which ‘the same’ consonant is produced when prefixed to different vowels. Sometimes the influence of the vowel is so disturbing that it seems necessary, or at least convenient, to regard the consonant as no longer ‘the same’, but as a variant of one of the forms which is, to some extent arbitrarily, taken as the basic one.

This impressionistic observation involves a vague distinction between the phonetic representation and the articulatory events in the vocal tract, certainly having an aim to explain allophonic variation before the high vowels (see section 2.3.1). Yet, the fact that the consonants in Japanese are more ‘prefixed to’ the vowels than those in English can be interpreted as indicating the status of the intervocalic consonant different between V-C and V-V timing.

Kiritani & Sawashima (1987) observed the production of VCV and CVC sequences
in Japanese and American English, using the X-ray microbeam technique. For Japanese variations in vowel articulation were analysed in the /C₁VC₂i/ sequences, while those in consonant articulation were examined in the /mV₁CV₂/ sequences: both sequences consisted of the five vowels and one of the four consonants /t, k, m, s/. The results indicate that: (i) in the /C₁VC₂i/ sequences the effects of C₁ upon the vowel are greater than those of C₂; and that (ii) in the /mV₁CV₂/ sequences the effects of V₂ upon the consonant are greater than those of V₁. For American English the effects of the preceding and following vowels were analysed for /s/ in the words such as seat, set, peace, and bus, that were embedded in a carrier sentence ‘It’s a ___ again’. It was found that: (i) the influence of V₁ was not significant for /s/ either in the syllable initial or final position; and that (ii) the effects of V₂ were significant for the syllable final /s/. For the sequences containing the syllable initial /s/, the tongue movement towards the following vowel started after the release of constriction. These results led Kiritani & Sawashima (1987; 148) to the conclusions that ‘in Japanese a consonant and the following vowel form a cohesive unit of articulatory movement’; and that in American English ‘the syllable initial consonant retains its identity more than in Japanese with regard to the influence of the following vowel’.

The findings of Kiritani & Sawashima (1987) have important implications for the proposed models of timing control, namely V-V timing and V-C timing. Their results for American English suggest that different phasing of the consonantal gesture depends on its position within the syllable, namely initial or final. In contrast, the intervening consonant in Japanese tends to be phased with the following vowel. These situations are summarised below.

(50a) V₁-- V₂  (50b) V₁-- V₂  (50c) V₁ [ V₂

These diagrams indeed reflect the results of Kiritani & Sawashima (1987) and help us to develop the idea of the two distinct types of gestural overlap. However, two points must be noted. First, different timing of consonantal gestures in (50a,b) can be explained by the effects of another prosodic category, namely foot, which is delimited from a stressed syllable to a syllable just before the next stressed syllable (e.g. Abercrombie, 1964). The syllable final /s/ in the utterance ‘It’s a bus again’ does correspond to the foot final position, and the syllable initial /s/ in ‘It’s a seat again’ corresponds to the foot initial position. Thus, we can reasonably say that the consonantal articulation is stronger, or more resistant to coarticular effects, in the foot initial position than in the foot final position. This suggests that there are close relationships between articulatory strength (or the degree of coarticulatory resistance) and the
prosodic domain. Second, the greater anticipatory effects for the following vowel in Japanese do not imply that there are no carryover effects from the preceding vowel. In other words, it is not necessary to suppose that V-V timing and V-C timing are mutually exclusive: the question is a matter of degree, or how consistently a language shows a tendency towards a particular timing type.

Magen (1984), in her pilot study, found that both English and Japanese do exhibit vowel-to-vowel coarticulatory effects, but the magnitude of coarticulation is greater in Japanese. The comparison was made for the VCV sequences with the vowels /i/, /a/ and one of the labial consonants /p, b, m/. The analyses of formant trajectories indicated that in English the V-to-V effects were blocked for stressed vowels; and that the effects of trans-consonantal vowels were clearly observed for the duration of an unstressed vowel, but not for that of a stressed vowel. The V-to-V effects in Japanese were found throughout the duration of the vowel, whether it was accented or not. Thus, it might be more precise to say that V-V timing and V-C timing are different from each other in terms of how the consonantal gesture is adjusted to the presumably continuous diphthongal gesture.

The model of V-V timing and that of V-C timing make different predictions for the timing of the vocalic gestures in the context where the intervening consonant is a geminate. On the one hand, the length of an intervening consonant would be expected to cause few effects on the timing of V-to-V gestures in a language with V-V timing. This is because the two vocalic gestures are directly phased with each other regardless of the consonantal gesture(s). On the other hand, the consonant length might well be expected to influence the timing of V-to-V gestures in a language with V-C timing. This is because the vocalic gestures are timed with respect to the consonantal gesture(s) in the medial position. These predictions were investigated by Smith (1992, 1995).

Using the X-ray microbeam technique Smith (1992, 1995) examined the movement of the tongue body and the lips for /mViCV/ and /mVCCV/ sequences where the vowels were asymmetrical combinations of /i, a/, and the consonant was one of /p, t, m, n/ and the homorganic geminate. These words were spoken by three speakers each of Japanese (i.e. V-C timing) and Italian (i.e. V-V timing). The results indicated that the contrast between single and geminate consonants showed up in a quite different way. For Japanese the articulatory plateau for V₁ was longer in utterances with a geminate consonant than in those with a singleton consonant: timing of the movement towards the target position of V₂ was effectively delayed. In contrast, such a delay was not found for a singleton-geminate contrast in Italian: the movement timing for the vowels remained unchanged regardless of the length of the medial consonant. Based on these results, Smith (1995) proposed that: a geminate consonant was modelled as one gesture with different parameter values from those for a singleton consonant; and that different articulatory manifestations of a geminate can be considered as
language-specific choices of coordinative patterns, namely V-C timing for Japanese and V-V timing for Italian.

To summarise, we have discussed articulatory and acoustic evidence for describing Japanese as a mora-timed rhythm. The description of rhythm depends on how we approach temporal regularities observed in the flow of speech. On the one hand, they can be viewed as the durational characteristic of certain units. On the other hand, what is rhythmic can be described as systematic patterns that generate perceptually isochronous units. We found that a mora-timed rhythm was characterised by: (i) the directionality and the domain of temporal compensation; and (ii) the pattern of overlap between vocalic and consonantal gestures (i.e. V-C timing). It is important to note that these properties are not only observed for the language with a mora-timed rhythm, but also for the others, namely languages that are described as a stress-timed rhythm or as a syllable-timed rhythm. Crucial to this, is to what extent a certain temporal property is consistently and systematically realised within a given language of a particular rhythm. There is another characterisation of rhythm. Ogden (1995; p. 228) mentions:

Rhythm, which is in part what moras and syllables are meant to explain, is not just a matter of timing slots, nor of syllables or moras alone. Rhythm is a product of temporal relationships...timing relations must be an integral part of the statement of phonetic expomency rather than part of the phonological statement per se.

In this view the durational and gestural coordination patterns are only a part of the general phonetic characteristics of (a given portion of) speech such as resonance and phonatory quality. Such a characterisation might become possible only if we obtained more information about the relationship between articulatory and acoustic aspects of spoken languages.

2.8 Summary and Conclusion
In this chapter, we sketched the articulatory and phonological characteristics of spoken Japanese. We described the syllable structure, the general pitch-accent system, and both segmental representations and their phonetic interpretations for the phonological inventory and allophonic variation, particularly: labialisation; palatalisation; and affrication. We argued that the ‘correlation of palatalisation’ (Trubetzkoy, 1958 /1962) was one of the crucial factors in Japanese phonology and was variously conceptualised in previous work. The related phenomenon of CyV variability was explained. Both palatalisation and CyV variability will be further discussed in Chapter 6 on the basis of the results of the EPG experiments. We showed that the lexical strata were important for the phonological description of Japanese. The phonetic form of loanwords is one of the decisive factors that cause the fluctuation and expansion of the standard pronunciation habits. Different explanations are provided from different frameworks of phonology: the distinction between the ‘innovative’ and
‘conservative’ dialect by phonemic phonology; the stratum-specific feature specification by transformational generative phonology; and the core-periphery structure of the phonological lexicon by constraint-based phonology. We proposed the incorporation of allophonic rules into the core-peripheral continuum of the lexical strata.

For segmental articulations we developed our discussion based on previous experimental results of physiological, acoustic and perceptual studies. The articulation of vowels was reinterpreted, in particular the lip gesture for [ui], which was characterised as a ‘vertical contraction’ (Laver, 1980, 1994), rather than unrounded. We argued that the articulatory gestures for Japanese diphthongs were generally more sequential than those for English diphthongs. We demonstrated this idea through the gestural analysis of ‘smoothing’ in British English (Wells, 1982) and glide formation in Japanese loanword phonology. Finally, we discussed the articulatory aspects of a mora-timed rhythm, relating it to the production of moraic consonants. We argued that it was possible to consider the articulatory regularities and the acoustic irregularities in the /-NC/- sequences as a modified version of Stetson’s articulatory syllable.

From the analyses and observations presented in this chapter, it is understood that there are complex and close relationships between the articulatory organisation and the phonological system of the language. This suggests that it is possible to describe phonological phenomena using relevant articulatory parameters in line with the parametric approach. In addition, this supports our idea that, while the vocal tract is constrained biomechanically, it is organised phonologically. Although the areas that we surveyed are limited, it is reasonable to conclude that understanding of the physiological systems certainly complements the linguistic conceptualisation of speech production and spoken language.
Chapter 3

The Articulation of Coronal Stop Consonants: a Review

3.1 Introduction

In this chapter we discuss the salient articulatory characteristics of lingual gestures involved in stop articulations [t, d, n, ŋ, r]. The primary aim is to establish a background for the EPG experiment described in the next chapter (Chapter 4). Reviewing previous descriptive and experimental work on [t, d, n, ŋ, r] in Japanese and other languages, we explore a number of questions related to the spatiotemporal patterns of articulatory movements. We focus particular attention on speakers’ control strategies and their relations among the five stop consonants. Assuming a two-component model of the tongue physiology, we examine previous articulatory descriptions and experimental results for each of the five consonantal articulations in detail. We also outline the principal articulatory features of lingual activities. We provide evidence that the spatiotemporal requirements are systematically organised for the consonantal articulations in the same phonetic category. Some hypotheses about the production strategies are developed and presented.

3.2 Coronal Stops and Their Feature Specifications

Stop articulations are characterised as having a stricture of complete oral closure by the active articulator against the passive articulator (Laver, 1994: 205). This provides us with another perspective on the consonants [t, d, n, ŋ, r] that have traditionally been described by different manner features: plosive (or stop), nasal and tap/flap. These articulations are stop articulations in the sense that they employ the tip/blade component of the tongue to create a central blockage of the airflow along the alveolar, or palatal, region of the vocal tract. It is therefore possible to describe the essential movements as an articulatory sequence: approach, hold, and release. The properties distinguishing the sounds are the presence, or absence, of voicing and the ‘aspectual processes’ (Laver, 1994): nasal and tap/flap. Such a characterisation assumes that there must be systematic diversities and similarities underlying the production of the consonants [t, d, n, ŋ, r].

A good place to start is a set of phonological features that are commonly used for describing the articulatory properties of the stop consonants [t, d, n, ŋ, r]. Their feature specifications are illustrated in (1) where ŋ is classified as an alveolopalatal nasal ŋ and also as a palatal nasal ŋ.
The features that specify the stricture and configuration of the active articulator reflect the traditional places and manners of articulation. The consonants in (1) share the feature \([-\text{continuant}]\) as they are produced with the airflow blocked in the oral cavity. The feature \([+\text{distributed}]\) is used to indicate both the laminality of dental articulations in Japanese \(\{t, d, n\}\) and that of an alveolopalatal \([p]\). The latter consonant, in contrast to a palatal nasal, is treated here as a complex segment, in which both the tongue blade and the dorsum are involved in forming the primary constriction. This issue of articulatory complexity will be discussed in detail in Chapter 6.

Whereas the features in (1) are indeed correct, it has been argued that they do not adequately characterise the contrast between the plosives and the tap/flaps. Hall (1997) supplements the \([-\text{continuant}]\) specification with the new feature \([\text{flap}]\). Inouye (1989), in contrast, considers \([r]\) as a contour segment, the manner feature of which is similar to affricates, being composed of the changing value, \([-\text{continuant}]\) and \([+\text{continuant}]\). Underlying these proposals is the point frequently made that the articulatory gesture for the tap/flap is momentary and its closure duration is typically shorter than the other stop consonants. Thus, the feature \([\pm\text{continuant}]\) that typifies the speed of the articulatory gesture, rather than the major class feature \([\text{sonorant}]\), is considered as accurate for the distinction between \([r]\) and the plosive. Furthermore, it is tacitly assumed, in the identification of various stop articulations, that the durational feature is consonant-specific. When the voiceless and voiced plosives are compared, the closure duration is frequently referred to as one of the distinctive articulatory parameters. Yet, the characteristic is rarely taken into account for the articulatory description of the nasal stop. In any case except for \([r]\), such a durational property in the actual realisation is left to the low-level phonetic rules and has not been examined seriously. When the production of speech is viewed as movement, the duration undoubtedly functions as one of the articulatory parameters. It can be asked how these parameters are

1) By the feature \([\text{distributed}]\) Chomsky & Halle (1968; 312ff.) refer to 'the length of the zone of contact' and intend to distinguish apical from laminal and retroflex from non-retroflex consonants. Although a large amount of tongue-palate contact is generally observed in laminal dentals and in alveolopalatales, the constriction length can vary language-specifically: apical constrictions can be long while laminal constrictions can be short (Keating, 1991). Therefore, we use the feature \([+\text{distributed}]\) to refer to the constriction place on the tongue, namely the blade.
organised; and what is considered as linguistically significant during the articulatory activities in question.

Research on coarticulation supports descriptions based on systematic articulatory control and the coarticulatory activities underlying those stop articulations. In the following discussion, the articulatory and coarticulatory patterning is characterised in terms of a two-component model of the tongue physiology. In the production of the coronal consonants, the tip/blade (anterior) component acts as the main articulator, whereas the dorsum (posterior) component is active both for the consonantal articulation and for the contextual vowel articulations. The two components are functionally divided but closely interrelated. The activities of the tip/blade component in the consonantal articulation constrain, to some extent, those of the dorsum component. Such a dependency relation, or the degree of coupling, is often associated with the extent to which the vowel articulation is overlapped by a given consonantal articulation. Our strategy will be to extract and examine the distinctive movements of the tip/blade component, the degree of dependency between the two components, and the degree of coarticulatory effects together with their temporal extent. Given this, these spatiotemporal manifestations are related to the factors of the articulatory control mechanisms: the consonant identity, and the speaker’s control strategies used in the production.

In the following sections the tip/blade activities and their relation to the dorsum activities during the production of [t, d, n] are explored. We consider oral and nasal stop articulations first, and then move on to tapped/flapped stop articulation.

3.3 The Articulation of the Japanese [t, d, n]

3.3.1 Configurational Characteristics

The consonants [t, d, n] and [p] have place-neutral articulatory gestures (Laver, 1994); the part of the tongue making a contact against the palate is that which lies directly under that location. The consonant [r], in contrast, may or may not be place-neutral: it depends on how much the tongue tip is curled back. It is reported that, while the three ‘place-neutral’ consonants [t, d, n] are described as sharing the same place label on the palate, the point of articulation on the tongue that is used for making a central articulatory closure differs language-specifically and idiosyncratically. It is not unrealistic to assume that a language has its own pronunciation routine, or ‘articulatory setting’ of the tongue (Honikman, 1964).

One focus of the EPG experiment in Chapter 4 is to investigate, in detail, the position and amount of the tongue tip and blade used for making the central constriction during the production of Japanese [t, d, n] as a function of the vowels. As defined by Catford (1977; 143), the tip is the central point of the forward edge of the tongue (the rim), and the blade is the part that lies opposite the teeth and alveolar ridge when the tongue is at rest, that
is, the tip plus about 10 to 15 mm along the central line. These anterior portions of the tongue are related to a feature such as apicality, or laminality, of the consonantal constriction. The importance of the feature lies in the fact that there is an interaction between the tip/blade and the dorsum components of the tongue; any activity tends to affect the overall spatial configuration of the tongue. We shall argue that the concept of language-specific tongue settings must be taken into account for understanding of the voicing and aspectual effects on the consonantal articulation, the phenomena that are commonly explained by appealing to the concept ‘force of articulation’.

Previous descriptions of Japanese oral and nasal stops have suffered from vagueness concerning the point on the tongue involved. Based on two parameters such as place of articulation on the palate and point of articulation on the tongue, the Japanese [t] and [d] have been traditionally described as apical-dental, a complete occlusion which is formed by pressing the tongue tip against the upper teeth (e.g. Hattori, 1984; Kawakami, 1977). The same specification is applied to [n]. In contrast, Edwards (1903; 27), comparing Japanese [t, d] with French and English, points out that:

Le son japonais tient le milieu entre le français et l'anglais: on appuie la pointe de la langue entre les dents d'en haut et les gencives, on ferme le passage en même temps avec une partie de la face de la langue qui est très aplatie.

Hattori (1984; 105, note 15) also notes the possibility that not only the apex but also some portion of the blade may be involved in making contact. Akamatsu (1997; 78f.) labels [t, d, n] as apico-dental/alveolar to reflect the participation of the tongue blade. The disagreement on the articulatory point on the tongue comes from the difference in the use of terminology and the lack of empirical support.

The pioneering EPG studies of Shibata (1968) and Fujimura et al. (1973) examined the spatiotemporal characteristics of various consonants in the vowel /a/ symmetrical context spoken by one Japanese speaker. Yet, the reports are limited to some specific consonants: the data for [n] is not available. Neither study explicitly discusses the configurational characteristics of the lingual gestures for [t] and [d]: Fujimura et al. (1973; 50) merely describe the two consonants as ‘an apical lingua-palatal contact’. By ‘apical’ they probably meant ‘apicolaminal’, because their EPG data shows a reasonable amount of contact at the front region. It is reported that the front contact for [t] is larger in degree and longer in duration than that for [d]. These spatial and temporal aspects of lingual contact agree with the results that are obtained by EPG techniques in various languages: Connell (1992) for Ibibio, Dixit (1990) for Hindi, Engstrand (1989) and Simada & Gauflin (1983) for Swedish, Farnetani (1989) for Italian. However, there is no further empirical evidence in Japanese to describe and discuss such a use of the lingual gestures, or coarticulatory effects that might
occur in the consonants.

The spatiotemporal characteristics above are commonly labelled as ‘tense vs. lax’ or ‘fortis vs. lenis’. It has been claimed that tense (fortis) consonants such as [t] are produced with greater pulmonic force and with greater pressure on the articulators involved, compared to lax (lenis) consonants such as [d] (e.g. Jacobson, Fant, & Halle, 1952; Malécot, 1968). Brown and Commerford (1979) found that intraoral air pressure rises before complete lingual closure, which is faster and greater in [t] than in [d], this characteristic being maintained until the release. Slis (1970) reported that intraoral air pressure decreases from [t] through [d] to [n]. These aerodynamic facts and the durational differences in the articulatory closure are considered as the measures for the articulatory energy, and promote the causal relationship among two consonant types, lingual pressure, and the amount of linguopalatal contact. The whole discussion of the ‘force-of-articulation’ is beyond the scope of this chapter. We will limit our discussion to the reliability of the linguopalatal contact degree as an indicator of the articulatory energy in the coronal consonant, and will consider whether the two plosives in Japanese are eligible for the tense/lax, or fortis/lenis, distinction.

Fujimura et al. (1973; 50ff.) asserted that ‘the consonant [t] was always “tenser” in the apical linguo-palatal contact than [d],’ on the ground that there is ‘a clear correlation between the degree of contact and the duration of closure’. The authors regard the two properties, the contact amount and duration, as an indicator of tenseness. It is implicitly assumed that the greater contact area in [t] reflects the involvement of the greater articulatory energy. Such a characterisation is also made by Shibata (1968). If the amount of intraoral pressure, linguopalatal contact, and the consonant types, are proportionally associated amongst themselves, that association will be one of the possible factors underlying the unity of stop articulations in general and useful in the phonetic description of the given consonant or a set of consonants.

However, the opposite result in the spatial aspects of [t, d] is reported by Dagenais et al. (1994). They obtained EPG recording from ten American English speakers and found that the voiced cognate had a greater amount of central contacts at the alveolar region, while lateral contacts remained relatively the same as in the voiceless cognate. We must notice here that the aerodynamic facts mentioned above are not necessarily reflected in the larger amount of tongue-palate contact. Dagenais et al., nevertheless, put the emphasis on one side of the concept ‘articulatory force’, saying that ‘...the greater length of contact for the voiced cognate could reflect a more relaxed tongue posture or a greater spread of tongue contact because of the comparatively smaller intraoral pressure’. The authors continue that ‘[a] more rigid tongue posture [for the voiceless stop] could result in the tongue not extending as far

---

2) See Catford (1977; 199) for a concise summary.
into the palatal vault, thus accounting for the shorter midline contact (231). They failed to account for what they referred to as ‘a more rigid tongue posture’; the argument becomes circular.

The lingual pressure data also brings into question the implicational relationship between the contact amount and the articulatory energy involved. McGlone et al. (1967) and Brown et al. (1973), using an artificial palate with pressure transducers, measured the tongue pressure at the maxillary central incisors and at the left and right maxillary first molars. The informants were presumably native speakers of American English. In the production of the sequence V1CV2 where the vowels were the symmetrical combinations of /i, a, u/ and the consonants were /t, d, n/, Brown et al. (1973) found that, while the average peak intraoral air pressure consistently decreased in the order of [t]>[d]>[n], the tongue pressure at the central articulatory closure exhibited a different trend from that at the lateral closure. The relative central lingual pressure values were found to be [n]>[t]>[d] in the /i/ and /u/ contexts. In the /a/ context, in contrast, they were found to be [d]>[t]>[n]; the lateral lingual pressure was proportional to the intraoral air pressure values. If such a relationship between aerodynamic and articulatory aspects is applicable to all the languages, the results indicate that the tongue pressure, particularly at the central closure, varies with surrounding vowels and individual speakers.

From these remarks, it appears that the direct causal relationships among intraoral air pressure, lingual pressure and linguopalatal contact degree are very complex and are not universal in the coronal articulation. If we carefully select one specific variable that may make a contrast between the voiced and voiceless plosives, then, as Catford (1977; 208) mentions, there is some justification for the retention of a parameter of tense/lax for the description of consonants. One such example can be found in Elugbe (1978). In Ghotuọ (a North-central Edo language of Bendel State, Nigeria) there is a phonological contrast between non-tapped and tapped articulation in alveolar nasal and lateral such as [năn3] ‘today’ and [nănã] ‘move like an ant,’ [lã] ‘flow’ and [lã] ‘be rotten’ (134). Elugbe demonstrated, using spectrograph and electro-aerometer, that these two phonologically contrastive classes were distinguished by the duration of the oral closure: the closure duration of tapped nasals and laterals ranges from about 35msec to 60msec, whereas the duration of the non-tapped counterparts is from 80msec to 130msec. This allows Elugbe to conclude that duration is a primary feature differentiating lenis (i.e. tapped) and non-lenis (i.e. non-tapped) sound, suggesting that the lenis consonants are the marked ones and the non-lenis consonants are not necessarily fortis. This study also implies that the phonetic correlations of tense/lax or fortis/lenis may differ language-specifically. It seems that such a dichotomy is most beneficial when it is closely related to phonological function. Thus, the traditional distinction, voiced and voiceless, is preferred for a language like Japanese, in which ‘a voiceless stop is
distinguished from a voiced stop in terms of a higher F0 at vowel onset' (Shimizu, 1996), rather than VOT values.

There is another aspect of speech production that has received little attention in previous studies. That is the feature of ‘articulatory settings’ (Honikman, 1964). If we assume, with Honikman (1964) that there is a language-specific framework for the overall arrangement and manoeuvring of the speech organs, then the tongue setting, among other factors, effectively constrains the freedom of movement. Such a hypothesis makes it possible to interpret the ‘unfamiliar’ spatial pattern reported by Dagenais et al (1994) as the reflection of phonetic individuality regulated for the tip/blade component of the tongue. In other words, the difference in the contact amount can partly be attributed to the different shaping activities of the tongue. Support for this view comes from Dart (1998).

In a direct palatographic and linguographic study of twenty-one French and twenty American English speakers, Dart (1998: 75) investigated the participation of the tongue tip and blade in the production of alveolar consonants [t, d, l, n, s, z]. Four different patterns of tongue contact are generalised: (i) apical, in which the very tip of the tongue makes contact; (ii) upper apical, in which the upper surface of the apex, not the rim, is involved; (iii) apicolaminal, in which both the apex and the blade make contact over a wide area; and (iv) laminal, in which only the blade makes contact but the tip is kept entirely out of the way, perhaps down behind the lower incisors. Dart (1998) found that, in spite of a great deal of individual variation even within one language, there were general trends, or preferences, of the articulations. The highest percentage of American English [t, d, n] is apical with alveolar articulations, with the next most common point on the tongue being laminal for [l], upper apical for [d] and [n]; French [t, d, n] are mostly apicolaminal with dental articulations, followed by laminal for the three consonants and upper apical for [d]. These trends would help to account for the data of Dagenais, et al. (1994) mentioned above: presumably [t] and [d] are articulated at the alveolar ridge with apical and apicolaminal modification respectively: the degree of the central contact turns out to be higher in [d].

It is reasonable to suppose that such a control of the tongue produces the difference not only in contact pattern but also in the overall shape. When the tongue tip is raised and pressed against the alveolar ridge as in English [t], the overall tongue shape may become concave to the palate. In contrast, when the tongue tip and blade touch the upper front teeth and alveolar ridge, the tongue becomes convex to the palate. This may give us an important insight on phonetic processes in a given language. Ladefoged & Maddieson (1996: 23) mention that ‘if a language has both an apical and a laminal stop consonant, then the laminal consonant is likely to be more affricated’. In addition, such dispositions of the tongue-point and coarticulatory effects are correlated. Recasens et al. (1993) report that, while the apicality, or laminality, of Catalan [n] varies with speakers, the laminal realisation shows more stable
tongue-palate contact than the apical realisation. Furthermore, important implications for aspects of so-called foreign accent can be reached. Jones (1960 * 142f.) shows the palatograms of British English /t/ and French /t/, suggesting that 'this articulation [a dental consonant production in French] produces a very unusual effect when used in English'.

Overall, the consonants [t, d, n] are described using the same place-label both on the palate and on the tongue, but this seems to be an oversimplification. The tongue-point at the central constriction varies language-specifically and idiosyncratically. At the same time, the contact pattern at the central constriction may be systematically controlled with reference to the consonant identity such as voiced/voiceless and nasal. Hence, the amount of contact depends partly on the tongue-point used and partly on a type of a consonant. In addition, we can say that articulatory settings, the tongue setting in particular, may have important benefits on the patterning of lingual gestures. Such a preference can be considered not only as an attribute of a given language but also as an idiosyncratic property of individual pronunciation. This suggests that, while there are certainly general production constraints on the stop gestures involved in [t, d, n], their different tip/blade activities may impose different degrees of constraints upon the dorsum activities specific to [t], [d], and [n]. Consequently, it is likely that the degree of the coupling effects between the two components of the tongue and articulatory overlap with the surrounding vowel gestures systematically vary with the consonant types.

3.3.2 Movement Characteristics of Japanese [t, d, n]

Turning now to the coarticulatory activities in the coronal consonants [t, d, n], we deal with intergestural coordination between the tip/blade and the body components of the tongue. The effects of the vowel articulation upon the formation of the tip/blade consonants can be viewed as anticipatory or carryover effects upon the tongue body. If the tip/blade articulation effectively constrains the tongue body movements during the given consonantal articulation, the patterns of coarticulatory variation vary with the consonant identity. This leads us to the question of how the consonants systematically differ with regards to the control of the two components. Previous studies that systematically analyse the coronal consonants as a class are very limited. We thus put together the facts relevant for our discussion.

Kent & Moll (1972), using the cinefluorographic method to detect the tongue movements in symmetric VCV sequences with the vowels /i, a, u/, reported that the displacement of the posterior tongue varied in the production of /g/ as a function of the vowel, but not in the production of /d/ and /z/. It is suggested that the movement of the tongue body is consciously controlled in the formation and release of the coronal consonants; the contextual vowel gesture is not freely incorporated in the consonantal gesture. This is interpreted as the evidence for a strong physical coupling effect between the two components.
of the tongue. In other words, the production of the coronal consonants, /d/ and /z/, imposes constraints, to some extent, on the tongue body movement.

Gay (1977) also applied cinefluorographic techniques for tracking tongue and jaw movements in the production of the sequence /kVCV/p/ which was composed of the vowels /i, a, u/ and the consonants /p, t, k/ in all possible combinations. It was found that, while the types of the intervocalic consonant affected the vowel-to-vowel movements, the relative gestural timing from the first vowel to the consonant were far more stable than that from the consonant to the second vowel. Anticipatory movements towards the second vowel always began shortly after closure for the consonant was attained; the anticipatory effects do not extend beyond the intervocalic consonant. On the other hand, carryover effects of the first vowel upon the consonantal production were found to be complicated: the effects were not adequately measurable in /p/; they did not appear in /t/; and they appear in a predictable way in /k/. These findings lead Gay to a proposal that the CV component of the VCV sequence might be organised as a basic unit. The coupling between the two components of the tongue is rather loose in the production of /t/.

The x-ray system used in Gay (1977) and Kent & Moll (1972) images the articulatory activities as horizontal and vertical movement contours; it cannot provide information about the spatial, or configurational, characteristics. Such a limitation of the technique may be closely related to the presupposition of Gay (1977), presumably shared by Kent & Moll (1972): while anticipatory coarticulation effects are essentially timing effects, carryover coarticulation effects are essentially mechanical effects and exist in the form of variability in target (or target feature) positions as a function of changes in phonetic context. Thus, the lack of carryover effects reported by Gay’s experiments means that the tip closure for /t/ is always made at the relevant point on the palate, even though the height of the tongue body varies with that of the preceding vowel.

Butcher & Weiher (1976), using the Kiel EPG system, recorded the production of V$_1$CV$_2$ nonsense words (C=/t, k/ and V=/i, a, u/) spoken by three German speakers. It was found that, while the two consonantal articulations show different coarticulatory activities, anticipatory movements towards a V$_2$ gesture were stronger than carryover movements from a V$_1$ gesture. Butcher & Weiher explain diverse manifestations of coarticulatory effects in the light of Ohman’s V-to-V model (1966). In the case of [t] the tongue tip, Butcher & Weiher (1976, 72) argue, is independently controlled and leaves the tongue body free to maintain the vocalic diphthongal gesture. In the case of [k], in contrast, the dorsum raising gesture towards the velum creates a great barrier to coarticulation across articulatory closure.

Farnetani et al. (1985) examined linguopalatal contact patterns during the production of intervocalic /t/ in a sequence of /tV$_1$TV$_2$/ (V=/i, a, u/), with or without stress on the second syllable, spoken by one Italian speaker. The degree of dependency between the
The results show that the relation between the tip/blade and the body is rather loose and that the carryover effects are stronger and more systematic than the anticipatory effects. It is suggested that the contextual vowel affects the height of the tongue body much more than that of the blade in /VtV/ productions; the tongue tip is free from the contextual effects. This agrees with the results of /t/ reported by Gay (1977) and Butcher & Weiher (1976). One puzzling effect that Farnetani et al. report is that anticipatory effects can extend to the pre-consonantal /u/, while carryover effects never extend to the post-consonantal vowel.

The issue of tongue control during the articulation of [n] has been given little attention. This situation contrasts with what we know about the velopharyngeal mechanisms for the nasal consonants in general. Using the Reading EPG system, Recasens et al. (1993) examined VCV sequences, where the vowels were [i, a, u] in symmetrical combination and the consonants were [n, l, j, n], spoken by five Catalan and three Italian speakers. The results indicate that the constriction shape for [n] varies with individual speakers ranging from apico-alveolar to apicolaminal-dental across the languages. The former type of realisation is likely to be more affected by the contextual vowel articulation, as the constriction place on the palate shifts forwards and backwards. It was found that the degree of the tongue dorsum involvement depends on the contextual vowels; it decreases in the order of [i]>[a]>[u] in both languages. It should be noted that, while the hierarchy is a general trend across the languages, the qualitative difference in the vowels produces language-specific effects for the contact patterns. These results suggest that the dorsum component adopts the gestures for the contextual vowels while the tip/blade component takes on the movements for the complete occlusion.

Two EPG studies facilitate our discussion. One is an investigation into a set of the coronal articulations within one language, and the other is a cross-language investigation into the articulatory nature of [t]. Connell (1992) obtained the EPG recording of VCV words in which the consonants are [t, d, n, k] in the vowel symmetrical contexts of [i, a, u], spoken by one Ibibio speaker. The five consonants were compared in terms of the degree of the tongue dorsum involvement and coarticulatory effects across the vowel contexts. It was found that the amount of the dorsal contact decreases along the hierarchy [n]>[k]>[t]>[d]. On the other hand, the constriction location and the overall contact become less stable in the order of [t]>[n]>[k]>[d]: in the production of the three consonants except [t], the constriction location is formed further forward in the context of /i/. Connell adds the comment that the behaviour of [n] is similar to that of [d]. The hierarchies demonstrate complex interactions. While the consonant [d] that is least affected in the dorsal involvement is most unstable in the constriction location and the overall contact, the consonant [t], which has the next smallest
dorsum contact, shows the most stable contact location and overall contact. This implies that the tip/blade articulation of [d] executes a stronger constraint upon the dorsum involvement than that of [t] even though the constriction place shifts with the types of the contextual vowels. The consonant [n] indicates the highest amount of dorsal contact since [n] involves the dorsum activity in its formation. It is worth noting that, based on the palatograms presented by Connell, Ibibio [t] can be considered as an apicocolaminal-dental articulation but [d] can be regarded as an apico-alveolar articulation. These configurational characteristics may explain why [t] has less coarticulatory variation than [d]. The more laminal the tongue-point becomes, the greater the contact stability attained. The hard palate provides more support for the apico-laminal constriction of [t] than for the apico-alveolar realisation of [d] (cf. Bladon & Nolan, 1977: Recasens et al., 1993). This suggests that there is a functional correlation between the feature of the tongue-point and the hard palate (Stone, 1995).

Farnetani et al. (1989) conducted a cross-language comparison of linguopalatal contact pattern using the Reading EPG system. They obtained the EPG recording of /bV'CV/ disyllabic nonsense words where the consonant was /t, k/ and the vowel were all possible combinations of /i, a/ spoken by French, Italian, and English speakers. The anticipatory and carryover effects were measured at the onset and offset of a central articulatory closure respectively. For [t] in the symmetrical contexts, Italian shows a stable positioning of the tip/blade component for the complete occlusion at the closure onset and the vowel effects are limited to the dorsum component, which is relatively free. In French, on the other hand, the constriction location is further retracted in the vowel /i/ context with the dorsum contact increased. However, in English the main constriction location is the same at the alveolar region (row 2) but both the posterior half of the tip/blade component and the dorsum component are raised in the vowel /i/ context. These results suggest that there are language-specific production strategies for the sounds described by the same phonetic symbol; the coupling between the two components is loose in Italian compared to French and English. In the asymmetrical contexts, it was found that there is also language-specific preference for the anticipatory and carryover effects. The overall contact pattern in Italian [t] is least affected at the onset of articulatory closure (i.e. anticipatory preferred), while in English only the contact pattern at the offset is affected (i.e. carryover preferred). French [t] shows the influences in the anterior half of the palate at both points in the constriction (i.e. no preference).

From previous studies that we have discussed, it is possible to discover two lines of investigation: first, into the articulatory nature of the tip/blade consonants as compared to the dorsum consonants; and second, into the articulatory and coarticulatory differences between some coronal consonants that differ in their manner of articulation. Although the direct comparisons between the studies above are difficult, the hypothesis that the tip/blade
activities in different coronal consonants constrain the dorsum activities is supported in essence. Such consonant-specific adjustments of the dorsum component are compatible with the configurational characteristics of the tip/blade component of the tongue that we have seen in the last section. Coarticulatory effects of the coronal consonants manifest in the spatiotemporal domain and are partly conditioned by the different degrees of the tongue dorsum raising and positioning for the vowel articulation. There is a potential link between the dependency relations of the two components and coarticulatory effects. As Farnetani et al. (1989) imply, one must distinguish between the coupling and coarticulatory effects within a language on the one hand, and those across languages on the other. Different languages show different directionality preference for coarticulatory effects. The question of how these articulatory parameters are systematically organised in one phonetic category has not been examined.

3.4 Movement Characteristics of the Japanese [n]

Trained phoneticians disagree over the description of the nasal consonant before a high front vowel [i] in Japanese. This nasal consonant is analysed as a conditional allophone of /n/ before /i/. The consonant in question is considered a palatal nasal [n] by Hattori (1984) and Kawakami (1977), but a kind of [n] with [i]-like posture by Edwards (1903) and Sakuma (1929). Edwards invented a phonetic symbol [n^r] for transcribing the nasal consonant and later Sakuma used [n]. Akamatsu (1997) also uses a special symbol for the consonant, characterising it as laminodorsoro-alveolopalatal nasal, as opposed to dorso-palatal nasal [n]. Akamatsu (1997, 122) provides the detailed phonetic description of the articulation as follows:

In articulating the Japanese [n], the tip of the tongue is not in contact with the upper front teeth or upper front teeth-ridge, but it is low and tucked away behind the lower front teeth...the blade of the tongue forms a closure with the hinder part of the teeth-ridge (hence lamino-alveolar) as well as the forward part of the front of the tongue with the forward part of the hard palate (hence dorso-palatal).

Takebayashi (1970) mentions that the quality of the consonant varies from [n^r] to [n]. In addition, it has been common practice to use the same phonetic symbol for the nasal sound before non-front vowels /a, o, u/, where unlike [n] before /i/ the consonant is phonologically contrastive and is analysed as a phonemic sequence of /n/ plus /j/.

The above remarks, even though they largely rely on impression and conjecture, suggest that the point of articulation moves forwards and backwards with large amplitude. It is generally agreed that the source of variability is the effect of palatalisation: the high front vowel /i/ or the palatal approximant /j/ influence the articulation of the preceding consonant. This view is apt to be misleading because it identifies articulatory events with the
phoneme-allophone relationship (Lass, 1984; 169ff; Wells, 1974). It provides no justification either for systematic control of articulators involved in the production of the consonant, or for extensive variability in the constriction place observed. This leads us to the question of how the consonant is actually articulated and why such a potential variation is involved in the given articulatory gestures.

As we pointed out above, the consonant [n] is considered as place-neutral: the tongue dorsum is raised and pressed against the palatal region with the tongue tip directed downwards. However, such a neutral status of [n] has been questioned. Catford (1977) characterises [n] as dorso-prepalatal nasal as well as dorso-palatal. The latter corresponds to a place-neutral gesture and is formed relatively further back. Catford implies that the consonant is more often produced as dorso-prepalatal than dorso-palatal. Although the difference between the two gestures lies in the point on the passive articulator, it is likely that the whole body of the tongue effectively moves further forward in the dorso-prepalatal realisation.

Recasens (1990) proposes another characterisation based on visual inspection of X-ray pictures and EPG experiments on the production of various palatal consonants in Czech, Slovak, and Catalan. Citing the side-view of X-ray data of [n] produced by a Czech speaker, Recasens points out that the consonant is formed primarily by pressing the tongue predorsally, rather than laminally [blade], against the intersection between the alveolar ridge and the front arch of the hard palate. The consonant [n] is thus classified as alveolopalatal; one of the four categories that are distinguished for consonants produced at the palatal zone. Although it is acknowledged that there is possible variability for the tongue blade and dorsum to participate in forming the complete occlusion, Recasens (1990; 277) asserts that their participation is mechanical, or automatic, as far as the tongue tip is folded down and the tongue dorsum is lifted up. This implicitly suggests that the 'true' palatal, namely place-neutral, realisation of [n] is atypical or unattested.

A previous EPG study by Miyawaki et al. (1974) examined /n/ in the VCV and VC/j/V sequences, where the vowels were possible combinations of /i/ and /u/, spoken by two Japanese speakers. It is reported that such spatial effects as an overall increase in contact and a 'backward shift' of the constriction were found in both /n/s at the closure mid-point. The constriction 'moves backwards,' since the authors attempt to describe the consonant with reference to the articulation of [n]. The EPG (the Reading system EPG2) full printouts of

Recasens (1990) proposes four categories, rather than a single class of palatal consonants: (i) [j] and [tʃ] are lamino-postalveolars; [n], [ʎ], [c], [ç], and [tʃ] are alveolopalatals; [j] and [ç] are front palatals; and [kʰ] and [kʰ(i)] are back palatals.

Keating (1988) also characterises [n] as alveolopalatal based on visual inspection of lateral X-ray data (a Czech speaker). Although the labelling is the same as that by Recasens (1990), the specification of the articulator(s) involved in the production is different. Keating (1988; 81) states that 'they [palatals] use the very back of the blade, and the large front part of the dorsum'. We will discuss these two views in detail in Chapter 5, where we deal with Japanese palatalisation.
Matsuno (1989), in contrast, indicate that the consonant displays a larger area of tongue-palate contact in the posterior regions, but no backward shift at the moment of maximum constriction. In the articulation of a sequence /ni/ the full tip closure similar to [n] is attained. It is worth pointing out that the onset (first frame) of the articulatory closure is made at rows 2-4 and the full tip closure is removed first towards the release. In contrast to [t, d, n, r] the consonant [n] involves the tongue dorsum component in its formation. It is expected that the degree of coupling effects between the two components is relatively strong. These findings by and large agree with the impressionistic phonetic descriptions above, suggesting that the sound commonly described by [n] in Japanese, when followed by [i], involves particular control strategies of the tongue.

The extent to which the larger amount of contact observed in palatal consonants is ‘actively’ controlled is controversial. Given Recasens’ proposal, the variations, with or without a ‘backward shift’ of the constriction place, in Japanese [n]-[n] are regarded as automatic consequences of the articulatory gesture for [p], in which the tongue predorsal region is actively pushed against the relevant place on the palate. This account is acceptable only if we assume that the articulation of [n] is identical across different languages and speakers of a language. Furthermore, it is in reality very difficult to assert that only a specific region of the tongue is under active control; the tongue predorsal is operative but the rest is not. Rather, it is more plausible to assume that different languages utilise different production strategies for the consonant. As far as the EPG data of Miyawaki et al. (1974) and Matsuno (1989) are concerned, the central constriction of Japanese [n] is formed slightly further forward, compared to that of Catalan and Italian (Recasens et al., 1993). The findings of Miyawaki et al. (1974) and Matsuno (1989) need further clarification particularly concerning the coarticulatory effects that may occur in the production of [n] as a function of the vowel. Such articulatory characteristics create important consequences for the feature specification of [n]. It can be either [+coronal] or [-coronal] depending on the location of a central constriction and the tongue-point involved. It would be more accurate that the coronality of a sound broadly described by the symbol [p] is determined language-specifically.

3.5 Tap/flap Articulation
A second focus of the EPG experiment is to examine specific production constraints especially for tap/flap articulation. In Japanese an alveolar tap [r] has been treated as a major allophone of the /r/ phoneme. Before turning to a closer examination of Japanese /r/, a few remarks should be made concerning tap/flap articulation in general.

Two properties are common characteristics of tap/flap articulation. They are the speed of articulatory gesture and the incomplete or lesser amount of tongue-palate contact. Ladefoged (1993: 168), for instance, mentions that ‘A tap or flap is caused by a single
contraction of the muscles so that one articulator is thrown against another. It is often just a very rapid articulation of a stop'. It is explicitly assumed that the articulatory mechanisms of taps and flaps are essentially the same as that of stops, except that the tongue movement is very rapid. Byrd (1993) acoustically measured the duration of various stop articulations using the TIMIT database of American English. It is reported that the mean duration of oral flaps is 29 msec, that is shorter than the mean duration of oral stops [t] (53 msec) and [d] (52 msec). For the second property, it is pointed out that a tap/flap may involve a lesser, or incomplete, contact between the articulators than the corresponding stop (Catford, 1977; 251, note 5; Connell, 1995; 46; Gimson, 1980; 207). Thus, taps and flaps generally stand for a manner of articulation that involves a brief incomplete contact between the articulators.

Some draw a distinction between the two articulations, tapped and flapped (e.g. Catford, 1977; Ladefoged & Maddieson, 1996; Laver, 1994). Ladefoged & Maddieson (1996; 231) suggest that taps are considered as a sound 'in which a brief contact between the articulators is made by moving the active articulator directly towards the roof of the mouth'; and that flaps are characterised as a sound 'in which a brief contact between the articulators is made by moving the active articulator tangentially to the site of the contact, so that it strikes the upper surface of the vocal tract in passing'. Catford (1977; 128f.) describes these gestures as 'flick' and 'transient' types of flap articulation respectively. Crucial in the distinction is the tip gesture involved in the momentary constriction; there are differences in the place of articulation which corresponds to two IPA phonetic symbols, [r] and [ɾ], respectively. The duration of both types of articulation, tap/flap or flick/transient, is 10-30 msec, which is shorter than that of most stops, 50-60 msec (Catford, 1977; 130).

The question arises as to where the articulatory nature of taps or flaps comes from. Given that the combinations of the three properties, such as the speed, the incomplete contact, and the retroflex gesture, constitute the articulatory nature of tap/flap articulation, there are at least two possible interpretations. If we assume that the speed, or the shorter duration, is the truly distinctive feature, a smaller area of contact between the articulators can be regarded as one of the consequences of the quick tongue movement. The different degree of a retroflex gesture, then, may be a by-product of the speed, or is independently motivated in a given language. On the other hand, it is possible to assume that the shorter duration is one of various means to achieve a lesser, or incomplete, contact between the articulators. This is suggested by Chomsky & Halle (1968; 318) is somewhat different: '...the tap [ɾ] [one-tap trill] may be produced by a different mechanism than the so-called "tongue flap" [D] which greatly resembles the tap [ɾ]. Whereas the latter [the tap [ɾ]] is the result of the aerodynamic mechanism just described [Bernoulli effect], it is quite possible that the tongue flap [D] is produced by essentially the same muscular activity that is found in the dental stop articulation, except that in the case of the tongue flap the movement is executed with great rapidity and without tension'. Their flap [D] is differentiated from the tap [ɾ] (one tap-trill) mainly by aerodynamic mechanisms but from the stop by the rapidity of the tongue gesture.

---

118
by Connell (1995). We must examine, to some extent, Connell’s claim.

Connell (1995) asserts, referring to Sawashima & Kiritani (1985) mentioned above, that ‘the shorter duration might be needed, as one of a variety of means, to generate such a contact pattern (47)’. In his EPG study of Ibibio Connell demonstrates that the stop [d] involves very brief complete closure in the (post)alveolar region and a ‘forward rolling’ movement, which, we will see, is very similar to the intervocalic [ɾ] in Japanese; and that the tap [ɾ], in contrast, consistently shows no central contact in the anterior region but only a few contacts at both sides in the posterior region. It was found that the duration of the approach phase is shorter but that of the hold and release phases and the total duration are longer in the tap. Connell suggests that such incomplete contact may be considered as an instance of ‘undershoot’ not by the greater velocity required for the tongue tip, but as a result of less muscular activity involved in the gesture (47). Therefore, a tap/flap is seen as a weaker, or lenis, form of a stop. Such a claim seems to miss the point. Rather, it is more plausible that Connell’s EPG data suggest that the articulation described by the same manner label has different attributes depending on the language.

The two possibilities above can be derived when we focus mainly on the articulatory activities during the hold phase. It is implicitly presupposed that, while tap/flap articulation is regarded as one particular type of articulation, it is essentially the same as stop articulation. This agrees with Laver’s (1994) characterisation that we have accepted.

Alternatively, when the overall articulatory process is taken into consideration, it can be argued that the three characteristics functionally aim at imposing a certain production constraint on the tongue, for instance, its overall shape. Furthermore, when we give attention to the activity in the release phase of tap/flap articulation, they are much more comparable with that of approximant or resonant articulation, than stop articulation. We develop this third possibility, critically reviewing previous treatments of Japanese tap/flap articulation.

3.6 Movement Characteristics of the Japanese [ɾ]
The Japanese /ɾ/ is mentioned as an example of the variphone, a sound which is unstable and susceptible to variation independently of the phonetic context (Jones, 1967 205f.). Although it has been a common practice to transcribe the Japanese /ɾ/ as an alveolar tap [ɾ], articulatory descriptions are quite divergent. They fail to grasp the accurate status of the gesture itself, causing difficulties with its classification, which are divided into two main groups.

We will argue that tap/flap articulation exhibits diverse characteristics from stop articulation in the release phase of articulatory sequence. We will propose that, if the third phase is split into

---

5) The third possibility is to consider Japanese /ɾ/ as a true variphone, the members of which are composed of [ə, ɾ, ʤ, ɭ] and their intermediate articulations (Jones, 1967 205ff.).
release and explosion\(^6\), the articulatory nature of taps or flaps, in relation to stop articulation, can become more approachable.

The first group of the classification is to view the initial /r/ as a kind of [d] but the medial /r/ as a kind of [ɾ]. Hattori (1984; 78; 83) states that ‘the initial /r/ can be regarded as a slack stop articulated by touching the tongue tip and the underside of the rim at the postalveolar region, while the medial /r/ is usually a tap hitting the alveolar ridge with the tip of the tongue [my translation]’. Despite such a tongue movement, the initial /r/ is not admitted to be a voiced retroflex stop [d] or a flap [ɾ]. Hattori’s description postulates a continuum of the closure strength associated to stop consonants. Without any empirical evidence, he adds the comment that ‘the (initial) /r/ needs lesser airflow and is weaker in closure strength and in explosion (78) [my translation]’. There is no clear indication of what is meant by ‘a slack stop’, either. If a ‘slackening’ articulatory closure implies a momentary blockage of the airflow, this may be related to an acoustic compactness, which is a general characteristic of [ɾ] on the spectrogram. Another possibility is that a loose-fitting seal in the oral cavity may not make a clear-cut sudden separation of the tongue from somewhere along the alveolar ridge\(^7\). Such an activity in the release phase creates a central path between the active and passive articulators, leading to slower outflow of the air, which might result in a percept of a humble voiced noise. In this case, the initial /r/ in Hattori’s description can be regarded as ‘a weak plosive’, rather than ‘a slack stop’.

The idea of ‘a weak plosive’ is developed by Okada (1991). It leads him to a proposal of [d] as a phoneme, which is articulated at the postalveolar place, rather than the retroflex. Okada (1993; 81f.) claims that the nature of the Japanese /r/ (his /ɾ/) lies in its manner, weak plosion or affrication\(^8\). ‘If an alveolar or a dental plosive /d/ is intentionally pronounced with affrication, it will be judged as /ɾ/ by a native-speaker of Japanese; in contrast, if /d/ is articulated at the postalveolar region, [ɾ] (his [ɾ]) and [d] are differentiated either with or without affrication [my translation]’. There are several objections to this. To start with, Okada assumes that the sound written by a letter ᵣ in Japanese is, on the whole, a very weak affricate, and is eager to clearly represent this phonetic nature using the symbol /ɾ/. His spectrograms of [ˈrai] (his [ˈɾai] ‘thunder’) show a release burst but no frication, immediately before the resonance bars for the following vowel. If such acoustic characteristics constitute his proof of ‘a weak affricate’, this is quite unsatisfactory. Okada’s

---

\(^6\) Arnold (1966) explicitly demonstrates that such a separation between articulatory and acoustic properties attains an accurate and comprehensive account of syllabic consonants in British English. Gimson (1980; 157ff. and note 2) also acknowledges the value of the sub-categorisation.

\(^7\) It is possible to assume that ‘a loose-fitting closure’ does not involve a lateral contact on both sides of the palate. Such a pattern is commonly observed in the open vowel context.

\(^8\) Here, the term affrication that is translated from ‘hashatsu [haˌsatsu]’ in Japanese clearly refers to acoustic quality, namely a voiced turbulent noise.
analysis forces us to believe that there must be potential confusion among [dak,kən] ‘rescue’, [rak,kən] ‘optimism’ and [dzak,kən] ‘casual thoughts’, which is not the case in mainstream Japanese. It is the lack of frication after the release burst which significantly represents one of the acoustic properties of ‘tap/flap’ articulation; this characteristic should not be interpreted as ‘incomplete affricate articulation’ or a ‘weak affricate’. We shall return to this point soon.

Furthermore, Okada asserted that ‘while an alveolar tap is frequently observed, it does not usually occur in the word initial position due to the very nature of the sound (82) [my translation]’. Such a definitional statement, or an image of the sound, merely excludes the possibility of the initial occurrence of [r]. Okada fails to account for either what the nature of tap articulation is or how it is distinguished from ‘a very weak affricate’. It might be true that the duration of a hit-and-run movement in initial position may not be so rapid as that in medial position. There might be a chance to associate such a ‘slower’ ballistic tap with affricate articulation. However, articulatory movements and their aerodynamic requirements for the proper affricates, in fact, are not compatible with those of the initial and medial /r/ (his /d/) in Japanese. All we learn from the same introspective experiment described above is that the initial [d] has an explosive release but the initial and the medial [r](his [d]) do not. This observation must be treated carefully.

The lack of frication after the release burst is a promising candidate for distinguishing [r] from [d] on the spectrogram. If we divide the third phase of an articulatory sequence into (articulatory) release and (acoustic) explosion, as Arnold (1966) suggests, then we can make our observation more explicit: the consonant [d] has both release and explosion in the third stage of articulation; in contrast, the consonant [r] has only release and no explosion after it. This suggests that the tap/flap articulation can be considered as having a dual status in its articulatory sequence. On the one hand, the articulation is a kind of stop articulation, as assumed by Laver (1994), in the sense that it involves a complete...
blockage of the airflow in oral cavity. On the other hand, the articulation in question is a kind of approximant (or resonant) articulation in which there is a release of the articulators but no explosion after it, and 'the airflow is non-turbulent when voiced (Catford, 1977; 122)'. More precisely, tap/flap articulation constitutes articulatory activity in the release phase more similar to that in approximant or resonant articulations. Recently, Okada (1999; 118) changes /d/ into /t/ which is formed at the postalveolar region and adds the 'convention' that ‘/t/…mainly occurs medially. Initially and after /N/, it is typically an affricate with short friction, [d*]’. Instead of defining the sound with a different phonetic category, it is more plausible to assume that Okada’s weak affricate is one of the possible realisations of tap/flap articulation.

Let us turn to the second group of the classification, in which the Japanese /r/, both in initial and medial positions, is regarded principally as a lingual tap [r], although various articulations that have been transcribed by different phonetic symbols are also acknowledged. We have already pointed out that the release phase of taps or flaps is significantly different from that of other stops, particularly [d]. The purpose here is to explore into the characteristics of the hold phase, that is, control of tongue movements required for a central articulatory closure.

In his classical but innovative work Edwards (1903; 33f.) mentions, although making no distinction in position, that:

...(r) japonais se forme en appuyant la pointe, ainsi qu’une partie de la face supérieure de la langue contre les alvéoles, plus près des dents que pour (j) anglais, dans une position assez voisine de celle que réclame (d) anglais. Le frottement est très faible en japonais. En prononçant l’(a) anglais la langue prend une position légèrement concave, moins accentuée pour l’(r) japonais. La pointe est un peu recourbée, de sorte que la face inférieure de la langue s’approche des dents. [my italics]

Devant (i) et (j), et quelquefois devant (e), on relève et on avance la langue. La pointe s’approche encore plus de la position exigée par (d) anglais. La langue n’est plus concave, mais la pointe est un peu recourbée vers le palais dur.
Pour l’oreille, cette variété de (r), dont le frottement est minime, ne se distingue guère de la plosive (d): rjo: (dragon), medeto: (proprice), sonnent quelquefois comme djo:, mereto:

It is obvious that, while Edwards labels both Japanese and British English /r/ as lingual fricative in his consonant chart (50f.), he attempts to relate the consonant to /d/, rather than /r/, in British English. Edwards provides examples of speech errors such as [ka,rada]→* [kadara] ‘body’ (his transcription), implicitly suggesting that the two consonants share the equivalent motor actions. Although such a dual status puzzles the phonetic judgment, there is no further discussion on the articulatory nature of Japanese /r/. We can speculate the source of the duality and attempt to integrate it into what we now refer to tap/flap articulation. For this
purpose, we must draw attention to ‘le frottement (literally friction, or the turbulent airflow)’ that is described as weaker in the Japanese /r/ than in the British English /r/. We shall argue that the term ‘le frottement’ is indicative of articulatory action, rather than its acoustic effect.

It has been recognised that there are two major allophones of the phoneme /r/ in British English. One is an approximant realisation such as in red and read. The other is a voiced or voiceless fricative realisation, the conditional allophone that occurs when preceded by [t] or [d] as in tree and dress. Sweet (1877; 38, §109) characterises these two realisations as squeezed /r/ and buzzed /r/ respectively, reflecting their distinctive acoustic quality; ‘...buzzed (r) is felt to be allied to the sibilants, especially (sh), while squeezed (r) is felt to be a weakened form of (rr)’. We are concerned with the approximant realisation, or squeezed /r/.

Jones’ (1960 ) description of the ‘fricative /r/' is a good starting point to seek the psychophysical parameters underlying Edwards’ observation of the Japanese /r/: ‘the most usual English r is a ‘fricative lingual’ sound. It is articulated by the tip of the tongue against the back part of the teeth-ridge, the main body of the tongue being kept low and the ‘front’ being held concave to the palate...and the whole tongue being laterally contracted (194)’. Gimson (1980 ; 205f.) states ‘the back rims of the tongue touching the upper molars, with the central part of the tongue lowered’. Jones distinguishes the fricative [ɹ] from the frictionless continuant [r] that is produced with a slightly wider constriction between the articulators and with less exhaling-force (205). In contrast, Gimson does not refer to the fricative [ɹ] in British English: ‘the most common allophone of RP /r/ is a voiced post-alveolar frictionless continuant (or approximant)’. It can be argued that the ‘fricative’ [ɹ] in Jones’ description refers to the articulatory characteristic of constrictive approximation during the production of [ɹ], rather than the acoustic characteristic of voiced or voiceless turbulent airflow.

There is evidence to suggest that the term ‘fricative, or friction,’ implies something other than the turbulent airflow. In his description of British English /r/, Sweet (1877; 37, §109) states that ‘[t]he characteristic feature of (r) is that the friction passage is formed as much as possible by the tip alone [my italics]’. The ‘friction’ almost certainly refers to the articulatory activities involved in the formation of /r/, since the articulation in question is a squeezed /r/. When Jones (1909; 30, §95) labels both [ɹ] and [z] as a voiced dental fricative, which is similar to what Edwards (1903) did for Japanese, it might be taken for granted that the two sounds unmistakably differ in their acoustic quality, namely the presence or absence of the turbulent airflow. Jones (1973, 108, §352) gives a comment seemingly contradictory to what we cited above from his Outline (1960 ): ‘[f]ricative r (or, alternatively, frictionless continuant r) is the usual sound given by Southern speakers to the written r when a vowel follows [italics mine]’. These descriptions suggest that the ‘fricative’ does not always correspond to the acoustic aspects of speech production that are normally anticipated by the term itself. This situation also indicates the lack of appropriate terminology at that time.
It is entirely fair to say that the term ‘fricative’, or ‘friction,’ was used to represent both the articulatory activities that make the airflow path more restricted and their potential acoustic result that generates the turbulent airflow. Before the term ‘approximant’ first used by Ladefoged (1964), it was the very nature of a given sound that suggested a correct referent in either the articulatory or the acoustic domain. There remains the question of why Jones describes the ‘fricative’ /r/ in addition to the frictionless continuant /r/. One possible answer may be that Jones tacitly intends to distinguish the articulation of the British English /r/ from that of other sounds in a class of rhotics such as taps, flaps or trills, by emphasising ‘fricative’ for the frictionless continuant. A ‘true’ fricative /r/ can be found in South African English as Wells (1982; 616, §8.36) mentions: an obstruent realisation of /r/ with the quality of a tap [r] or a post-alveolar fricative [ʃ], rather than the approximant.

We must now return to the Japanese /r/ that is described by Edwards (1903). We have good reason to suppose that Edwards efficiently refers to the articulatory gesture, ‘constrictive approximation,’ using the term ‘le frottement’. This makes it possible to approach a specific aspect of the articulatory gestures involved in the Japanese /r/. The articulatory activities for [ʃ], as described by Jones and Gimson above, effectively make a hollow on the surface of the tongue as if it were a spoon. It is likely that the tongue body movement is highly constrained in the articulation of the British English [ʃ]. In contrast, the constraints imposed on the Japanese /r/ vary with the phonetic context. Edwards states that the overall tongue shape becomes much less concave in the context of /i/ and /j/. This allows Edwards to judge both constrictive approximation and concave tongue shape for the Japanese /r/ lesser in degree than that in the British English /r/; hence his estimation for the overall concavity and contraction of the tongue. In addition to such a tongue configuration, the presence of the (complete or incomplete) articulatory closure suggests that there are predictable similarities to a voiced apico-alveolar stop consonant in British English. We have already pointed out that the explosion after the release of the articulators is absent in tap/flap articulation. It is also possible to assume that the tip/blade activities effectively constrain the tongue dorsum activities, even if the constriction place on the palate varies with the contextual vowels. This view is consistent with the observation that there is a dip on the front area of the tongue, the articulatory characteristic that is common between [ʃ] and [r] (Jones, 1909* 30, §95; Gimson, 1980 ; 207). We shall return to this point soon.

Uldall (1974) listened to various words containing /r/ spoken by six native speakers of Japanese and inspected the spectrograms of the recordings. The sound is transcribed as a

124 It can be argued whether the fricative quality in south African English [ʃ] is attributed only to a particular segment or to a number of consecutive segments. This is because impressionistically a frication more often extends to the surrounding segments. This suggests that the characteristic can be regarded as one of prosodic features, or, alternatively, breathy voice quality. I am grateful to John Wells for explaining this point. (2000, August, p.c.).
tap/flap with or without a complete articulatory closure, a retroflexed tongue shape, or a longer closure; a very short [d] or [l]. Uldall (1974; 512f.) mentions another source of difficulty in the judgements: 'when a complete closure on the teeth ridge but not at the sides, we perceive sometimes [r] and sometimes [l]. When a tap articulation does not make a complete closure on the teeth ridge, we hear sometimes [r] and sometimes [l], the timing of the very brief sound being appropriate for what we think of as a tap, but an actual closure not always being present'. She further suggests that 'some of the “free variation” is in the perception by Western listeners'. Since the place, point, and duration of articulatory closure vary a great deal, Uldall supposes that those properties are not vital to perception, saying that 'so long as there is a tap articulation and the difference from /d/ is maintained'.

The two, [r] and [d], in Japanese are phonologically contrastive, yielding the following minimal-pairs: ['Uisiukiu'] 'risk' and [' dTsmkiu'] 'disk', ['fe,k^içi'] 'history' and ['de,k^ici'] 'drowning', ['fa,km'] 'easy' and ['da,km'] 'hold', ['fo,km'] 'six' and ['do,km'] 'poison', ['dmm'] 'do' and [' rmm'] 'roux'. Akamatsu’s description of the Japanese /r/ inspired the focus of the EPG experiment in the next chapter. Akamatsu (1997; 105) points out that there are two articulatory patterns of [f] that normally and freely alternate with each other irrespective of the phonetic context. One pattern is that 'the tip of the tongue is quickly raised, poised at a short distance behind the teeth-ridge, then is made to shoot forward and downward to hit (i.e. tap), just once, the...

This observation is helpful particularly when we consider learning and teaching English pronunciation. It is often difficult for non-native speakers to monitor their own pronunciations with confidence. How they sound to a native-speaker’s ear is very important information to assist the learning process. In this connection we may note, in passing, the two observations for Japanese /r/ by Ellis (1874: partly cited in Sweet, 1877) and Sweet (1877).

The symbol (‘r)...means very short (l)...followed by trilled (r). My teacher [a native speaker of Japanese] seemed unable to pronounce (r) with an entirely free tongue. He involuntarily struck the palate first, and although he seemed to remove the tongue immediately, he produced so much of an (l) effect, that the real (r), also very briefly trilled, became obscured. (Ellis, 1874, 1133).

This (r) may be represented by (\{d\}r) or (\{l\}r). An unaccustomed ear hears it as something between (r), (l), and (d). The Japanese pronounce all foreign (l)s and (r)s as this sound, so that when a Japanese says 'a little man,' it sounds to an English ear like 'literally man'. (Sweet, 1877, 85, §244)

From the viewpoint of articulatory gestures, it is likely that the two consonantal articulations, [l] and [r], contrast each other in the control of the tongue dorsum component. Jones (1960; 174, §665) points out that while the tip is touching the teeth-ridge or teeth, the main part is free to take up any position, and in particular it may take up any vowel-position', and describes various colourings of [l] by the notations such as [l̥, l̥', l̥, l̥', l̥̊, l̊'. Hence, the phonetic quality of [l] in British English is usually ‘clear’ before [i] and [j] but ‘dark’ elsewhere (see McCarthy (1944; 122, §441 for median clear [l]). This suggests that the dorsum activity is least constrained to effect various resonance characteristics, and that the physical coupling between the two components is loose. The same, as we will discuss, is not true for the articulation of Japanese [r].
teeth-ridge’. The other is a variant [r] that ‘starts with the tip of the tongue already lightly and momentarily placed against the teeth-ridge (the rims of the tongue on both sides being in contact with the side teeth-ridge), and the tip of the tongue is then made to be blown downwards by outgoing air (106)’. These two patterns, although Akamatsu neglects them, correspond to the distinction proposed by Catford (1977) and Ladefoged & Maddieson (1996): the first [r] described by Akamatsu can be considered as a flap (transient type), whereas a variant [r] can be considered as a tap (flick type). It can be asked whether such a different use of the tongue tip may require different control of the whole body of the tongue.

In a series of cineradiographic tracings of [f] in American English ‘flap’ and Spanish ‘tap’, Monnot & Freeman (1972; 413) have shown that, while the tip of the tongue prepares to go to the alveolar ridge just before the closure for [r] in wa[r]er, the tongue tip in Spanish [r] shows no preparatory activity in the production of Ibé[r]ica. This result leads Ladefoged & Maddieson (1996; 232) to a claim that such an anticipatory activity is characteristic of American English flap. This claim, strictly speaking, is inconclusive. Since different vowels, back and front, precede the target consonants, flap and tap, the tongue dorsum gesture for the front vowel may affect timing control of the tip activity in Spanish tap. Rather, we would like to draw attention to the moment in which [r] is articulated with the tip touching on the alveolar ridge. It is worth noting that the blade and/or the pre- and medio-dorsum of the tongue are depressed with the tongue tip raised in American English [r]. Similarly, during the articulatory closure of the Spanish [r], the tongue dorsum is already shaped for the following high front vowel [i] and only the tip is involved in the contact, the consequence of which is to make the tongue concave. In either case, it is reasonable to assume that, while the two gestures, tap and flap, may be different in their use of the tongue tip, or a retroflex gesture, they functionally aim at making the tongue concave to the palate. In effect, such a gesture makes a hollow on the blade and/or pre-dorsum region of the surface of the tongue.

This assumption of the functional equation between tap and flap articulations is compatible with the description by Edwards (1903) discussed above and is implicitly shared by Akamatsu (1997). Without referring to Uldall (1974), it is claimed that the very presence of tap articulation is crucial for distinguishing [r] from [d] (Akamatsu, 1997; 106). The reason for the claim is that [r] is shorter in duration and has the characteristic hollow on the front of the tongue, the feature missing in [d]. The same point is discussed by Gimson (1980; 207). In his discussion of phonetic variants of British English /r/, Gimson remarks that ‘the articulation of [r] differs from that of /d/ in that the contact for [r] is of shorter duration and less complete than that of /d/, the central hollowing of the tongue typical of [i] being retained’.

A series of EPG experiments has been performed by Sudo et al. (1982, 1983) and
Sawashima & Kiritani (1985). They used the utterance \( V_1 CV_2 V_1 CV_2 \), where the vowels are possible combinations of \( V_1 /i, e, a, o, u/ \) and \( V_2 /i, e, a, o, (u)/ \) and two consonants are either \(/d/s\) or \(/r/s\). There are four major findings: (i) during the production of \(/r/\), there are two types of tongue-palate contact which proceed posterio-anterioirly or vice versa; (ii) articulatory closure is not always attained particularly in \(/a/\) and \(/o/\) contexts, but is most frequently achieved in \(/i/\) context; (iii) \(/r/\) has a larger inter-utterance variability of the two particularly in the anterior region; (iv) the contextual variation between the two consonants is not statistically significant for the total amount of contact. Sawashima & Kiritani (1985) compare the maximum contact pattern and time course between \(/r/\) and \(/d/\) for three speakers. The general trend is that \(/d/\) is wider in contact at the maximum point, has less variation in the anterior portion of the palate and is longer in closure duration than \(/r/\). This tendency, as the authors report, was not observed in all the speakers. The anterior trajectory of \(/r/\) is similar to that of \(/d/\) but an amount of contact in the anterior region is smaller than \(/d/\). These findings support the general characteristics of tap/flap articulation discussed in section 3.5 above. Yet, the study fails to attain either a satisfactory account of the articulatory nature of the Japanese \(/r/\) or the points discussed in our hypothesis above.

Given that tap/flap articulation functionally aims at making a characteristic dip on the surface of the tongue, we can say that the tongue dorsum raising gesture should be controlled properly. This is because the degree of the raising gesture effectively limits the freedom of the anterior activity for making a characteristic hollow. In other words, the raising of the dorsum component is antagonistic to the rapid flipping movement of the tip/blade component. This implies that the physical coupling effects are relatively strong: the two components are not freely co-produced. This hypothesis allows us to predict two kinds of coarticulatory patterning, or resolution of articulatory antagonism, between the gestures for the contextual vowels and those for \([r]\).

One possible result is articulatory blending: the gestures for a given vowel blend with those for the consonantal articulation. Ladefoged (1993; 169) mentions that many speakers of American English use a retroflex flap, rather than an alveolar tap, in a sequence such as hard up. This blending can be explained by the characteristic gestures for \(/r/\) in American English. Its articulation involves three principal gestures such as raising the tongue dorsum towards the mid-palatal region (i.e. bunched), pulling the tongue root towards the pharynx (i.e. pharyngealised), and rounding the lips. In addition, the tongue tip is optionally raised towards the alveolar region, a gesture that is often observed in strong syllables (Dellattre & Freeman, 1968; Ladefoged & Maddieson, 1996). Such articulatory movements for \(/r/\) in the pronunciation of \(/h\alpha:\text{rd} \text{ $\Lambda p$}/\ cause the preceding vowel to become retroflexed, shifting the constriction place of the following consonant \([r]\) (usual realisation in post-stressed position) further back. A whole process can be formulated as \(['h\alpha:\text{d} \text{ $\Lambda p$}]\) or \(['h\alpha:\text{d} \text{ $\Lambda p$}]->
[ˈhɑːɬ t ʌp]→[ˈhɑːɬ ʌp]. Notice here that the articulatory blending in this case is carryover in nature and is presumably motivated in a language with the system of rhoticised vowels in the phonology.

Another possibility is articulatory compounding, in which the contextual vowel gesture and/or the consonantal gesture may be altered in its timing and degree. Since the dorsum raising gesture, particularly for a high front vowel /i/, may not be compatible with the tip/blade gesture involved in either tapped or flapped articulation, the effects may be anticipatory as well as carryover. Such an antagonistic gesture by the tongue is parallel to the degree of coarticulatory resistance in the given consonantal gesture.

### 3.7 Summary and Hypotheses

A brief review of previous descriptive and experimental work was set out with the assumption that the articulation of [t, d, n, ji, r] is stop articulation with or without voicing and aspectual processes, nasal and tap/flap. We have discussed the five consonantal articulations in terms of interplay between the two components of the tongue, tip/blade gesture and dorsum gesture. With such a perspective, we have attempted to approach and exemplify the issues of speaker’s control strategies underlying the given consonantal articulation. It has been demonstrated that our knowledge of this aspect in Japanese is very limited in scope and coverage. There is little empirical evidence for the spatiotemporal coordinative pattern during the production of Japanese [t, d, n, ji, r]. Since those consonants are distinguished each other by their traditional manner of articulation categories in previous studies of Japanese and other languages, it prevents us from approaching conspicuous characteristics shared by the lingual articulatory gesture.

The assumption that the three different manner categories are subsumed into the stop articulation depends on the presence or absence of a complete blockage of the airflow by the active articulator in the oral cavity. Another categorisation is also possible for the consonants [t, d, n, ji, r]. As pointed out in section 3.6, the release phase of ‘tapped/flapped’ stop articulation is not compatible with those of ‘plosive’ stop articulation: the former involves release but no explosion after it; the latter, in contrast, involves those articulatory and acoustic stages. This is also true for ‘nasal’ stop articulation, since the velopharyngeal port is open during the production of the consonant. When we focus on the articulatory event at the release phase, the consonants [t, d] are still stop, or obstruct, articulations, but the consonants [n, ji, r] are resonant articulations.

The evidence from both phonetic judgments by trained phoneticians and measurements by various experimental techniques reveals that speaker’s control strategy in the five consonantal articulations is a complex issue in the speech production mechanisms. There are three remarkable features. First, the two components of the tongue are physically
interconnected but functionally independent during the production of the given consonantal articulation. Second, the coordination between the tip/blade and dorsum components is systematically constrained by a number of factors. As elementary production requirements, voicing and aspectual processes (i.e. nasal and tap/flap) impose specific constraints upon the tongue activities. This suggests that there are consonant-specific control strategies for the two functional divisions of the tongue by a speaker. Furthermore, the freedom of the movements is more restricted, or modified, by language-specific settings on the tongue and by idiosyncratic preferences. Third, there is a potential link between the degree of the physical coupling and the coarticulatory effects, both anticipatory and carryover.

The specific question to be explored in the experiment concerns the factor(s) underlying the unity of stop articulations. We may notice that simple phonetic observation we reviewed implicitly shows a resemblance between the sounds, rather than a substantial divergence. One can safely state that, while each sound has its own distinctive properties, these are in a sense very subtle. Although phonetic symbols typically represent the independent existence of a particular sound, the human judge assumes, or relies on, mutual parameters in both psychological and physical domains for measuring and characterising a sound or a set of sounds. We do not believe that a single invariant phonetic correlation is sufficient information for phonetic categorisation. There are potential candidates for phonetic parameters, the combinations of which allow us to formulate two hypotheses concerning the target consonants \([t, d, n, p, r]\) in Japanese.

First, the stop articulations \([t, d, n, p, r]\) in Japanese constitute a continuum in terms of the contact degree and the closure duration. This hypothesis is an extension of Fujimura et al. (1973). It is generally recognised that the articulatory closure duration is longer in the voiceless alveolar stop \([t]\) than that in the voiced cognate \([d]\). As reviewed in section 3.3.1, Fujimura et al. (1973) claimed that the degree of contact and the duration of articulatory closure were clearly correlated in the production of \(/t/\) and \(/d/\); the longer the articulatory closure, the higher the tongue-palate contact. If such a correlation is regarded as one of the attributes of stop consonant articulation, then the consonant \([n]\) with its distinct aspectual feature must have been considered as having the characteristic. Furthermore, this raises the question about the status of another stop consonant \([r]\); whether it is part of a continuum of the other stops, particularly \([d]\), in Japanese; or whether it is a separate unrelated sound.

Second, if a speaker consciously coordinates the two components of the tongue, their dependency relations are consonant-specifically determined. Our review suggests that the tip/blade activities are effectively connected to the dorsum activity. It has been observed in various languages that the tip/blade gesture allows the tongue body to freely accommodate the vocalic gesture in the production of \([t]\). In contrast, the tongue body movement is relatively restricted in the production of \([r]\). We would like to know whether such a
dependency, or interarticulator coordination, between the two components is linear in nature.

These hypotheses are of relevance in phonetic categorisation. Further examination may provide an articulatory basis for considering a phonetic category, coronal stop consonants. The experiment that follows is an attempt to gain information on these features concerning lingual articulation and coarticulation of [t, d, n, n, r] in Japanese.
Chapter 4

Lingual Articulation and Coarticulation of Coronal Stops [t, d, n, j, f]

4.1 Introduction

An experiment was carried out to investigate the general articulatory and coarticulatory characteristics of coronal stop articulations as a function of the vowel contexts. One of the principal aims, as well as presenting the articulatory and coarticulatory characteristics of the coronal stops in Japanese, was to examine various EPG correlates for the production of [t, d, n, j, f] discussed in Chapter 3. In particular, an attempt was made to identify distinctive articulatory properties that were not explained by the feature specifications. Based on the measurements of the spatiotemporal parameters, speakers' control strategies were inferred and some organisational characteristics for the lingual gestures for the five consonants were explored. The correlations between the following EPG parameters were studied to ascertain: (i) the degree of the front (tip/blade) and the central (medio-dorsum) contacts; (ii) the articulatory duration of the hold phase (i.e. a complete closure); and (iii) the articulatory duration of a consonantal gesture (i.e. the duration of the front region contact detectable by EPG).

In section 4.2 the experimental procedure is described in detail. In section 4.3 the results are presented and discussed under the three headings: the configurational characteristics (4.3.1 and 4.3.2); coarticulatory characteristics (4.3.3); and the durational characteristics of closure and consonantal gesture (4.3.4). In section 4.4 correlations of articulatory parameters are examined. The consonants [t, d, n, j, f] are readily specified by both active and passive articulators that are involved in forming the articulatory closure. We shall argue that the articulatory correlates of linguistically contrastive events are not realised solely from the activities of the tongue tip/blade, but also from those of the tongue dorsum and the systematic interconnections between the two components. Finally, we conclude the chapter with a brief summary in section 4.5.

4.2 Data Collection and Analysis

4.2.1 Speech Items

The speech items used in this experiment were VCV disyllabic (two-mora) words that contained the consonants [t, d, n, j, f] and the vowels /i, a, u/ in all possible combinations. They are mostly nonsense words. Two points must be noted. First, the sequences of /ti, di, tu, du/ in this experiment correspond to [t'j, d'i, t'u, d'u], forms not found in native Japanese
words but common in the pronunciation of loanwords such as ['tlii] ‘tea’, ['paatlii] ‘party’, ['d'inaa] ‘dinner’, ['aid'iai] ‘idea’, ['tutu'] ‘two’, [ta'tutut] ‘tattoo’, [du,uy] ‘do-it-yourself’ and ['çinduuu] ‘Hindu’. Second, the data for [n] is restricted to VCV sequences, the second vowel of which is either /a/ or /u/: this restriction is related to allophonic variation of /n/ (see Chapter 2). Thus, we made the comparisons in the full-vowel contexts for [t, d, n, r] and in non-front-vowel contexts for [n].

4.2.2 Measurements

The analysis was carried out on the acoustic and EPG signals. As shown in Figure 4-1, the surface of the artificial palate was divided into three major regions: front, central, and back. The front region includes dentalveolar, alveolar and postalveolar areas; the central and back regions correspond to palatal and postpalatal areas respectively. The configurational and spatiotemporal parameters were calculated based on these three regions.

![Figure 4-1: Arrangement of Electrodes and Three Regions](image)

Figure 4-2 illustrates schematically the three phases of stop articulation and the points in time selected for quantifying the configurational, spatial, and temporal characteristics of the VCV sequences.

![Figure 4-2: Measurement Points in Time for the EPG Data](image)

Three points in the hold phase, MAX, C1, and C2, were identified from the EPG signal. The point of maximum lingual constriction, MAX, was the EPG frame showing the maximum number of on-electrodes. The point C1 (the closure onset) was defined as the first frame showing full central contact or at least four central contacts in any one row (Hardcastle, 1984). In addition, the frame showing one or more central electrodes within the front region was also regarded as a criterion for identifying the point C1 of the sequences with [r] (Recasens, 1991).
C2 (the closure offset) was the last frame showing articulatory closure; the frame immediately prior to the removal of the central contact. The acoustic recordings were consulted when necessary.

The general configurational properties and their V-to-C coarticulatory effects were analysed for the three regions at the point of maximum lingual constriction (MAX). Maximum linguopalatal contact is indicative of the degree of lingual displacement of a given consonantal articulation. Contact configurations for the target consonants were averaged over five repetitions and were presented in the form of EPG prototypical palatograms that show the frequency of the electrode activation in various shades.

The degree of coarticulatory sensitivity was inferred from the percentage of electrode activation presented by each target consonant at the MAX. The activation values were calculated for the three regions, front, central and back, separately. Since the major constriction of the target consonants was formed in the front region, it is possible, on the basis of the percentage of contact in the other regions, to infer the degree of tongue dorsum elevation and the interplay between the two components of the tongue.

Two measurements of articulatory duration were carried out for a sequence of the three phases in Figure 4-2; one for the hold phase only and the other for the three phases as a whole. The former represents the duration of articulatory closure for a given consonant. It was measured from the temporal interval between the points C1 and C2. The latter, the total duration of the three phases, is referred to as the duration of a consonantal gesture, or the gestural duration. This was measured as duration of contact change in the front region; from the point at which the tongue tip/blade sets in motion for making a closure during V1 to the point at which it ends the releasing movement during V2. This is regarded as the duration of a tip/blade activity for a given consonant.

A common method for this measurement is to visually inspect the EPG printouts and to find out the particular frames corresponding to detectable changes in contact pattern (Butcher & Weiher, 1976; Butcher, 1989; Farnetani et al. 1985). This method is rather crude even if the articulatory measurement is associated with the offset and onset of voicing for the vowels. These acoustic edges are normally parallel with the (EPG) articulatory edges in an open vowel context, but not in a high vowel context, particularly /i/ where the ending point is hardly definable. This means that the tongue has not yet moved away from the hard palate, although the release burst and resumption of formant structure acoustically mark the ‘abrupt’ change from a consonant to a vowel, or vice versa.

In this experiment, therefore, a consistent CV cut-off point was set at 20% of linguopalatal contact. The percentage value was obtained in the following way. The first step was to get the percentage value in the front region activation for the steady state of the preceding and the following vowels, by looking at articulatory trajectories of all the tokens.
and calculating the mean activation value. Then this mean value was modified by the visual inspection of the trajectory data in order to cover all the starting points of contact change in the articulation of the target consonants. Thus, the 20% value indicated a lowest, or minimally deviated, starting (and finishing) point of the changes in tongue-palate contact pattern. The duration of consonantal gesture was defined as the temporal interval above that cut-off point.

A series of two-way and one-way ANOVAS were applied for identifying and isolating the factors affecting the variability of the amount of linguopalatal contact. Two-way ANOVAS used the consonant type and the vowel contexts as the factors. This enabled us to find out whether either of the factors caused a significant difference in the percentage of the electrode activation. The two factors were separately examined in one-way ANOVAS. In addition, pair-wise comparisons by t-test were conducted to identify the significant relation between particular consonant (or vowel) pairs.

4.3 Results and Discussion

4.3.1 Linguopalatal Contact Configurations and Maximum Extensions

The first step in examining the articulatory characteristics of the target consonants was to identify an overall tendency for the variability of the contact amount in the three regions. The percentage values were calculated separately for front, central, and back regions for maximum extension of linguopalatal contact. The contact values in the front region can be read as an index of the central articulatory closure and those in the central and back regions can be interpreted as an index of the effects of the contextual vowels. The values were averaged over five repetitions and are presented for each of the two speakers. Figure 4-3 presents the mean values of the MAX linguopalatal contact for the five target consonants [t, d, n, ji, r] in the vowel symmetrical contexts (see Appendix 1 for the descriptive statistics).

A two-way ANOVA was performed to identify the factors affecting the mean percentage of on-electrodes in each region. For effective comparison, the five target consonants were categorised into two groups, [t, d, n, r] and [t, d, n, p, r], the analyses for which were based on the three vowel contexts (i.e. /i, a, u/) and two vowel contexts (i.e. /a, u/) respectively. The ANOVA results for the group [t, d, n, r] are summarised in Tables 4-1 and those for the group [t, d, n, p, r] are in Table 4-2.

The three regions show different trends for the two factors. As seen in Tables 4-1 and 4-2, the consonant type has significant effects both in the full-context comparison and in the non-front-vowel context comparison. In contrast, the significance of the vowel effects varies with the speaker: this can be seen in Table 4-2 which shows these contrasting results for the central region. Significant interaction between the two factors was found in all the regions except in the central region of speaker MN in the non-front vowel contexts.
Figure 4-3: Mean percentage values of lingualpalatal contact in the three regions by the target consonants for the two speakers: X-axis=the VCV sequences; Y-axis=contact amount.

Table 4-1: Two-way ANOVA results for the effects of the consonant types and the vowel contexts on lingualpalatal contact of [t, d, n, r]

<table>
<thead>
<tr>
<th>Factors</th>
<th>Front</th>
<th>Central</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPEAKER MN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 VOWELS</td>
<td>F(2,48)</td>
<td>3.44**</td>
<td>267.25**</td>
</tr>
<tr>
<td>4 CONSONANTS</td>
<td>F(3,48)</td>
<td>99.38**</td>
<td>269.91**</td>
</tr>
<tr>
<td>INTERACTION</td>
<td>F(6,48)</td>
<td>3.43**</td>
<td>15.45**</td>
</tr>
<tr>
<td>SPEAKER TM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 VOWELS</td>
<td>F(2,48)</td>
<td>1.54</td>
<td>501.74**</td>
</tr>
<tr>
<td>4 CONSONANTS</td>
<td>F(3,48)</td>
<td>207.23**</td>
<td>311.94**</td>
</tr>
<tr>
<td>INTERACTION</td>
<td>F(6,48)</td>
<td>17.65**</td>
<td>20.59**</td>
</tr>
</tbody>
</table>

Table 4-2: Two-way ANOVA results for the effects of the consonant types and the vowel contexts on lingualpalatal contact of [t, d, n, r] in /a_a/ and /u_u/ contexts

<table>
<thead>
<tr>
<th>Factors</th>
<th>Front</th>
<th>Central</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPEAKER MN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 VOWELS</td>
<td>F(1,40)</td>
<td>2.66</td>
<td>2.454</td>
</tr>
<tr>
<td>5 CONSONANTS</td>
<td>F(4,40)</td>
<td>50.35**</td>
<td>265.72**</td>
</tr>
<tr>
<td>INTERACTION</td>
<td>F(4,40)</td>
<td>3.101*</td>
<td>1.54</td>
</tr>
<tr>
<td>SPEAKER TM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 VOWELS</td>
<td>F(1,40)</td>
<td>1.065</td>
<td>121.68**</td>
</tr>
<tr>
<td>5 CONSONANTS</td>
<td>F(4,40)</td>
<td>92.74**</td>
<td>202.19**</td>
</tr>
<tr>
<td>INTERACTION</td>
<td>F(4,40)</td>
<td>15.83**</td>
<td>32.27**</td>
</tr>
</tbody>
</table>

* = p < 0.05; ** = p < 0.01; unmarked = nonsignificant
In the following discussion we discuss the above results in detail and attempt to contrast the consonants in terms of the maximum extension of linguopalatal contact in the three regions, front, central, and back. EPG prototypical palatograms that contain the activation of the electrodes averaged over five repetitions were made for the MAX constriction of each consonant. The palatograms shown here consist of 62 square locations. Each square is shaded and indicates the frequency of activation over five repetitions: 'black' 100-80%; 'grey' 79-60%; 'striped' 59-40%; 'dotted' 39-20%; and 'white' 19-0%. The darker area delimits the relatively stable contact for the production of a given consonant. This allows us to study configurational aspects of the consonantal articulation and possible coarticulatory effects from the surrounding vowels. It also allows us to infer the point of articulation on the tongue. We first consider the configurational and coarticulatory characteristics of [t, d, n, r], and then move on to those of [n] in relation to the other consonants.

4.3.2 The Consonants [t, d, n, r]: an overview

If the activity of the tip/blade component is consonant-specifically determined in the coronal stop articulations, that of the dorsum component is effectively constrained by the consonant type. This assumption is examined by comparing the variability of contact in the symmetrical contexts. In this context the point vowels /i/, /a/, and /u/ impose extreme demands on the control of the tongue body. Speakers, in principle, attempt to keep a higher position for the sequences /iCi/ and /uCu/. In the /aCa/ sequences, on the other hand, the continuity of a lower position may be required. Since the requirement of the dorsum elevation is extreme and specific for each symmetrical context, the contact variability in the posterior two regions across the consonants can be considered as an act of self-maintenance on the given consonantal articulation: the maintenance act is required for successfully producing a distinctive activity of the tip/blade component. In short, the degree of coproduction between the two components of the tongue may vary with the extent by which the given consonantal articulation constrains the dorsum activity.

Table 4-3 summarises the results of the one-way ANOVA showing how the target consonants differ and which vowel is most influential. The consonant types have significant effects on the maximum extension of the tongue-palate contact in all the three regions. We may summarise the general trend in each region across the three vowel contexts by the following three points. First, the amount of the front region contact decreases in the order [n]>[t]>[d]>[r], while the differences between [n] and [t] and those between [d] and [r] tend to be levelled out. Second, in the central region the consonant [r] involves the smallest contact degree of the four consonants. Third, the back region contact of [n] tends to be larger than the others.
Table 4-3: One-way ANOVA Results for the Effects of the Consonant Types and the Vowel Contexts on Linguopalatal Contact of [t, d, n, r]

(a) The Effects of the Consonant Type

<table>
<thead>
<tr>
<th>Speaker MN</th>
<th>Regions and F-Values†‡</th>
<th>F(3,16)</th>
<th>F(3,16)</th>
<th>F(3,16)</th>
<th>F(3,16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/Ci/</td>
<td>Front</td>
<td>64.61**</td>
<td>26.78**</td>
<td>4.8*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central</td>
<td>72.74**</td>
<td>146.27**</td>
<td>41.94**</td>
<td></td>
</tr>
<tr>
<td>/Ca/</td>
<td>Front</td>
<td>14.02**</td>
<td>170.85**</td>
<td>9.83*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/Cu/</td>
<td>Front</td>
<td>114.16**</td>
<td>108.73**</td>
<td>12.24**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central</td>
<td>53.05**</td>
<td>291.09**</td>
<td>68.37**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Back</td>
<td>76.12**</td>
<td>36.02**</td>
<td>7.64*</td>
<td></td>
</tr>
<tr>
<td>/Ci/</td>
<td>Front</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/Ca/</td>
<td>Front</td>
<td>9.70**</td>
<td>175.80**</td>
<td>15.44**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central</td>
<td>19.46**</td>
<td>102.31**</td>
<td>120.00**</td>
<td></td>
</tr>
<tr>
<td>/Cu/</td>
<td>Front</td>
<td>43.00**</td>
<td>1162.66**</td>
<td>53.6**</td>
<td></td>
</tr>
</tbody>
</table>

†**=p<0.05; *=p<0.01; unmarked = non-significant
‡The relation between the consonant pairs was assessed by unpaired t-test; the criterion for the significance was p<0.01. The symbol > indicates a significant difference but the symbol = indicates no significant difference.

The contextual vowels also have significant effects on the proportion of on-electrodes in all the three regions, except the front region of /VtV/ sequences. There is a general trend across the consonants: the contact amount in the three regions becomes greater in the /i/ vowel context than in the /a/ and /u/ contexts. The data for [t], [d], and [n] in the front region agree with this trend. The opposite result, however, is obtained for [r], the greatest contact of which is achieved in the open vowel context. This suggests that the height of the tongue dorsum typical for the close vowels imposes specific constraints on the tip/blade gesture of the tap.

We shall develop our discussion on articulatory and coarticulatory characteristics of each consonantal articulation based on the results presented in Table 4-3 above. The EPG palatograms of [t, d, n, r] are given in Figures 4-4-1 and 4-4-2. Those for [r] will be presented later.
Figure 4-4-1 Linguopalatal Configurations at the MAX Point for /t, d, n/ in the Vowel Symmetrical Contexts /i _ i/, /a _ a/, and /u _ u/ (Speaker MN): 'black' 100-80%; 'grey' 79-60%; 'striped' 59-40%; 'dotted' 39-20%; 'white' 19-0%.
Figure 4-4-2 Linguopalatal Configurations at the Max Point for /t, d, n/ in the Vowel Symmetrical Contexts /i_ i/, /a_ a/, and /u_ u/ (Speaker TM): 'black' 100-80%; 'grey' 79-60; 'striped' 59-40%; 'dotted' 39-20%; 'white' 19-0%.
4.3.2.1 [t] and [d]
The two consonants, [t] and [d], are specified by two parameters: the dental place on the palate, which is the maximum extension of the tongue; and the apicolaminal point on the tongue, which involves both the tongue tip and blade. It can be seen from Figures 4-4-1 and 4-4-2 that the central articulatory closure for [t] and [d] occurs along the alveolar region with complete lateral closure. This central-lateral complete occlusion is always realised and stable across all the vowel contexts. Both the tip and blade of the tongue participate in the formation of the central articulatory closure. It was confirmed in the post-experimental interview that the rim and the tip of the tongue touched the central maxillary incisor teeth. The side contacts for speaker MN, particularly in row 4, extend towards the midsagittal line in all the vowel contexts. In contrast, the side contact is quite variable for speaker TM.

This idiosyncratic difference in the central closure suggests that the laminal constriction is more resistant to the coarticulatory effects from the surrounding vowels. Because of the regular attainment of the side contacts in rows 3 and 4, the differences in the amount of contact for [t] and [d] produced by speaker MN is not significant (see Table 4-3): [t=0.34, p=0.37, df=8] for [iti] vs. [ata]; [t=0.48, p=0.34, df=8] for [iti] vs. [uttu]; [t=1.94, p=0.04, df=8] for [idi] vs. [uduu]; [t=1.41, p=0.09, df=8] for [ada] vs. [udui]. In contrast, there are significant differences for the tokens of speaker TM: [t=3.93, p=0.002, df=8] for [iti] vs. [ata]; [t=4.54, p=0.0009, df=8] for [uduu] vs. [ada].

There is a group trend for [t] and [d] in the front region. Although both consonants involve an apicolaminal closure along the alveolar region, [d] occupies a reduced area: the length of the central constriction is two rows for /d/ and three rows for /t/. Such a characteristic is generally observed for the central and back regions. The lateral contact for [t] largely extends towards the midsagittal line in the /i/ vowel context. This extension is more variable in [d]. These differences could be attributed to the effects of voicing. They are reflected in the overall contact pattern, not only in the degree of the tip/blade contact but also in that of the lateral contact. This generally agrees with the results obtained by a number of studies in various languages, with the notable exception of Dagenais et al. (1994; see section 3.3.1). It appears that, while the dorsum gesture adjusts its height well with that for the contextual vowels in [t], this adjustment is constrained in [d] by the voicing requirement. We will discuss this point further in section 4.4.3.

4.3.2.2 [n]
The consonant that is commonly transcribed by the symbol [n] is characterised as an alveolopalatal nasal [n]. It is articulated with the blade on the alveolar and postalveolar regions, accompanying a large amount of dorsal contact. It is supposed that the closure is made mainly by pressing the posterior half of the blade and some front part of the
anterodorsum against the intersection between the alveolar and postalveolar areas. This tongue gesture is naturally facilitated by directing the tongue tip downwards. It can be seen that the frontmost (row 1) contact tends to be withdrawn, while the posterior contact (rows 5-8) increases significantly. Consequently, the place of primary constriction shifts slightly further back, in contrast to that of [t, d] and [n]. It is likely that the articulation in question may vary from [n^] to [\textipa{n}], the phonetic symbols marking whether there is a tip closure or not.

An alveolopalatal nasal \([n^\textipa{1,2}]\) generally exhibits the largest amount of percentage contact in all the regions. When the contact value was contrasted with that of [t, d, r], the four consonantal articulations were found to be substantially divergent in the contact values in the central region. The involvement of the tongue dorsum decreases in the order [n]>[t]>[d]>[r]. The vowel effects tend to be non-significant for speaker MN (see Table 4-3(b)): \([t=1.89, p=0.04, df=8]\) for \([\textipa{ni}]\) (72.5%) vs. \([\textipa{m\textipa{n}}]\) (67.50%); \([t=1.39, p=0.10, df=8]\) for \([\textipa{a\textipa{n}}]\) (62.50%) vs. \([\textipa{m\textipa{n}}]\). In contrast, significant differences were found for speaker TM: \([t=7.21, p<0.0001, df=8]\) for \([\textipa{ni}]\) (85.00%) vs. \([\textipa{a\textipa{n}}]\) (62.50%); \([t=2.82, p=0.011, df=8]\) for \([\textipa{a\textipa{n}}]\) vs. \([\textipa{m\textipa{n}}]\) (54.16%).

This disparity implies two different, but interrelated, characteristics of the articulation of [n]. First, the consonant actively involves the raising gesture of the tongue dorsum that is certainly used by the vowel articulation. One natural consequence of this is that the influence from the preceding and following vowels is blurred. Thus, the consonant is considered as highly resistant to coarticulatory effects from the surrounding vowels. This is evident in the data of speaker MN. Second, as suggested in Recasens et al. (1993), the consonant readily accommodates the ‘height’ of the contextual vowel: the higher position of the tongue body that is required in the production of [n] sensitively reacts to the lower position of the contextual vowel gesture. These paradoxical qualities, coarticulatory resistance and height-dependent V-to-C coarticulation, are predicted from the articulatory nature of the consonant in question. However, when the variability of contact in the front region, as well as the central and back regions, is considered, it is certain that there are language-specific and idiosyncratic factors involved in the production of [n]. We shall give a more accurate account of the characteristics later in the analysis of coarticulatory activity.

4.3.2.3 \([r]\)

Figure 4-5 illustrates the pattern of linguopalatal contact for [r] at the three points, onset, MAX, and offset of the central articulatory closure. It is specified as a voiced apico-alveolar tap. The tongue tip (and presumably very anterior part of the blade) forms the complete central closure in the alveolar region. Contrary to a common assumption, the amount of contact in the front region is rather large. Although the contact is often made at the frontmost row similar to [t], [d], and [n], the rim and tip of the tongue do not touch the upper teeth. This
was verified in the post-experimental interview. The tongue tip/blade is very slightly retracted and/or retroflexed to become concave to the hard palate.

(a) **Speaker MN**

<table>
<thead>
<tr>
<th>ONSET (C1)</th>
<th>MAX</th>
<th>OFFSET (C2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/iri/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ara/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/uru/</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) **Speaker TM**

<table>
<thead>
<tr>
<th>ONSET (C1)</th>
<th>MAX</th>
<th>OFFSET (C2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/iri/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ara/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/uru/</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4-5** Linguopalatal configurations at the MAX point for /r/ in the vowel symmetrical contexts /i_ i/, /a_ a/, and /u_ u/: ‘black’ 100-80%; ‘grey’ 79-60%; ‘striped’ 59-40%; ‘dotted’ 39-20%; ‘white’ 19-0%.
The consonant [r] demonstrates unusual characteristics. The contact in all the regions across the vowel contexts is the smallest of the five consonantal articulations. One exception was found in the front region of the open vowel context: the degree of contact is larger than, or the same as, that in [d] but smaller than those in [t] and [n].

There are inter-speaker variations in the movement of the tongue tip/blade. It is evident in the /a/ vowel context that the tongue tip moves from row 2-3 through row 1-2 to row 1 (speaker MN, Figure 4-5(a)). This forward movement, however, becomes less clear in the other vowel contexts. This may be due to the tongue dorsum involvement in the contextual vowels: the activity relatively fixes the posterior part of the tongue to the hard palate and consequently constrains the freedom of transient movement of the tongue tip. In contrast, such a ‘rolling’ movement was not observed for the other speaker (speaker TM, Figure 4-4(b)) as the central contact stays on the same row(s) throughout the three points of the articulatory closure: the tip of the tongue flicks once at the anterior part of the alveolar ridge, rather than hits there in passing.

For the anterior regions of [r], the EPG palatograms reveal that the complete central closure is constantly attained at the MAX point by speaker MN but not by speaker TM; closure tends to become incomplete in the /u/ context. It is also evident in Figure 4-3 that, while mean percentage of front contact levels out in the data of MN, it substantially changes in that of TM. These characteristics are statistically significant (see Table 4-3); for speaker MN, the amount of front contact in [r] is greater than or the same as that in [d]. For speaker TM, this is true only in the open vowel context: [ara] (68.66%), [iri] (50.66%), and [ururu] (51.33%); [ara] vs. [ada] (63.33%), [t=2.13, p=0.03, df=8]. In addition, this idiosyncrasy is echoed by the contextual effects. The relations among vowel pairs are non-significant for MN, /a/=/u/=/i/ ([F(2,12)=0.19, p=0.827]) but are significant for TM, /i/>/a/=/u/ ([F(2,12)=43, p<0.0001]).

For the posterior regions of [r], the EPG palatograms indicate coarticulatory patterns similar to [t, d], but different in degree: the dorsum elevation typical for the contextual vowels is the smallest. The full lateral closure is not made throughout the articulatory closure in the open vowel context (TM), while the lateral contact commences after the central closure (MN). The contact degree in the central region is the smallest across the vowel contexts (Table 4-3): the consonant hierarchy is [n]>[t]>[d]>[r] and the vowel hierarchy is [i]>[u]>[a]. This is consistent for both speakers. When this dorsal activity is considered with the above observations of a tip/blade activity, it suggests that the dorsum component may cause a conflict with the tip/blade component in the articulation. One of the consequences manifests as an incomplete articulatory closure in the front region. This view potentially contrasts with the common assumption that an incomplete (or lesser degree of) central closure may be a truly distinctive characteristic (e.g. Connell, 1995). We shall return to this point later.
Closure for [n] in the /a/ and /u/ symmetrical contexts is formed along the alveolar region with complete lateral contact. Both the tongue tip and blade are involved: the consonant is specified as a voiced apico-laminal dental nasal. It can be seen, in the EPG palatograms in Figures 4-4-1 and 4-4-2, that the length of a central closure for [n] in the /a/ context is quite similar to that of [t] in the same vowel environment: the length of the complete occlusion extends from row 1 to 3. In contrast, the central closure for [n] in the /u/ context is reduced by one row to become closer to that for [d]. It appears that, while the length of apicolaminal closure for the two plosives varies with the presence and absence of voicing, the apicolaminal closure for the nasal depends on the height of the vowels, /a/ vs. /u/.

Table 4-4 presents the ANOVA results comparing [n] with the other consonants in the two vowel contexts. In the front region, the contact value is similar to that of two plosives, suggesting that the nasal consonant [n] exhibits a lingual gesture much more similar to [t] and [d] than to [n]. In the central and back regions there is a different trend depending on the contextual vowels. The effect of the /u/ vowel is reflected in the back region, yielding the relation /a/</u/.

**Table 4-4: One-Way ANOVA Results for the Effects of the Consonant Types and the Vowel Contexts on Lingual Palatal Contact of [t, d, n, r]**

(a) The Effects of the Consonant Types

<table>
<thead>
<tr>
<th>SPEAKER MN</th>
<th>FRONT (F(4,20))</th>
<th>CENTRAL (F(4,20))</th>
<th>BACK (F(4,20))</th>
</tr>
</thead>
<tbody>
<tr>
<td>/aCa/</td>
<td>n&gt;t=n&gt;r&gt;d 77.69**</td>
<td>p&gt;t=n&gt;d&gt;r 125.40**</td>
<td>n&gt;t=d&gt;r 41.08**</td>
</tr>
<tr>
<td>/uCu/</td>
<td>t&gt;n&gt;r=d 11.86**</td>
<td>p&gt;t=n&gt;d=r 141.38**</td>
<td>p&gt;t=d=n=r 9.50**</td>
</tr>
</tbody>
</table>

(b) The Effects of the Vowel Contexts

<table>
<thead>
<tr>
<th>SPEAKER MN</th>
<th>FRONT (t (df=8))</th>
<th>CENTRAL (t (df=8))</th>
<th>BACK (t (df=8))</th>
</tr>
</thead>
<tbody>
<tr>
<td>/VtV/</td>
<td>a=u 1.00</td>
<td>a=u 0.00</td>
<td>a&lt;u 6.00**</td>
</tr>
<tr>
<td>/VdV/</td>
<td>a=u 1.41</td>
<td>a=u 1.00</td>
<td>a&lt;u NA</td>
</tr>
<tr>
<td>/VpV/</td>
<td>a=u 2.15</td>
<td>a=u 1.39</td>
<td>a=u 2.12</td>
</tr>
<tr>
<td>/VrV/</td>
<td>a=u 0.14</td>
<td>a&lt;u 12.79**</td>
<td>a&lt;u 7.60**</td>
</tr>
<tr>
<td>/VnV/</td>
<td>a=u 4.80**</td>
<td>a=u 1.00</td>
<td>a&lt;u NA</td>
</tr>
</tbody>
</table>

The relation between the consonant pairs was assessed by unpaired t-test; the criterion for the significance was p<0.01. The symbol > indicates a significant difference but the symbol = indicates no significant difference.
Apart from the consonantal articulation, the idiosyncratic differences in the production of the vowel \([\text{ui}]\) are clearly reflected in the effects of the contextual vowels in Table 4-4(b). For speaker MN there is no significant difference in degree between /a/ and /u/ in the central region: one exception is the case of \([r]\) and this is explained by the production requirements discussed above. In contrast, for speaker TM, the degree of contact in the central region is consistently larger in the /u/ vowel: the non-significant difference in \([n]\) is due to its active involvement of the tongue dorsum for making a primary constriction. Japanese \([\text{ui}]\) is typically articulated, with the lips compressed horizontally, at a more central area in the phonetic vowel space. The results suggest that the constrictive approximation is formed slightly further forward for speaker TM than speaker MN.

Informal listening to the actual pronunciation and tape recordings suggests that the vowel produced by speaker TM sounds much more fronted than that of speaker MN. This is particularly evident in the utterance of /utu/, which may be impressionistically transcribed as [tuttü–urti]. Such a variation in quality reflects the medio-dorsum activities involved in the vowel production and this results in a centralised quality being created after the release of the consonantal closure. Because the EPG data in other experiments shows the contact only at both sides of row 7 and 8, the constriction place for \([\text{ui}]\) is readily variable with the type of the consonantal articulation.

4.3.3 Anticipatory and Carryover Effects
Anticipatory and carryover effects were analysed for the three regions, front, central and back at the MAX point, comparing contact patterns and the amount of activated electrodes that were derived from specific contexts. The anticipatory effects were analysed for \(V_1CV_2\) sequences where \(V_1\) was fixed to one of the three vowels /i, a, u/. A MAX contact pattern of /ita/, for instance, is considered as the anticipatory influence of \(V_2/a/\) on /t/ in the context of \(V_1/i/\). The carryover effect, in contrast, was examined for \(V_1CV_2\) sequences where \(V_2\) was fixed to one of the three flanking vowels /i, a, u/. For example, a string of /atu/ is regarded as the carryover effect of \(V_1/a/\) on /t/ in the context of \(V_2/u/\).

4.3.3.1 Coarticulatory Activities in \([t, d, n, ji, r]\): an Overview
Figures 4-6-1 and 4-6-2 indicate the anticipatory effects upon \([t, d, n, ji, r]\), while Figures 4-7-1 and 4-7-2 indicate the carryover effects (see Appendix 2 for the descriptive statistics). A series of two-way ANOVAS were conducted, in which the vowel contexts were coded as anticipatory or carryover environment. The factors analysed for this examination are the consonant type and the three flanking vowels. Similarly to the comparison made for the symmetrical sequences in the previous section, the consonants are divided into two groups: \([t, d, n, r]\) and \([t, d, n, ji, r]\). The results are summarised in Tables 4-5 and 4-6 respectively.
Figure 4-6-1: Anticipatory Effects in the Three Regions for the Consonants [t, d, n, j, r] Produced by Speaker MN
X-axis: the VCV sequences containing the five target consonants
Y-axis: measures of the contact amount at the MAX point (%)
**FIGURE 4-6-2: ANTICIPATORY EFFECTS IN THE THREE REGIONS FOR THE CONSONANTS [t, d, n, j, r] PRODUCED BY SPEAKER TM**

X-axis: the VCV sequences containing the five target consonants

Y-axis: measures of the contact amount at the MAX point (%)
Figure 4-7-1: Carryover Effects in the Three Regions for the Consonants [t, d, n, j, r] Produced by Speaker MN
X-axis: the VCV sequences containing the five target consonants
Y-axis: measures of the contact amount at the MAX point (%)
FIGURE 4-7-2: CARRYOVER EFFECTS IN THE THREE REGIONS FOR THE CONSONANTS [t, d, n, j, r]
PRODUCED BY SPEAKER TM
X-axis: the VCV sequences containing the five target consonants
Y-axis: measures of the contact amount at the MAX point (%)
<table>
<thead>
<tr>
<th>Factors and Vowel Contexts</th>
<th>Anticipatory</th>
<th>Carryover</th>
</tr>
</thead>
<tbody>
<tr>
<td>/iCV/</td>
<td>REGIONS AND F-VALUES</td>
<td>REGIONS AND F-VALUES</td>
</tr>
<tr>
<td>3 Vowels F(2,48)</td>
<td>22.48**</td>
<td>14.26**</td>
</tr>
<tr>
<td>4 Consonants F(3,48)</td>
<td>115.46**</td>
<td>111.97**</td>
</tr>
<tr>
<td>Interaction F(6,48)</td>
<td>11.67**</td>
<td>1.88</td>
</tr>
<tr>
<td>/uCV/</td>
<td>FRONT</td>
<td>CENTRAL</td>
</tr>
<tr>
<td>3 Vowels F(2,48)</td>
<td>13.01**</td>
<td>48.46**</td>
</tr>
<tr>
<td>4 Consonants F(3,48)</td>
<td>82.13**</td>
<td>231.77**</td>
</tr>
<tr>
<td>Interaction F(6,48)</td>
<td>2.20</td>
<td>5.23**</td>
</tr>
<tr>
<td>/aCV/</td>
<td>FRONT</td>
<td>CENTRAL</td>
</tr>
<tr>
<td>3 Vowels F(2,48)</td>
<td>1.70</td>
<td>112.50**</td>
</tr>
<tr>
<td>4 Consonants F(3,48)</td>
<td>72.90**</td>
<td>522.10**</td>
</tr>
<tr>
<td>Interaction F(6,48)</td>
<td>2.09</td>
<td>28.64**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factors and Vowel Contexts</th>
<th>Anticipatory</th>
<th>Carryover</th>
</tr>
</thead>
<tbody>
<tr>
<td>/iCV/</td>
<td>REGIONS AND F-VALUES</td>
<td>REGIONS AND F-VALUES</td>
</tr>
<tr>
<td>3 Vowels F(2,48)</td>
<td>10.34**</td>
<td>17.24**</td>
</tr>
<tr>
<td>4 Consonants F(3,48)</td>
<td>304.76**</td>
<td>286.86**</td>
</tr>
<tr>
<td>Interaction F(6,48)</td>
<td>5.20**</td>
<td>4.86**</td>
</tr>
<tr>
<td>/uC/</td>
<td>FRONT</td>
<td>CENTRAL</td>
</tr>
<tr>
<td>3 Vowels F(2,48)</td>
<td>9.90**</td>
<td>24.18**</td>
</tr>
<tr>
<td>4 Consonants F(3,48)</td>
<td>258.57**</td>
<td>192.11**</td>
</tr>
<tr>
<td>Interaction F(6,48)</td>
<td>8.92**</td>
<td>4.18**</td>
</tr>
<tr>
<td>/aCV/</td>
<td>FRONT</td>
<td>CENTRAL</td>
</tr>
<tr>
<td>3 Vowels F(2,48)</td>
<td>3.01</td>
<td>375.42**</td>
</tr>
<tr>
<td>4 Consonants F(3,48)</td>
<td>81.68**</td>
<td>874.66**</td>
</tr>
<tr>
<td>Interaction F(6,48)</td>
<td>5.12**</td>
<td>41.71**</td>
</tr>
</tbody>
</table>

†= p<0.05; **= p<0.01; unmarked = nonsignificant
### Table 4-6: Two-Way ANOVA Results for Anticipatory and Carryover Effects on [t, d, n, j, r] in /VCa/ and /VCu/ Sequences

#### (a) Anticipatory Effects

<table>
<thead>
<tr>
<th>Speaker Mn</th>
<th>Regions and F-Values†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>iCV</td>
</tr>
<tr>
<td></td>
<td>Front</td>
</tr>
<tr>
<td>2Vowels</td>
<td>F(1,40)</td>
</tr>
<tr>
<td></td>
<td>**</td>
</tr>
<tr>
<td>5Consonants</td>
<td>F(4,40)</td>
</tr>
<tr>
<td></td>
<td>**</td>
</tr>
<tr>
<td>Interaction</td>
<td>F(4,40)</td>
</tr>
<tr>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>

#### (b) Carryover Effects

<table>
<thead>
<tr>
<th>Speaker Mn</th>
<th>Regions and F-Values†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VCi</td>
</tr>
<tr>
<td></td>
<td>Front</td>
</tr>
<tr>
<td>3Vowels</td>
<td>F(2,60)</td>
</tr>
<tr>
<td></td>
<td>**</td>
</tr>
<tr>
<td>5Consonants</td>
<td>F(4,60)</td>
</tr>
<tr>
<td></td>
<td>**</td>
</tr>
<tr>
<td>Interaction</td>
<td>F(8,60)</td>
</tr>
<tr>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speaker Mn</th>
<th>Regions and F-Values†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Front</td>
</tr>
<tr>
<td>3Vowels</td>
<td>F(2,60)</td>
</tr>
<tr>
<td></td>
<td>**</td>
</tr>
<tr>
<td>5Consonants</td>
<td>F(4,60)</td>
</tr>
<tr>
<td></td>
<td>**</td>
</tr>
<tr>
<td>Interaction</td>
<td>F(8,60)</td>
</tr>
<tr>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>

†*= p<0.05; **= p<0.01; unmarked = nonsignificant
It can be seen from Tables 4-5 and 4-6 that the effects of the consonant type in the three regions are consistently significant throughout the two groups of comparisons. This generally supports the hypothesis that there are consonant-specific articulatory controls for the two functional division of the tongue, both the tip/blade and the dorsum components. The effects of the flanking vowels, in contrast, switch their significance depending on the particular fixed vowel and region. In the group of [t, d, n, r], Table 4-5, where all combinations of the three vowels are used as the fixed and flanking vowels, the significant interactions are observed in all the regions of the anticipatory and carryover contexts. The exceptions for this trend are found in the anticipatory contexts: the front region of /aCV/ sequences and the central region of /iCV/ sequences (speaker MN); the back region of /iCV/ and /uCV/ sequences (speaker TM). In the group of [t, d, n, r], Table 4-6, where the V2, either as a fixed or a changing vowel, is restricted to /a/ and /u/, the interaction between the two factors tends to be non-significant in the anticipatory contexts. In the carryover contexts, in contrast, the interaction is found to be significant in all three regions. This suggests that the V-to-C carryover effects are more consistent and extensive than the anticipatory effects upon a given consonantal articulation.

In the following sections further comparisons were made separately for [t, d, f] and [n] in the three-vowel contexts and for [n] in the non-front-vowel contexts. The EPG palatograms at the three temporal intervals are presented. These are paralleled in the particular combinations of the vowel contexts: /aCi/ and /iCa/, /aCu/ and /uCa/, and /iCu/ and /uCi/. This parallel representation effectively indicates the changing patterns of linguopalatal contact as a function of a particular vowel context. When we compare /aCi/ with /iCa/ sequences we notice that their temporal evolutions of contact, particularly in the posterior region, are portrayed as a mirror image. This is also true for the other two combinations, namely /aCu/ and /uCa/, and /iCu/ and /uCi/. However, it must be noted that the nature and extent of the contact pattern are specific to a given consonantal articulation.

4.3.3.2 [t, d, r]
Figures 4-8-1 to 4-10-2 illustrate the EPG palatograms of [t, d, r] that were measured at the three temporal points, namely the onset, the MAX, and the offset of the articulatory closure. It can be found that the two plosives constantly attain a complete articulatory closure at the dentalveolar and alveolar regions (rows 1 to 3). The constriction length is always wider in [t] than in [d]. The tap, in contrast, shows different tip/blade gestures for forming a central closure. Both flicking and transient movements are interchangeably used by the two speakers. Whereas a full closure along the alveolar region is regularly made at the MAX point by speaker MN, a closure tends to be incomplete in the utterances made by speaker TM.
Figure 4-8-1: [t, d, r] at the three temporal points of articulatory closure in /aCi/ and /iCa/ contexts (speaker MN): ‘black’ 100-80%; ‘grey’ 79-60%; ‘striped’ 59-40%; ‘dotted’ 39-20%; ‘white’ 19-0%.
Figure 4-8-2: [t, d, r] at the Three Temporal Points of Articulatory Closure in /aCi/ and /iCa/ contexts (Speaker TM): 'black' 100-80%; 'grey' 79-60%; 'striped' 59-40%; 'dotted' 39-20%; 'white' 19-0%.
Figure 4-9-1: [t, d, r] at The Three Temporal Points of Articulatory Closure in /aCu/ and /uCa/ contexts (Speaker MN): ‘black’ 100-80%; ‘grey’ 79-60%; ‘striped’ 59-40%; ‘dotted’ 39-20%; ‘white’ 19-0%.
Figure 4-9-2: [t, d, r] AT THE THREE TEMPORAL POINTS OF ARTICULATORY CLOSURE IN /aCu/ AND /aCa/ CONTEXTS (SPEAKER TM): 'black' 100-80%; 'grey' 79-60%; 'striped' 59-40%; 'dotted' 39-20%; 'white' 19-0%.
Figure 4-10-1: [t, d, n] At The Three Points of Articulatory Closure In /iCi/ AND /uCi/

           CONTEXTS (SPEAKER MN): 'black' 100-80%; 'grey' 79-60%; 'striped' 59-40%; 'dotted' 39-20%;
           'white' 19-0%.
FIGURE 4-10-2: [t, d, r] AT THE THREE POINTS OF ARTICULATORY CLOSURE IN /\Cu/ AND /\Gi/ CONTEXTS (SPEAKER TM) 'BLACK' 100-80%; 'GREY' 79-60%; 'STRIPED' 59-40%; 'DOTTED' 39-20%; 'WHITE' 19-0%.
A one-way ANOVA was conducted to test for differences in contact degree as a function of the flanking vowels in the anticipatory and carryover contexts. The comparisons were made for the consonant types and for the three flanking vowels in each of the three regions. The results are shown in Table 4-7 below.

**Table 4-7: One-way ANOVA Results for Anticipatory and Carryover Effects on [t, d, r] at the MAX Lingualpalatal Contact**

(a) **Anticipatory Effects**

<table>
<thead>
<tr>
<th>[Consonant Type]</th>
<th>Regions and F Values†</th>
<th>Speaker MN</th>
<th>Front F(2,42)</th>
<th>Central F(2,42)</th>
<th>Back F(2,42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ICV/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>t&gt;d=r</td>
<td>25.58*</td>
<td>t&gt;d=r</td>
<td>39.08*</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>t&gt;r=d</td>
<td>56.33*</td>
<td>t=d=r</td>
<td>6.80*</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>t&gt;r=d</td>
<td>86.14*</td>
<td>t&gt;d=r</td>
<td>9.86*</td>
</tr>
<tr>
<td>/UCV/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>t&gt;d=r</td>
<td>179.73*</td>
<td>t&gt;d=r</td>
<td>117.33*</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>t&gt;r=d</td>
<td>89.04*</td>
<td>t&gt;d=r</td>
<td>30.37*</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>t&gt;r=d</td>
<td>56.26*</td>
<td>t&gt;d=r</td>
<td>12.51*</td>
</tr>
<tr>
<td>/ACV/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>t&gt;d=r</td>
<td>6.62*</td>
<td>i=a=u</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>t&gt;r=d</td>
<td>2.27</td>
<td>i&gt;a=u</td>
<td>24.97*</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>t&gt;a=r</td>
<td>0.60</td>
<td>i&gt;u=a</td>
<td>21.70*</td>
</tr>
</tbody>
</table>

(b) **Carryover Effects**

<table>
<thead>
<tr>
<th>[Consonant Type]</th>
<th>Regions and F Values†</th>
<th>Speaker MN</th>
<th>Front F(2,42)</th>
<th>Central F(2,42)</th>
<th>Back F(2,42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/VCi/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>t&gt;d=r</td>
<td>49.82*</td>
<td>t&gt;d=r</td>
<td>28.35*</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>t&gt;r=d</td>
<td>33.80*</td>
<td>t&gt;d=r</td>
<td>4.04**</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>t&gt;r=d</td>
<td>64.37*</td>
<td>t&gt;d=r</td>
<td>3.47**</td>
</tr>
<tr>
<td>/VCu/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>t&gt;d=r</td>
<td>49.23*</td>
<td>t&gt;d=r</td>
<td>36.22*</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>t&gt;r=d</td>
<td>39.68*</td>
<td>t&gt;d=r</td>
<td>10.88*</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>t&gt;r=d</td>
<td>43.62*</td>
<td>t&gt;d=r</td>
<td>11.06*</td>
</tr>
<tr>
<td>/VCa/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>u=i=a</td>
<td>0.72</td>
<td>i&gt;u=a</td>
<td>7.17**</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>u=i=a</td>
<td>1.60</td>
<td>i&gt;u=a</td>
<td>49.10**</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>i=a=u</td>
<td>2.50</td>
<td>i&gt;u=a</td>
<td>71.93**</td>
</tr>
</tbody>
</table>

(Three Flanking Vowels)
In the front region the effects of the consonant type overrule the variability of the amount of tongue-palate contact both in the anticipatory and carryover contexts. The contact amount of [t] is consistently larger than that of [d] or [r] in both the anticipatory and carryover contexts: a common hierarchy is [t]>[d]>[r]. The difference between [d] and [r] tends to be non-significant for speaker MN; [r] usually has the smallest value for speaker TM. In contrast, the effects of the flanking vowels are non-significant. One exception for this is the /iCV/ sequence produced by speaker MN: [F(2,42)=6.628, p=0.003]. However, the different amounts of contact caused by changing the V₂ turn out to be statistically non-significant.

The front contact of the consonant [r] exhibits the highest variability of the three target consonants. The mean percentage across the asymmetrical contexts is 70% (ranging from 60.00% to 78.00%) for speaker MN, while it is 56.12% (ranging from 35.33% to 72.67%) for speaker TM. The activation value of [r] is higher for speaker MN and tends to be statistically compatible with that of [d]. This tendency is found only in the sequences with the /a/ vowel for speaker TM. It appears that the dorsum gesture for the open vowel causes little interference to the tip/blade gesture. Furthermore, speaker MN regularly forms a full closure along the alveolar region in all the phonetic contexts; and speaker TM, in contrast, tends to miss a complete occlusion when high vowels, either /i/ or /u/, are included in a VCV sequence (see Figures 4-8-2, 4-9-2 and 4-10-2). These differences, which were also observed in symmetrical contexts (see Figure 4-5), suggest that speakers employ idiosyncratic strategies to resolve the antagonistic gestures of the tongue (dorsum).

In the central region, in contrast, both factors, the consonant type and the flanking vowels, affect variability in the degree of linguopalatal contact. Each consonantal articulation involves distinctive control parameters for the tongue pre- and medio-dorsum. This, in turn, constrains the elevation of the tongue body specific to the contextual vowel articulation. The patterns, [t]=[d]>[r] and [i]>[u]=[a], are commonly found for both anticipatory and carryover contexts.

There are two remarkable differences between the two speakers. One is that the distinction between [d] and [r] tends to be levelled out in the sequences /uCV/ and /VCu/ for speaker MN. The other is that the difference between /u/ and /a/ tends to be significant for speaker TM, namely /u/>/a/. These differences may partly be due to the individual feature observed before. The former tendency for the relation between [d] and [r] can be accounted for if we assume that the tongue dorsum activity for the vowel /u/ provides good support for, and less interference to, the tip/blade activity. In the production of /u/ the tongue postdorsum is raised towards and is pushed laterally against the intersection between the hard palate and soft palate (row 8). Such local contact anchors one end of the tongue body to the palate, giving a stable base to the flicking (or transient) movement of the anterior portion of the tongue. The tip and the very anterior part of the blade create the appropriate contact by hitting
the alveolar ridge. Consequently, the overall activity involves the natural amount of contact in the central region. This interpretation assumes that, in the production of [r], the activity in the back region causes less interference to the front region activity than the activity in the central region. In addition, it is likely that the regular attainment of complete closure for [r] (speaker MN) gives support to the dorsum movement of the tongue. This in effect may permit greater freedom to accommodate the preceding and following vowel articulations. The latter tendency for the relation between /u/ and /a/ is related to the more frontal realisation of [u] by speaker TM, a characteristic that we discussed before.

Both anticipatory and carryover effects of the flanking vowels become indistinct when the sequence contains /i/ as the fixed vowel. Farnetani et al. (1985) report similar results for Italian /iCV/ sequences. In the production of /iCV/ sequences the fixed vowel /i/ masks the potential contact variation that is derived by the changing vowels. The anticipatory effects of the vowels are statistically non-significant, /i//= /a//= /u/; for speaker MN, /i/ (55.83%) vs. /a/ (54.72%), [t=0.31, p=0.37, df=28]; /a/ vs. /u/ (49.02%), [t=1.78, p=0.04, df=28]; for speaker TM, /i/ (62.5%) vs. /a/ (55.55%), [t=1.02, p=0.15, df=28]; /a/ vs. /u/ (50.55%), [t=0.73, p=0.23, df=28]. Similarly, in the sequence /VCi/ for speaker TM, the fixed V2 /i/ overrides the carryover effects of the flanking vowels, /u/ and /a/: the pattern is /i/> /u//= /a/: /i/ (62.50%) vs. /u/ (45.27%), [t=3.29, p<0.001, df=28]; /u/ vs. /a/ (42.22%), [t=0.65, p=0.25, df=28]. These results indicate that anticipatory effects are blocked and carryover effects are weakened when they are measured at the point of maximum constriction. This can be explained by the location-to-shape relationship for the vowel articulation. The (pre- and) medio-dorsum of the tongue, which is specified for a high front vowel, is not extensively involved in the formation of the two vowels, /a/ and /u/. While the raising gesture for /i/ overlaps with the lowering gesture for /a/, it readily accommodates with the vowel /u/, the gesture of which involves the raising of the tongue medio-dorsum.

In the back region, similar to the results in the central region, both the consonant type and the flanking vowels reveal significant effects on the different amounts of tongue-palate contact. One exception is the effects of the flanking vowels in the /iCV/ sequences in the anticipatory context: as we discussed above, the preceding high front vowel overshadows the anticipatory effects for the following /a/ and /u/. The patterns, [t]=[d]>[r] and [i]>[u]>[a], are again predominant both in anticipatory and carryover contexts.

In summary, the overall results that correspond to the tongue body activities for the production of /t, d, r/, as a function of context, demonstrate that the consonant type generally overrules the variation for linguopalatal contact. The relation between the consonants and between the flanking vowels varies with the given region and the type of the fixed vowel. The front region reflects the consonant effects more than the vowel effects, yielding the recurring patterns /t/> /d/> /r/ and /i//= /u//= /a/. Yet, both the preceding and following vowels tend to
affect /r/ in particular. The central and back regions are influenced both by the consonant type and by the flanking vowels. Consequently, the hierarchical relations turn out to be /t/=/d/>/r/ and /i/>/u/>/a/. These hierarchies are shared both in the anticipatory context and in the carryover context, and are influenced by the types of the fixed vowels. When an utterance contains the high front vowel /i/, either as a fixed or a flanking vowel, the tongue body is kept high and makes ineffective the appearance of /a/ and /u/ effects. The statistical results prove that carryover effects are more extensive and systematic than anticipatory effects. This confirms the fact that there are consonant-specific control strategies for the two functional divisions of the tongue.

4.3.3.3 [ŋ]

Figures 4-11-1 and 4-11-2 show the EPG palatograms at the three temporal intervals during the production of [ŋ]. The contact increases from back to front but decreases from front to back along the alveolar ridge. This observation becomes explicit when the tongue configuration, both at the onset and at the offset of the articulatory closure, is considered. The central articulatory closure at the MAX point extends from row 1 to row 5. However, the anterior and posterior ends of this extension are variable depending on the vowel context and speaker.

Table 4-8 summarises the results of a one-way ANOVA. The comparisons were made for all combinations of the three point vowels. The results indicate that the vowel-dependent effects in the anticipatory contexts tend to be non-significant for the three regions. This characteristic is generally observed in the carryover contexts for speaker MN but not for speaker TM. This divergence of significant vowel-dependent effects is due to the nature of the two coarticulatory effects and to the potential variability of the constriction location mentioned above.

In the front region the opposite results are observed for the significant vowel-dependent effects: the sequences with /i/ tend to be significant for speaker MN but are non-significant for speaker TM: for speaker MN, /i/ (98.00%) vs. /u/ (92.66%) [t=5.05, p<0.0001, df=8]; /i/ vs. /a/ (86.66%), [t=2.71, p=0.013, df=8]; for speaker TM, /i/ (85.00%) vs. /a/ (80.00%), [t=1.43, p=0.09, df=8]; /i/ vs. /u/ (85.83%), [t=0.25, p=0.40, df=8]. Although a similar tendency was observed for [t, d, r], it differs in its nature as the anticipatory and carryover effects in the other two contexts are substantially divergent between the two speakers. The EPG palatograms suggest that, while speaker MN tends to withdraw the frontmost contact when the sequence contains the open vowel, speaker TM tends to maintain that contact in the same situation.
Figure 4-11-1: [n] AT THE THREE TEMPORAL POINTS OF ARTICULATORY CLOSURE IN THE VOWEL ASYMMETRICAL CONTEXTS (SPEAKER MN): ‘black’ 100-80%; ‘grey’ 79-60%; ‘striped’ 59-40%; ‘dotted’ 39-20%; ‘white’ 19-0%.
Figure 4-11-2: [p] at the three temporal points of articulatory closure in the vowel asymmetrical contexts (Speaker TM): 'black' 100-80%; 'grey' 79-60%; 'striped' 59-40%; 'dotted' 39-20%; 'white' 19-0%.
The variability in the tip participation can also be accounted for when we consider language-specific and speaker-specific production characteristics. First, the primary constriction of the Japanese [n] is formed slightly further forward than a palatal nasal in other languages. This characteristic becomes clear if compared with the Catalan [n]. As discussed in Chapter 3, Recasens (1990) claims that the tip and dorsum mechanically, or automatically, take part in the constriction on the assumption that the more restricted area on the tongue, namely the tongue pre-dorsal region, is actively controlled. The Japanese [n] is described by the same place label, alveolopalatal, as the Catalan [n], but the two [n]s differ in both the place on the palate and the point on the tongue. In Recasens et al. (1993) the EPG palatograms (the Reading System) of three informants show that, while the posterior end of the central contact extends to row 5, the anterior end does not reach more than row 2. All the variations, both individual and contextual, occur within this range, namely row 2 to row 5. Thus, Recasens et al. (1993, 222) assert that the Catalan [n] is formed by pressing the back of the blade and/or the predorsum against the postalveolar and prepalatal regions. In contrast, the anterior and posterior ends of the central closure in the Japanese [n] spread out from row 1 (without the tip contact at the upper teeth) to row 4 at most. There is a language-specific range difference in the central articulatory closure between Catalan (rows 2-5) and Japanese (rows 1-4). We can identify the Japanese [n] as a palatalised alveolar, but the Catalan [n] as an alveolarised palatal. It seems reasonable to suppose that the Japanese [n] has its centre of articulatory gravity slightly further forward along the surface of the tongue. This helps to
explain the variable contact at row 1 in the Japanese [n]. In addition, it is likely that the vowel context may cause a specific influence on both the tip/blade and dorsum components.

The second point about the high variability of tip contact is that the actual realisation, or more precisely the ‘focal’ location of a complete occlusion, differs considerably between the two speakers. It is possible to infer idiosyncratic control strategies by inspecting the black area of the EPG palatograms at the onset and offset of the articulatory closure. The central articulatory closure common to all the vowel contexts is formed primarily at rows 3-4 for speaker MN. For speaker TM, in contrast, the typical closure is made slightly further forward, namely at rows 2-3. It is reasonable to suppose, given that the two components of the tongue are interconnected, that the more fronted realisation effectively reduces the freedom for the tip gesture and, in turn, provides greater freedom for the dorsum gesture. The more retracted realisation, on the other hand, may produce the reverse effects on each component. These considerations, together with the first point above, are based on potential synergetic relations between the two components of the tongue. This synergism, as expected, provides a key to understand the next point, namely the variability in the central region contact.

In the central region the anticipatory effects of the flanking vowels are largely identical for both speakers. This is also found in the carryover effects for speaker MN. These tendencies are explained naturally by the fact that the articulation of an alveolopalatal [n] involves the same zone on the tongue as the vowel. Accordingly, the raising gesture of the tongue body for [n] is relatively unaffected and overrides the activities for the surrounding vowels. As mentioned before (section 4.3.2.2), the positioning of the tongue, typical of [n], leads Recasens, et al. (1993) to a speculation on the V-to-C coarticulatory effects: an alveolopalatal is more sensitive to the difference in height of the tongue body than that in its backness. The EPG data presented by Recasens, et al. reveals that in the Catalan [n] the amount of contact in the palatal region is significantly larger in the /i/ and /u/ symmetrical contexts than in the /a/ symmetrical context. This is naturally explained by the ‘height-dependent coarticulation’ hypothesis: while the higher tongue position for [i] effectively keeps the tongue body close to the roof of the mouth, the anticipation for the vowel [a] causes the earlier lowering of the tongue body.

Two instances in our results are compatible with the prediction of the height-dependent coarticulation. The /aCV/ sequence produced by speaker TM exhibits the height effects, namely /u/=/i/>=/a/: \[t=0.31, p=0.37, df=8\] for /u/ (65.83%) vs. /i/ (65.00%); \[t=5.09, p=0.0004, df=8\] for /i/ vs. /a/ (54.16%). Another instance is the /VCi/ sequence spoken by speaker MN: \[t=1.5, p=0.08, df=8\] for /u/ (75.00%) vs. /i/ (72.30%); \[t=3.83, p=0.002, df=8\] for /i/ vs. /a/ (65.00%). It seems reasonable to suppose that the lesser degree of contact is caused by the earlier (or delayed) lowering (or raising) of the tongue body for the sequence with the open vowel. In addition, the same effects are observed for the back region of /aCV/
and /VCa/ sequences (speaker TM). However, not all the variability is satisfactorily explained by the idea of the height-dependent coarticulation; in particular, the overall difference between the speakers.

The idea of the height-dependent coarticulation would accord with the speaker-dependent focal point of articulatory closure discussed above. Given that an alveolopalatal is articulated with the tongue blade and dorsum, it is predicted by a natural synergism that the constriction location determines the degree of freedom for the tongue tip and post-dorsum: the focal point on the palate serves as a pivotal function. The constriction place of the Japanese [n] is slightly further forward than that of the Catalan [n]. In addition, the two speakers in this experiment differ in their pronunciation habits: the main constriction is further forward in speaker TM (rows 2-3) than in speaker MN (rows 3-4). It follows from these language- and speaker-specific characteristics that the height-dependent variability becomes more obvious in the case where the constriction location moves forwards. This explains why the utterances produced by speaker TM tend to reveal vowel-related variability, carryover effects in particular, in all the three regions.

4.3.3.4 [n]

Figures 4-12-1 and 4-12-2 present a series of EPG palatograms of [n] at the onset, MAX, and offset of articulatory closure. As mentioned before, the second vowel of V/n/V sequences is restricted to the non-front vowels /a/ and /u/. It can be seen that the formation and release of tongue-palate contact is similar to that of the two plosives. The contact at row 1 is constantly attained at the MAX point. The length of central articulatory closure ranges from row 1 to row 3 for speaker MN and from row 1 to row 2 for speaker TM. It is necessary to consider the tongue configuration varying with the surrounding vowels in relation to the distinctive nasal resonance.

Because of the restriction for the second vowel, the anticipatory effects were analysed for the two non-front vowels. Carryover effects, on the other hand, were analysed for the three point-vowels as in the analyses of the other consonants. The results of one-way ANOVA are summarised in Table 4-9.
Figure 4-12-1: [n] at the three temporal points of articulatory closure in the vowel asymmetric contexts (speaker MN): 'black' 100-80%; 'grey' 79-60%; 'striped' 59-40%; 'dotted' 39-20%; 'white' 19-0%.
Onset (C1)  |  Max  | Offset (C2)

/ina/  |  |  

/inu/  |  |  

/anu/  |  |  

/una/  |  |  

Figure 4-12-2: [n] at the Three Temporal Points of Articulatory Closure in the Vowel Asymmetrical Contexts (Speaker TM): ‘black’ 100-80%; ‘grey’ 79-60%; ‘striped’ 59-40%; ‘dotted’ 39-20%; ‘white’ 19-0%.
Table 4-9: One-Way ANOVA Results for Anticipatory and Carryover Effects on [n]
In The /Vna/ and /Vnu/ Sequences

(a) Anticipatory Effects

<table>
<thead>
<tr>
<th>SPEAKER MN CONSONANTS</th>
<th>REIGNS AND VALUES†‡</th>
<th>F(3,36)</th>
<th>F(3,36)</th>
<th>F(3,36)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/iCV/</td>
<td>n=t&gt;d=r</td>
<td>25.52**</td>
<td>45.86**</td>
<td>25.81**</td>
</tr>
<tr>
<td>/uCV/</td>
<td>t=n&gt;d=r</td>
<td>29.86**</td>
<td>10.29**</td>
<td>10.56**</td>
</tr>
<tr>
<td>/aCV/</td>
<td>t=n&gt;r&gt;d</td>
<td>104.31**</td>
<td>31.98**</td>
<td>29.69**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VOWELS</th>
<th>FRONT</th>
<th>F-value</th>
<th>CENTRAL</th>
<th>F-value</th>
<th>BACK</th>
<th>F-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>/iCV/</td>
<td>a&gt;u</td>
<td>2.91**</td>
<td>a=u</td>
<td>1.70*</td>
<td>a=u</td>
<td>0.75</td>
</tr>
<tr>
<td>/uCV/</td>
<td>a&gt;u</td>
<td>0.61</td>
<td>a=u</td>
<td>0.71</td>
<td>u&gt;a</td>
<td>1.99*</td>
</tr>
<tr>
<td>/aCV/</td>
<td>a&gt;u</td>
<td>0.13</td>
<td>a=u</td>
<td>1.23</td>
<td>u&gt;a</td>
<td>2.47**</td>
</tr>
</tbody>
</table>

(b) Carryover Effects

<table>
<thead>
<tr>
<th>SPEAKER MN CONSONANTS</th>
<th>REIGNS AND VALUES†‡</th>
<th>F(3,56)</th>
<th>F(3,56)</th>
<th>F(3,56)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/VCi/</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/VCI/</td>
<td>n=t&gt;d=r</td>
<td>37.50**</td>
<td>5.29**</td>
<td>4.20**</td>
</tr>
<tr>
<td>/VCIa/</td>
<td>t=n&gt;r&gt;d</td>
<td>47.22**</td>
<td>3.02*</td>
<td>6.49**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VOWELS</th>
<th>FRONT</th>
<th>F-value</th>
<th>CENTRAL</th>
<th>F-value</th>
<th>BACK</th>
<th>F-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>/VCI/</td>
<td>a&gt;u</td>
<td>0.41</td>
<td>i&gt;a=u</td>
<td>51.51**</td>
<td>i&gt;u&gt;a</td>
<td>50.06**</td>
</tr>
<tr>
<td>/VCIa/</td>
<td>i=a=u</td>
<td>3.80*</td>
<td>i&gt;u=a</td>
<td>101.47**</td>
<td>i&gt;u&gt;a</td>
<td>55.15**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPEAKER TM CONSONANTS</th>
<th>REIGNS AND VALUES†‡</th>
<th>F(3,56)</th>
<th>F(3,56)</th>
<th>F(3,56)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/VCi/</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/VCI/</td>
<td>t=n&gt;d&gt;r</td>
<td>35.70**</td>
<td>8.46**</td>
<td>21.58**</td>
</tr>
<tr>
<td>/VCIa/</td>
<td>t=n&gt;d&gt;r</td>
<td>38.29**</td>
<td>7.65**</td>
<td>11.66**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VOWELS</th>
<th>FRONT</th>
<th>F-value</th>
<th>CENTRAL</th>
<th>F-value</th>
<th>BACK</th>
<th>F-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>/VCI/</td>
<td>a&gt;u=i</td>
<td>2.49</td>
<td>i&gt;u&gt;a</td>
<td>21.14**</td>
<td>i=u=a</td>
<td>6.30**</td>
</tr>
<tr>
<td>/VCIa/</td>
<td>i=u</td>
<td>0.21</td>
<td>i&gt;u&gt;a</td>
<td>42.43**</td>
<td>i&gt;u&gt;a</td>
<td>28.34**</td>
</tr>
</tbody>
</table>

†**=p<0.05, **=p<0.01, unmarked = non-significant
†The relation between the consonant pairs and between the vowel pairs was assessed by unpaired t-test; the criterion for the significance was 0.01. The symbol > indicates a significant difference but the symbol = indicates no significant difference. A comma between the consonants means that the statistical value was not specified.
Differences in the anticipatory effects depend heavily on the fixed vowels rather than the flanking vowels /a/ and /u/. When the high front vowel /i/, either as a fixed or as a flanking vowel, takes part in a VCV syllable, it gives the same outcome that has been mentioned before: the manifestation of the other vowels is overridden. Consequently, the flanking vowel effects in the /iCV/ context turn out to be non-significant throughout the three regions and the relationship between /a/ and /u/ levels out. This is true for speaker TM but not for speaker MN: the front and central regions show significant effects in the /iCV/ sequence. The fact that the pronunciation habit of /u/ is different for each speaker is clearly reflected in the central and back regions of the /anV/ sequences. In this sequence the tongue starts its movement from a relatively low position and naturally parallels the degree of the dorsum elevation for the following vowel. Significant difference was found only in the back region for speaker MN. For speaker TM, in contrast, both central and back regions revealed significant differences.

The carryover effects are more systematic than the anticipatory effects. As seen in Table 4-9 (a) the behaviour of [n] in the central and back regions varies with the speakers: [n] parallels [t] for speaker MN but [d] for speaker TM. In contrast, Table 4-9 (b) indicates the essential patterns [t]=[n]=[d]>[r] and /i/>/u/>/a/ in the central and back regions. This suggests that the preceding vowel more systematically influences the position of the tongue body for the consonantal articulations.

The fact that the two components of the tongue are effectively modified in their shapes suggests that the oral stricture does not merely make a side-chamber to the main nasal cavity but is adjusted to effect specific vowel-colouring to the nasal resonance.

### 4.3.4 Articulatory Closure Duration and Front Region Duration

#### 4.3.4.1 Complete Articulatory Closure: Symmetrical Sequences

Table 4-10 displays duration of articulatory closure for [t, d, n, r] as a function of vowel context for the two speakers.

<table>
<thead>
<tr>
<th>CONSONANTS</th>
<th>/iCi/</th>
<th>/aCa/</th>
<th>/uCu/</th>
<th>Pooled /a,u/</th>
<th>Pooled /i, a, u/</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ViV]</td>
<td>76.13</td>
<td>4.05</td>
<td>84.07</td>
<td>3.77</td>
<td>92.01</td>
</tr>
<tr>
<td>[VdV]</td>
<td>35.74</td>
<td>2.76</td>
<td>43.03</td>
<td>9.06</td>
<td>51.63</td>
</tr>
<tr>
<td>[VnV]</td>
<td>80.10</td>
<td>15.41</td>
<td>58.91</td>
<td>2.76</td>
<td>74.80</td>
</tr>
<tr>
<td>[VnV]</td>
<td>NA</td>
<td>66.86</td>
<td>2.76</td>
<td>60.24</td>
<td>63.89</td>
</tr>
<tr>
<td>[VrV]</td>
<td>31.11</td>
<td>6.45</td>
<td>22.50</td>
<td>2.70</td>
<td>34.42</td>
</tr>
<tr>
<td>SPEAKER MN</td>
<td>/iCi/</td>
<td>/aCa/</td>
<td>/uCu/</td>
<td>Pooled /a,u/</td>
<td>Pooled /i, a, u/</td>
</tr>
<tr>
<td>[ViV]</td>
<td>76.13</td>
<td>4.05</td>
<td>84.07</td>
<td>3.77</td>
<td>92.01</td>
</tr>
<tr>
<td>[VdV]</td>
<td>35.74</td>
<td>2.76</td>
<td>43.03</td>
<td>9.06</td>
<td>51.63</td>
</tr>
<tr>
<td>[VnV]</td>
<td>80.10</td>
<td>15.41</td>
<td>58.91</td>
<td>2.76</td>
<td>74.80</td>
</tr>
<tr>
<td>[VnV]</td>
<td>NA</td>
<td>66.86</td>
<td>2.76</td>
<td>60.24</td>
<td>63.89</td>
</tr>
<tr>
<td>[VrV]</td>
<td>31.11</td>
<td>6.45</td>
<td>22.50</td>
<td>2.70</td>
<td>34.42</td>
</tr>
</tbody>
</table>

171
The results of a two-way ANOVA showed significant effects for the group of consonants \([t, d, n, r]\) in the three-vowel contexts: for speaker MN, 3 vowels \([F(2,48)=13.72, p<0.0001]\); 4 consonants \([F(3,48)=202.34, p<0.0001]\); for speaker TM, 3 vowels \([F(2,48)=23.87, p<0.0001]\); 4 consonants \([F(3,48)=130.25, p<0.0001]\). Their interaction was also significant: \([F(6, 48)=5.65, p<0.0001]\) for speaker MN; \([F(6,48)=5.17, p<0.0001]\) for speaker TM. For the group of consonants \([t, d, n, r]\) in the /a/ and /u/ contexts, significant main effects were found: for speaker MN, 2 vowels \([F(1,40)=20.82, p<0.0001]\); 5 consonants \([F(4,40)=145.87, p<0.0001]\); for speaker TM, 2 vowels \([F(1,40)=4.70, p=0.036]\); 5 consonants \([F(4,40)=144.16, p<0.0001]\). Their interaction was significant: \([F(4,40)=5.31, p=0.0015]\) for speaker MN; \([F(4,40)=6.49, p<0.0001]\) for speaker TM.

Single-factor randomised design ANOVA were applied to isolate the sources of significant interactions. It was found that the effects of vowel context upon closure duration were not significant for either speaker: in the three-vowel contexts, \([F(2,57)=1.13, p=0.328]\) for speaker MN; \([F(2,57)=2.89, p=0.063]\) for speaker TM; in the /a/ and /u/ contexts, \([F(1,38)=2.24, p=0.142]\) for speaker MN; \([F(1,38)=0.63, p=0.43]\) for speaker TM. In contrast, the effects of consonants were significant: in the group of \([t, d, n, r]\), \([F(3,56)=103.61, p<0.0001]\) for speaker MN; \([F(3,56)=57.53, p<0.0001]\) for speaker TM; in the group of \([t, d, n, r]\), \([F(4,45)=79.97, p<0.0001]\) for speaker MN; \([F(4,45)=91.75, p<0.0001]\) for speaker TM.

There are three dominant, time-related, patterns. Figure 4-13 below demonstrates the mean closure durations of \([t, d, n, r]\) pooled across the three-vowel contexts on the one hand, and those of \([t, d, n, r]\) in the /a/, /u/-pooled context on the other. A voiceless plosive \(/t/\) had a longer duration than its voiced counterpart \(/d/\). A tap was the shortest of the five. The nasals \([n]\) and \([\eta]\) were intermediate between the two plosives. It is worth noting that the duration of an alveolopalatal nasal and that of a dentalveolar nasal were not substantially different.

**Figure 4-13:** MEAN CLOSURE DURATIONS FOR \([t, d, n, r]\) IN THE VOWEL SYMMETRICAL CONTEXTS (POOLED SPEAKER CONTEXT); ‘GREY’=/a, u/-POOLED CONTEXT; ‘WHITE’=3-VOWEL CONTEXT; standard deviations are \(/t/\) (10.89), \(/t/2\) (8.38), \(/d/1\) (12.28), \(/d/2\) (8.43), \(/nj/1\) (13.37), \(/nj/2\) (9.14), \(/r/1\) (6.74), \(/r/2\) (7.37), and \(/n/\) (7.76).

X-axis: the five target consonants in the different contextual vowels
Y-axis: measures of the articulatory closure duration in milliseconds

172
Based on the data in Figure 4-13, a series of ANOVA were performed to identify the differences in articulatory closure duration for the five consonants \([t, d, n, j, r]\). The results are summarised in Table 4-11. It can be seen that the articulatory closure duration decreases in the progression \([t] > [p] > [d] > [r]\). The consonant \([n]\), although the vowel context is limited only to \([a, u]\), comes between \([t]\) and \([p]\): the durational hierarchy is \([t] > [p] > [n] > [d] > [r]\).

<table>
<thead>
<tr>
<th>TWO-WAY ANOVA Factors</th>
<th>F value</th>
<th>P value</th>
<th>Relation between the consonants (t-test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Vowels /i, a, u/</td>
<td>F(2,108)=9.90</td>
<td>p=0.0001</td>
<td></td>
</tr>
<tr>
<td>4 Consonants [t, d, j, r]</td>
<td>F(3,108)=150.35</td>
<td>p&lt;0.0001</td>
<td>(t&gt;n&gt;d&gt;r)</td>
</tr>
<tr>
<td>Interaction</td>
<td>F(6,108)=3.07</td>
<td>p=0.008</td>
<td></td>
</tr>
<tr>
<td>2 Vowels /a, u/</td>
<td>F(1,90)=12.12</td>
<td>p=0.0007</td>
<td></td>
</tr>
<tr>
<td>5 Consonants [t, d, n, j, r]</td>
<td>F(4,90)=136.14</td>
<td>p&lt;0.0001</td>
<td>(t&gt;n&gt;p&gt;d&gt;r)</td>
</tr>
<tr>
<td>Interaction</td>
<td>F(4,90)=3.42</td>
<td>p=0.0117</td>
<td></td>
</tr>
</tbody>
</table>

*The relation between the consonant pairs was assessed by unpaired t-test; the criterion for the significance was 0.01. The symbol > indicates a significant difference.

4.3.4.2 Complete Articulatory Closure: Asymmetrical Sequences

Mean closure duration of the five consonants in asymmetric sequences is given in Table 4-12.

<table>
<thead>
<tr>
<th>SPEAKER MN</th>
<th>/aCi/ msec</th>
<th>/iCu/ msec</th>
<th>/tCu/ msec</th>
<th>/uCu/ msec</th>
<th>/uCi/ msec</th>
<th>/iCu/ msec</th>
<th>/iCu/ s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>[t]</td>
<td>136.37</td>
<td>3.96</td>
<td>105.25</td>
<td>5.54</td>
<td>115.18</td>
<td>0.83</td>
<td>90.69</td>
</tr>
<tr>
<td></td>
<td>131.73</td>
<td>3.27</td>
<td>109.89</td>
<td>3.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[d]</td>
<td>95.99</td>
<td>2.12</td>
<td>78.77</td>
<td>4.20</td>
<td>84.07</td>
<td>1.51</td>
<td>60.24</td>
</tr>
<tr>
<td></td>
<td>89.37</td>
<td>2.34</td>
<td>80.10</td>
<td>1.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[n]</td>
<td>NA</td>
<td>NA</td>
<td>129.75</td>
<td>7.66</td>
<td>107.90</td>
<td>2.88</td>
<td>88.04</td>
</tr>
<tr>
<td></td>
<td>146.96</td>
<td>9.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[r]</td>
<td>133.06</td>
<td>4.86</td>
<td>60.24</td>
<td>1.48</td>
<td>60.24</td>
<td>0.83</td>
<td>67.52</td>
</tr>
<tr>
<td></td>
<td>143.65</td>
<td>4.27</td>
<td>69.51</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPEAKER TM</td>
<td>/aCi/ msec</td>
<td>/iCu/ msec</td>
<td>/tCu/ msec</td>
<td>/uCu/ msec</td>
<td>/uCi/ msec</td>
<td>/iCu/ msec</td>
<td>/iCu/ s.d.</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>[t]</td>
<td>137.03</td>
<td>5.94</td>
<td>88.04</td>
<td>2.30</td>
<td>99.30</td>
<td>1.22</td>
<td>80.10</td>
</tr>
<tr>
<td></td>
<td>129.09</td>
<td>5.61</td>
<td>104.59</td>
<td>3.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[d]</td>
<td>100.62</td>
<td>4.39</td>
<td>78.77</td>
<td>1.30</td>
<td>72.15</td>
<td>1.30</td>
<td>52.96</td>
</tr>
<tr>
<td></td>
<td>89.90</td>
<td>1.87</td>
<td>88.04</td>
<td>1.34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[n]</td>
<td>NA</td>
<td>NA</td>
<td>111.21</td>
<td>3.84</td>
<td>85.39</td>
<td>2.86</td>
<td>69.51</td>
</tr>
<tr>
<td></td>
<td>116.51</td>
<td>4.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[r]</td>
<td>99.96</td>
<td>4.32</td>
<td>56.27</td>
<td>1.58</td>
<td>59.58</td>
<td>1.58</td>
<td>50.97</td>
</tr>
<tr>
<td></td>
<td>113.20</td>
<td>3.42</td>
<td>59.58</td>
<td>1.87</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A two-way ANOVA showed significant main effects and interaction: for \([t, d, n, r]\), 6 vowel contexts \([F(5,96)=49.96, p<0.0001]\), 4 consonants \([F(3,96)=144.85, p<0.0001]\), and interaction \([F(15,96)=17.40, p<0.0001]\) (speaker MN); 6 vowel contexts \([F(5,96)=78.03, p<0.0001]\), 4 consonants \([F(3,96)=346.54, p<0.0001]\), and interaction \([F(15,96)=7.76, p<0.0001]\).
\( p < 0.0001 \) (speaker TM): for \([t, d, n, r]\), 4 vowel contexts \([F(3,80)=17.30, p<0.0001]\), 5 consonants \([F(4,80)=77.56, p<0.0001]\), and interaction \([F(12,80)=5.42, p=0.0001]\) (speaker MN); 4 vowel contexts \([F(3,80)=52.12, p<0.0001]\), 5 consonants \([F(4,80)=296.55, p<0.0001]\), and interaction \([F(12,80)=8.39, p<0.0001]\) (speaker TM). We first consider the durational characteristics of the consonants and then move on to the effects of the vowel contexts.

In Figure 4-14 the same durational hierarchy as in the symmetric contexts is found for both speakers. A one-way ANOVA indicated a significant effect: for \([t, d, n, r]\), \([F(3,116)=27.68, p<0.0001]\) (speaker MN) and \([F(3,116)=70.17, p<0.0001]\) (speaker TM); for \([t, d, n, j, r]\), \([F(4,95)=37.54, p<0.0001]\) (speaker MN) and \([F(4,95)=83.57, p<0.0001]\) (speaker TM).

One noticeable difference between the two speakers is the duration of \(/r/\). It is about twice as long for speaker MN (60.13ms) than for speaker TM (32.77ms). There is a significant difference between the two speakers: \(/r/\) \([t=7.86, p<0.0001, df=58]\) (in the three-vowel context, Figure 4-14(a)). Standard deviation, in contrast, reveals much more variability in TM’s utterances than in MN’s. This may be related to idiosyncratic characteristics of the tip/blade gesture when making a constriction. We have observed above that, while the complete closure is regularly attained in MN’s production, it tends to be lost in TM’s. It is hypothesised that the regular occurrence of complete closure generates a relatively...
long contact duration during tap articulations produced by speaker MN; whereas the higher variability and shorter duration of speaker TM’s articulations occurs because of irregularities in the constriction pattern.

The relation between the particular consonant pairs also reflects the individual characteristics of \([r]\). In the three-vowel contexts (i.e. Figure 4-14(a)) the consonants reveal a significant hierarchy \([t]>[n]=[d]>[r]\) for speaker MN and \([t]=[d]>[n]>[r]\) for speaker TM (at the \(p<0.01\) significance level). In the non-front vowel contexts (i.e. Figure 4-14(b)) there is no significant divergence between the duration of \([n]\) and that of \([r]\) in MN’s production \((t=0.64, p=0.26, df=38)\), yielding the pattern, \([t]=[n]>[d]>[n]=[r]\). In contrast, that relation is significant for speaker TM: \([t]=[n]>[d]>[n]>[r]\); \([n]\) vs. \([r]\), \([t=13.85, p<0.0001, df=38]\). In addition, it is worth noting that the consonants \([n]\) and \([n]\) reveal different durational properties. This suggests that the duration may not be a distinctive factor in the production of the nasal consonant. It must be noted that the closure duration becomes longer when the sequence contains the high vowel \(/i/\).

Turning now to the effects of the vowel contexts, Figure 4-15 summarises the mean closure duration as a function of the vowel contexts averaged across the two speakers. To explain contextual effects on the closure duration, the comparison in terms of the position of the high vowels \(/i, u/\) may be helpful. In the production of \([t]\) and \([d]\), the sequences with \(/i/\) in the \(V_2\) position (grey bars) are longer than the other sequences; when the sequences contain \(/a/\), those with the high vowel, either \(/i/\) or \(/u/\), in the \(V_2\) position, tend to be longer. The limited comparison for \(/n/\) shows that, while the consonant \([n]\) exhibits the same result for the sequence with \(/u/\) as the \(V_2\), a different result is found for \(/i/\). The duration becomes longer when the \(V_1\) is \(/i/\).

Figure 4-16 represents the closure duration of \(/t/\) calculated separately for the two speakers. The general trend observed in \(/t/\) and \(/d/\) data is more clearly found for speaker MN than for speaker TM. This is because the closure duration of TM’s tokens is generally shorter. There is one opposite result in the relation between \(/ari/\) and \(/ira/\) (MN, dotted bars).

From these observations one general point becomes clear: the vowel height and backness contribute to vary the closure duration. The tongue body movements from low to high and from back to front lengthen the closure duration. In contrast, those from high to low and from front to back shorten it. This tendency is evidently represented in the results of t-tests given in Table 4-13: \(/t/\) and \(/d/\) of speaker MN and \(/t/\) of speaker TM. The behaviour of \(/r/\) is unusual and this can be closely related to the tighter constraint on the tongue body movement in the production and to the pronunciation habits of each speaker.
Figure 4-15: Mean Articulatory Closure Durations and Standard Deviations for [t, d, n, n] in the Vowel Asymmetrical Sequences (Pooled Speaker Context): X-axis = the target consonants; Y-axis = measures of the contact duration in milliseconds; standard deviation is bracketed.

Figure 4-16: Mean Articulatory Closure Durations and Standard Deviations for [r] in the Vowel Asymmetrical Sequences: standard deviation is bracketed.

Table 4-13: t-Test Results for the Contextual Effects on the Articulatory Closure Duration

<table>
<thead>
<tr>
<th>SPEAKER MN</th>
<th>[t]</th>
<th>[d]</th>
<th>[r]</th>
<th>[p]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONCEPTS</td>
<td>Relation (df=8)</td>
<td>Relation (df=8)</td>
<td>Relation (df=8)</td>
<td>Relation (df=8)</td>
</tr>
<tr>
<td>/a_i/ vs. /i_a/</td>
<td>3.08**</td>
<td>2.46*</td>
<td>3.53**</td>
<td>9.66**</td>
</tr>
<tr>
<td>/a_u/ vs. /u_a/</td>
<td>7.05**</td>
<td>7.20**</td>
<td>3.11*</td>
<td>2.20**</td>
</tr>
<tr>
<td>/u_i/ vs. /i_u/</td>
<td>3.11**</td>
<td>2.33*</td>
<td>1.29</td>
<td>11.55**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPEAKER TM</th>
<th>[t]</th>
<th>[d]</th>
<th>[r]</th>
<th>[p]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONCEPTS</td>
<td>Relation (df=8)</td>
<td>Relation (df=8)</td>
<td>Relation (df=8)</td>
<td>Relation (df=8)</td>
</tr>
<tr>
<td>/a_i/ vs. /i_a/</td>
<td>5.19**</td>
<td>1.71</td>
<td>0.23</td>
<td>6.41**</td>
</tr>
<tr>
<td>/a_u/ vs. /u_a/</td>
<td>6.74**</td>
<td>2.71*</td>
<td>1.17</td>
<td>2.15*</td>
</tr>
<tr>
<td>/u_i/ vs. /i_u/</td>
<td>2.59*</td>
<td>1.00</td>
<td>3.80**</td>
<td>9.29**</td>
</tr>
</tbody>
</table>

The symbol >, or <, indicates a significant difference but the symbol = indicates no significant difference.

**p<0.05, ***p<0.01, unmarked = nonsignificant

176
4.3.4.3 Duration of Front Region Contact

Table 4-14 summarises the mean duration of front region contact both in symmetric and in asymmetric contexts. This is the overall duration of the three articulatory stages, approach, hold, and release, and is considered as the duration of consonantal articulation.

| TABLE 4-14: MEAN VALUES AND STANDARD DEVIATIONS OF THE DURATION OF CONSONANTAL GESTURE |
|-----------------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| SPEAKER MN                              | VOWEL CONTEXTS AND DURATION |
| CONSONANTS                               | /t/Ci/ | /a/Cu/ | /a/Cu/ | /u/Cu/ | Pooled /a,u/ | Pooled /i, a, u/ |
| [VtV]                                   | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  |
| 132.40                                  | 7.76   | 99.96 | 1.48  | 121.14 | 9.24  | 110.55 | 12.79 | 117.83 | 15.36  |
| [VdV]                                   | 104.59 | 5.53  | 72.82 | 6.19  | 88.70 | 2.76  | 80.76 | 9.51  | 88.70 | 14.22  |
| [VbV]                                   | N.A.   | 83.41 | 4.31  | 88.04 | 7.62  | 85.72 | 6.32  | N.A.   |        |
| [VjV]                                   | 205.22 | 24.32 | 94.00 | 5.01  | 141.00 | 11.60 | 117.50 | 26.16 | 114.52 | 35.75  |
| [VrV]                                   | 56.93  | 3.62  | 39.05 | 2.76  | 50.97 | 7.97  | 45.01 | 8.43  | 48.98 | 9.12   |
| SPEAKER TM                              | VOWEL CONTEXTS AND DURATION |
| CONSONANTS                               | /t/Ci/ | /a/Cu/ | /a/Cu/ | /u/Cu/ | Pooled /a,u/ | Pooled /i, a, u/ |
| [VtV]                                   | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  | msec  | s.d.  |
| 149.61                                  | 4.31   | 77.45 | 8.30  | 102.61 | 8.10  | 109.89 | 31.65 | 109.89 | 31.65  |
| [VdV]                                   | 122.47 | 6.19  | 57.59 | 5.01  | 70.17 | 6.36  | 63.88 | 8.55  | 83.41 | 29.58  |
| [VnV]                                   | N.A.   | 77.45 | 5.53  | 78.77 | 4.90  | 78.11 | 4.98  | N.A.   |        |
| [VjV]                                   | 159.54 | 18.03 | 84.73 | 6.45  | 99.30 | 13.24 | 92.01 | 12.46 | 146.74 | 49.40  |
| [VrV]                                   | 41.04  | 4.44  | 38.39 | 2.90  | 31.77 | 4.44  | 35.08 | 4.98  | 37.07 | 5.48   |

The results of two-way ANOVAs showed significant effects: for 3 vowels, [F(2,48)=132.12,p<0.0001] (MN) and [F(2,48)=225.79,p<0.0001] (TM); for 4 consonants (i.e. [t, d, n, r]), [F(3,48)=295.96,p<0.0001] (MN); [F(3,48)=267.80,p<0.0001]; and for interaction, [F(6,48)=26.17 p<0.0001] (MN); [F(6,48)=21.92,p<0.0001]. The comparison in the /a, u/ contexts also indicated significant effects: for 2 vowels, [F(1,40)=114.66,p<0.0001] (MN) and [F(1,40)=22.06,p<0.0001] (TM); for 5 consonants (i.e. [t, d, n, r]), [F(4,40)=185.96, p<0.0001] (MN) and [F(4,40)=109.43,p<0.0001] (TM); and for interaction, [F(4,40)=14.84, p<0.0001] (MN) and [F(4,40)=7.57,p=0.00012] (TM).
The comparison between consonant types was conducted by one-way ANOVA and the results are given in Table 4-15 below.

### TABLE 4-15: ONE-WAY ANOVA RESULTS FOR THE DURATION OF CONSONANTAL GESTURE

#### (a) Symmetrical Sequences

<table>
<thead>
<tr>
<th>SPEAKER MN</th>
<th>FACTORS</th>
<th>F VALUE</th>
<th>P VALUE</th>
<th>The Relation between the consonants (t-test)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 CONSONANTS [t, d, p, r]</td>
<td>F(3,56)=23.76</td>
<td>p&lt;0.0001</td>
<td>p=t&gt;d&gt;r</td>
<td></td>
</tr>
<tr>
<td>5 CONSONANTS [t, d, n, p, r]</td>
<td>F(4,45)=53.31</td>
<td>p&lt;0.0001</td>
<td>p=t&gt;n&gt;d&gt;r</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPEAKER TM</th>
<th>FACTORS</th>
<th>F VALUE</th>
<th>P VALUE</th>
<th>The Relation between the consonants (t-test)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 CONSONANTS [t, d, p, r]</td>
<td>F(3,56)=35.31</td>
<td>p&lt;0.0001</td>
<td>p=t&gt;d&gt;r</td>
<td></td>
</tr>
<tr>
<td>5 CONSONANTS [t, d, n, p, r]</td>
<td>F(4,45)=39.09</td>
<td>p&lt;0.0001</td>
<td>p=t&gt;n=d&gt;r</td>
<td></td>
</tr>
</tbody>
</table>

#### (b) Asymmetrical Sequences

<table>
<thead>
<tr>
<th>SPEAKER MN</th>
<th>FACTORS</th>
<th>F VALUE</th>
<th>P VALUE</th>
<th>The Relation between the consonants (t-test)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 CONSONANTS [t, d, p, r]</td>
<td>F(3,116)=105.37</td>
<td>p&lt;0.0001</td>
<td>p=t&gt;d&gt;r</td>
<td></td>
</tr>
<tr>
<td>5 CONSONANTS [t, d, n, p, r]</td>
<td>F(4,95)=61.60</td>
<td>p&lt;0.0001</td>
<td>p=n=t&gt;d&gt;r</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPEAKER TM</th>
<th>FACTORS</th>
<th>F VALUE</th>
<th>P VALUE</th>
<th>The Relation between the consonants (t-test)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 CONSONANTS [t, d, n, p, r]</td>
<td>F(3,116)=109.13</td>
<td>p&lt;0.0001</td>
<td>p=t&gt;d&gt;r</td>
<td></td>
</tr>
<tr>
<td>5 CONSONANTS [t, d, n, p, r]</td>
<td>F(4,95)=73.10</td>
<td>p&lt;0.0001</td>
<td>p=n=t&gt;d&gt;r</td>
<td></td>
</tr>
</tbody>
</table>

†The relation between the consonant pairs was assessed by unpaired t-test; the criterion for the significance was 0.01. The symbol > indicates a significant difference but the symbol = indicates no significant difference.

The general trend is that each target consonant has a distinctive temporal interval for the tip/blade movement, yielding the durational hierarchy [t]>[d]>[r]. The property of the two nasals, [n] and [p], varies with individual speakers: the consonants exhibit durational characteristics similar to [t] for speaker MN; the durational characteristics are closer to [d] for speaker TM. Although ‘closure’ duration of [n], as observed before, was shorter than the two plosives, the duration of the front region appears to be the longest of the five consonantal articulations. This suggests that the shorter ‘gestural’ duration is not always necessary to attain the shorter ‘closure’ duration. The relation between the two temporal-intervals is not proportional in the articulation of [n], but is proportional in the articulation of [r], in which both gestural and closure durations are relatively short.

Turning to the effects of the contextual vowels, Figure 4-17 summarises the activation period of the front region pooled across the speakers. The gestural duration of [r] is shown separately for each speaker in Figure 4-18. It can be seen that the contextual vowel, particularly a high front vowel /i/ either as V₁ or V₂, affects the consonantal duration. In asymmetric contexts, a pattern similar to that in closure duration can be observed: the sequences with the high vowels, /i/ or /u/, in the V₂ position tend to be longer than the others; those with the high front vowel in the V₂ position are longest. The articulatory duration of [r]

178
reflects the idiosyncratic features noticed in the closure duration; the duration is shorter and displays smaller divergence in speaker TM. The comparisons between the specific pairs of vowel contexts in Table 4-16 demonstrate the same relationship pattern as in Table 4-13 for closure duration.

**Figure 4-17:** Mean Front Region Durations for [t, d, n] in the Vowel Asymmetrical Contexts (Pooled Speaker Context): X-axis = the target consonants; Y-axis = measures of the duration in milliseconds. Standard deviations are: for speaker MN, /ati/ (12.33), /ita/ (11.66), /atu/ (8.99), /uta/ (8.08), /ati/ (14.40), /itu/ (9.20), /adi/ (11.04), /ida/ (5.85), /adu/ (7.68), /uda/ (5.88), /adi/ (8.44), /idi/ (5.88), /ani/ (12.33), /inju/ (11.66), /anju/ (8.99), /unju/ (8.08), /uni/ (14.40), /inju/ (9.20), /una/ (14.87), /una/ (13.22), /ina/ (11.26), /inu/ (10.02).

**Figure 4-18:** Mean Front Region Duration for [r] in the Vowel Asymmetrical Contexts: X-axis = the target consonants; Y-axis = measures of the duration in milliseconds. Standard deviations are: for speaker MN, /ari/ (10.87), /ira/ (11.12), /aru/ (7.62), /ura/ (5.53), /uri/ (8.63), /iru/ (7.17); for speaker TM, /ari/ (5.43), /ira/ (2.96), /aru/ (1.81), /ura/ (3.77), /uri/ (2.96), /iru/ (5.92).
From these observations and those in the previous section, it seems reasonable to suppose that there is some kind of mathematical relationship between the duration of articulatory closure and that of consonantal gesture. Increases in one durational property are generally associated with increases in the other as the type of consonant changes, although not all the consonants show this pattern. Given that two durational properties are inherent in the articulation of coronal stops, each consonantal articulation will be controlled with their own activation intervals. It could be argued that this is not a linear correlation.

The inquiry above can also be linked to the question concerning the relationship between the closure duration and the amount of tongue-palate contact. If these are the related parameters, as claimed by Fujimura et al. (1973) for Japanese [t, d] and Farnetani et al. (1989) for Italian [t], then other stops should be distributed along such a continuum. The claim will be further supported by the results of the duration of consonantal gestures. For examinations of these properties, it must be noted that in the articulation of /r/, the contact degree varies greatly with individual speakers but the duration is generally shortest of the four consonants.

4.4 Searching for Linguistically Significant Articulatory Parameters

In the previous sections we have extracted and examined various EPG correlates of the consonantal articulations in the coronal region. Based on the data presented so far, this section explores how a set of articulatory parameters are organised in the coronal stop consonants [t, d, r, n, p]. Four parameters were chosen for the regression analysis (the Pearson product-moment coefficient of linear correlation): (i) the closure duration, (ii) the gestural duration (i.e. the contact duration in the front region), (iii) the degree of front contact; and (iv) the degree of central contact. Three correlations were examined for all the target consonants.
The cluster of the plots shows divergence: one develops diagonally while the other develops horizontally. The former, the diagonal clustering, indicates that the closure duration and the gestural duration are identical. This clustering includes the /a/ and /i/ tokens of [t, d], the /a/ tokens of [n], and the /u/ tokens of [p]. In contrast, the latter, the horizontal clustering, demonstrates that the closure duration is shorter than the gestural duration. This clustering involves the /a/ tokens of [t, d, n] and the /a/ and /i/ tokens of [p]. Such a divergent development, as we will see below, was not obtained for the [r] utterances. It must be noted that some plots appear above the diagonal clustering: the gestural duration is shorter than the closure duration. Those plots reveal obvious inaccuracies of measurement. Because the rapid onset of the articulatory closure tends to occur in the /a/ vowel context, this movement leaves few traces of the tongue-palate contact approaching towards the complete closure. In this case it is difficult to measure the contact interval for the gestural duration accurately. Also, it should be reminded here that the exact beginning and finishing points of the lingual articulation are not detected by the EPG technique: it records the facts about tongue-palate contact. Therefore, we must allow the gestural duration in this experiment to be an approximation of the actual time interval that is required for the activation of the tongue. In the analysis shown in Figures 4-19 and 4-20, the data plots in question are included, since they do not substantially affect the statistical results.
with the speakers pooled. The fits and slopes are shown in all the Figures. By a good fit, it is meant that one parameter increases proportionally to the other. The higher the $r^2$ value, the more linear the relation between the parameters.

### 4.4.1 Gestural Duration and Closure Duration

The durational analysis in the previous section generally indicates that each consonant has its own characteristics in the closure duration and gestural duration. If we consider the two durational properties as a natural attribute of coronal stop articulations, it can be asked how the value of closure duration will vary over that of articulatory duration. This is shown in Figure 4-19.

In Figure 4-19, the durational parameters specific to individual consonantal articulations of [t, d, r, n, p] together make up a continuum. Given the linear relationship, the duration of consonantal articulation accounts for up to 35% of the variation in closure duration as indicated by the $r^2$ value. Although closure duration tends to be proportional to gestural duration, it must be noticed that there are two kinds of clustering: one reflects a proportional increase, while the other exhibits a relatively steady state of closure duration at around 70msec. This will be clarified if we consider the target consonants separately.

**Figure 4-19: Relationship Between Closure Duration and Gestural Duration (All the Consonants Across the Speakers): n=420**

Figure 4-20 summarises the two durational parameters separately for the five target consonants. The first thing one notices is that the correlations vary greatly with the consonant type. This is closely related to the coarticulatory characteristics of a given consonant. Since the duration of articulatory closure includes transitions to and from the articulatory closure, the tongue body height for the preceding and the following vowels affects the timing of the involvement of the tip/blade gesture. Such a contextual effect is chiefly caused by the high
Figure 4-20: Relationship Between Closure Duration and Gestural Duration (each consonant across the speakers): n=90 for [t], [d], [r], and [n]; n=60 for [n].

Arrows in the graphs of [t, d, n] indicate the /u/ tokens (see text for explanation): an arrow in [p] indicates the tokens of Vn/i/ syllables.
vowel /u/. When the utterance contains that vowel, the closure duration tends to be unchanged, but the gestural duration tends to be longer. As marked by an arrow in Figure 4-20, it was found that: (i) /utu/ tokens exhibited a larger range of variation in the closure duration than in the gestural duration; (ii) the duration of closure for /udu/ tokens was relatively stable; and (iii) the plots of /unu/ utterances were isolated and develop horizontally. In contrast, such a contextual effect was not obtained for [r], most tokens of which were tightly clustered within a shorter durational range, showing the highest correlation of the five.

Another interesting effect is found for [n]. The plots exhibit two diverse developments, yielding a lower correlation value. One progression is composed of the /Vni/ tokens and the other is composed of the /Vnja/ and /Vnju/ tokens. In the latter group the closure duration appears to be tightly controlled at about 60msec. When those two groups were analysed separately, they showed tolerable values of correlation: for /Vni/, $r^2=0.92$ [$F(1,28)=341.36, p<0.0001$]; for /Vnja/ and /Vnju/, $r^2=0.24$ [$F(1,58)=18.76, p<0.0001$]. It can be argued that the consonant [n], although transcribed by the same symbol, involves different timing strategies with reference to the following vowel. In /Vni/ sequence the tip/blade component for making a constriction is activated with a relatively long and continuous movement of the dorsum component that realises the following high front vowel. In contrast, in the /Vnja/ and /Vnju/ sequences, it is not plausible to assume such a constant movement of the dorsum component: the activation of the tongue dorsum may be shorter. This is compatible with the configurational characteristics of the consonant in question: the two components are actively involved in the constriction formation. Furthermore, the temporal diversity is related to the different phonological status of [n]: it is non-contrastive before a high front vowel but is contrastive before non-front vowels. This suggests that the control of the articulators is organised phonologically. This point will be further discussed in Chapter 6 by examining Japanese palatalisation.

Overall, the results provide evidence that the duration of closure and that of consonantal gesture are consonant-specifically determined for the coronal stops. Some require a tight temporal control for both parameters, while others depend on phonetic contexts and (possibly) phonological status. It must be noted that the shorter closure duration does not necessarily correspond to the shorter gestural duration.

### 4.4.2 Closure Duration and Front Contact Degree

The hypothesis that the stop articulations [t, d, n, p, r] constitute a linear continuum in terms of the contact degree and the closure duration is examined. This is an extension of the observation made by Fujimura et al. (1973): those two articulatory parameters clearly reveal a linear correlation in the production of /t/ and /d/, with articulatory closure increasing and the tongue-palate contact becoming higher. Farnetani et al. (1985) also found equivalent results
for Italian /t/ in a changing vowel environment. Our result for [t, d, n, j, r] across the two
speakers is given in Figure 4-21 below.

**Figure 4-21:** Relationship between Front Contact Degree and Articulatory Closure Duration (all the consonants across the speakers): n = 420.

It can be seen that the plots take a general developing pattern that the degree of front
contact increases as the closure duration increases. The $r^2$ value implies that 47% of the
variation in the front contact degree is accounted for by the closure duration. In general, the
two parameters are linearly ordered within the coronal stop articulations; we do find support
for the set of coronal consonants hypothesised above.

The linear correlation predicts that the contact degree becomes smaller if the closure
duration is shorter. However, not all the consonants reflect such a property and indeed, some
show distinctive characteristics for the parameters in question. The analyses of each of the
five consonants are shown in Figure 4-22. Although the $r^2$ value of individual consonants is
smaller than that of the coronal consonants considered as a set, the pattern of clustering was
found to be consonant-specific.

One noticeable exception to the linear relation is the production of [r]; the front
contact degree is seen to be as large as that in [d]. As explained in configurational and
coarticulatory characteristics, the tip/blade gesture varies with phonetic contexts and with
individual speakers. We have found that, while a complete occlusion is regularly attained by
speaker MN regardless of vowel environments, a closure tends to be incomplete in speaker
TM’s production particularly in the high vowel contexts. The relatively tight and thin
clustering in the vertical dimension reflects these contextual and idiosyncratic variations. The
closure duration is actually the shortest of the five, but the contact degree is not smallest.
Figure 4-22: Relationship between front contact degree and articulatory closure duration (each consonant across the speakers): n=90 for [t], [d], [r], and [p]; n=60 for [n].
The lowest correlation was obtained for [n]. This is probably because the tip/blade gesture is so firmly involved in the articulation that only a minimal additional increase is evidenced as the closure duration changes. Recall that the /Vnja/ and /Vnju/ sequences involve shorter closure and gestural durations than the /Vni/ sequences. Therefore, over a greater area, there is a tight clustering on the one hand and a poor clustering on the other. Similar configurational characteristics also reflect different $r^2$ values between [t] and [n]. The former attains a larger and more stable amount of front contact than the latter, even though the closure duration varies with the vowel contexts. The tight clustering of [d] suggests that it constantly maintains the two parameters regardless of phonetic contexts.

These consonant-specific characteristics raise a question particularly concerning the relationship between [d] and [r]. The two consonants are commonly characterised as having equivalent motor actions, differing only in the speed. The result of the comparison is given in Figure 4-23, which includes the comparison between [t] and [d] for reference.

![Figure 4-23: [t, d, c] Relationship Front Contact Degree and Articulatory Closure Duration (selected consonant pairs across the speakers): n=180.](image)

The results show that the linear correlation is weaker in a cluster of /d/ and /r/ than in that of /t/ and /d/. This implies that the common observation mentioned above is an underestimation of the articulatory gestures involved. The weaker correlation comes from the distinctive durational properties specific to /d/ and /r/, since the distribution of the contact degree is in a similar range. On the other hand, the higher variability in the contact amount questions whether the speed is the most influential factor in the articulation of [r]. We will return to this point in the next section.

Overall, the regression analyses present evidence for a linear correlation between the amount of tongue-palate contact in the front region and the closure duration when the target consonants are considered as a set. This linear continuum is composed of contrasting grades of the $r^2$ values that are specified for each consonant.
4.4.3 Intergestural Coordination

We have noted that the two components of the tongue, the tip/blade and the dorsum, are systematically constrained in the coronal consonants in question. The linear correlation between the two functional divisions of the tongue is examined here.

Figure 4-24 demonstrates that 30% of the variation of the front contact is due to the variation of the central contact. Although the $r^2$ value is weak, the correlation is still considered as linear. This weakness, in turn, suggests the possibility that the parameter for intergestural coordination is a distinguishing parameter of the coronal consonants.

![Figure 4-24: Relationship between front contact degree and central contact degree (all the consonants across the speakers): n=420.](image)

Figure 4-25 presents the results of the regression analyses separately for each consonant. The five target consonants fall into two groups in terms of the $r^2$ values; the values are much weaker for [t, r, n] than for [d, n]. The similarities of the $r^2$ values between the particular consonants suggest that, although physiologically different control strategies are used for intergestural coordination, they all attain functionally the same manifestation of coupling effects between the two components of the tongue.

For [t], the tip/blade component of the tongue is independent of the dorsum component. The length of the tip/blade occlusion is so large and steady that the deviations are very limited. As observed in both anticipatory and carryover effects the tongue body elevation apparently reflects the changes in contact amount in the central region. What the weaker $r^2$ value reflects is this relative independency of the two components that are evidently involved in the constriction formation.

In the production of [n], the $r^2$ values implies that both tongue blade and dorsum are actively involved in the constriction formation. The results exhibit a tight clustering at a relatively high area. However, as observed before, the involvement of the full-tip closure is
Figure 4-25: Relationship between front contact degree regressed against central contact degree (each consonant across the speakers): n=90 for [t], [d], [r], and [n]; n=60 for [n].
variable: the consonant is articulated with or without it. If the tip is regularly directed downwards, the inverted relation can be obtained; the front contact decreases as the central contact increases. These coupling effects, or correlations between the two components of the tongue, lead us to further considerations in relation to the feature specifications of the consonants in question.

In the articulation of [r], a bare minimum correlation was obtained. The raising gesture of the tongue body is consciously controlled with respect to the tip/blade gesture. As the clustering indicates, while the activation in the front region varies greatly with phonetic contexts and speakers, that in the central region is highly restricted within the range 0-50%. It has been assumed that the speed, or shorter duration, is a most relevant property of [r] (e.g. Catford, 1977). Given this, the degree of front contact would correlate perfectly with the duration of closure in the production of [r], yielding the correlation that can reasonably extend to the other stop articulations. This is not supported by our results.

What is more significant in our data, from the viewpoint of speakers' control strategy, is the relative inactivation of the central region. This suggests that the production constraint is more strongly imposed on the dorsum component for [r] than for the other consonants. Connell (1995) claims that a lesser or incomplete contact is more relevant than shorter duration. However, it is quite possible that the lesser amount of front contact is simply the result of the synergetic interplay between the two components of the tongue. The raising gesture of the tongue body is extremely antagonistic to the tongue tip/blade gesture in the production of [r]. This means that a speaker consciously chooses a particular way of resolving antagonistic gestures on a case-by-case basis; and that an incomplete closure is one of various means to sort out such a conflict. This view, in turn, raises a question about the widely accepted relationship between [r] and [d].

The higher $r^2$ value in the production of [d] suggests that the two components of the tongue are equally involved in forming a constriction, maintaining the distinctive voicing feature. In his cineradiographic study Perkell (1969; 21) reports that the pre- and medio-dorsum are depressed during the complete closure of [d] spoken by an American English speaker. In addition, the greater widening in pharynx width was found in the voiced plosive relative to its voiceless counterpart [t]. Perkell (1969; 35) speculates two possible interpretations for the potential link between the voicing feature and the depression observed. The voicing in [d] requires the expansion of the vocal tract. Because the vocal tract is constricted by the tongue tip/blade, the airflow passing the approximated vocal cords must increase supraglottal pressure. This effectively causes the vocal tract to expand to attain a large enough volume for voicing (e.g. Kent & Moll, 1969; Bell-Berti, 1975; Westbury, 1983). According to this view, Perkell continues, the depression on the tongue dorsum during the closure is part of the expansion of the entire vocal tract. Alternatively, the widening activities
by the dorsum and root of the tongue are accomplished by an active contraction of tongue musculature. There is evidence in favour of this interpretation and it is closely related to the similarity between [d] and [r].

The articulatory trajectories of /iti/ and /idi/ syllables are given in Figure 4-26 below. Each panel contains two trajectories, anterior (rows 1-4) and posterior (rows 5-8), of the five repetitions.

![Figure 4-26: Articulatory Trajectories of /iti/ and /idi/ Syllables](image)

X-axis = Time in frames; Y-axis = measures of the percentage contact (%)
It can be seen that the posterior trajectory exhibits a movement typical for each consonantal articulation. Note in passing that the ‘trough phenomenon’ (Engstrand, 1989) of the dorsum contour is observed. In the production of the /idi/ syllable, the tongue-palate contact in the posterior region commences its decrease immediately after the anterior contact attains its peak (i.e. a maximum articulatory closure). In the production of the /iti/ syllable, in contrast, the posterior trajectory tends to remain constant throughout the peak, keeping the tongue position of the preceding high vowel. This earlier reduction of the dorsum activity in the [d] production, as well as the relatively empty space around the middle of the EPG palatogram, can be considered as suppression effects similar to Perkell’s data. Although our EPG data in Japanese shows similar results for the dorsum depression as in American English, there is not much evidence to decide the matter. However, the result allows us to speculate the potential link between [d] and [r].

Tap articulation explicitly demonstrates the earlier reduction of the dorsum activity that was found for [d]. Figure 4-27 illustrates articulatory trajectories of the five productions of /iri/ syllables spoken by two speakers.

Evidently the posterior trajectory commences its substantial decrease before the anterior trajectory attains its peak. In other words, the position of the tongue dorsum for the preceding /i/ is lowered before the tongue tip/blade comes into action; namely flicking or transient activity for [r]. This earlier lowering explicitly reveals that the dorsum component is tightly constrained in the temporal domain of the articulation, as well as in the spatial domain.
When considered from the viewpoint of the tongue configuration, the depression of the tongue dorsum activity during the closure of [d] can be regarded as functionally the same as the earlier reduction observed for [r]. Either activity makes a hollow on the surface of the tongue. This possibility, as reviewed in sections 3.5 and 3.6 in Chapter 3, is not explored either by Gimson (1980) or by Akamatsu (1997). Furthermore, Ladefoged (1993; 168) asserts that 'it [a tap or flap] is often just a very rapid articulation of a stop'. However, it is more plausible to assume that the manifestation of motor equivalence is not limited to one aspect of articulation such as lingual closure. The EPG evidence suggests that the voicing feature in [d] and the tighter constraint on the dorsum movement in [r] functionally generate the same effect on the surface shape of the tongue, the depression or a hollow. This assumption can account for the alternation between the two consonants commonly found in child's speech and adult's careless pronunciations. Furthermore, our findings suggest that carryover effects of the preceding vowel are not automatic, or mechanical, as commonly assumed (e.g. Gay, 1977), but are consciously controlled by the speaker. In addition, the stability of [r] implies that a consonant that strongly resists the vowel effects, affects both the preceding and the following vowel gestures.

In summary, the examination of the linear correlation for the four articulatory parameters provides evidence for the potential factors responsible for the underlying unity of the coronal stop articulations. Given the linear relationship, the pairs of parameters reveal different degrees of significance along with consonant-specifically determined values. If we assume that differing grades of significant correlation are associated with the distinctiveness of the given consonantal articulation in the category, as far as the EPG correlates are considered, then the parameter of intergestural coordination is a better candidate than the other two.

4.5 Summary and Conclusion

In this chapter the articulatory and coarticulatory characteristics of the coronal stop consonants have been characterised in terms of the two-component model of the tongue physiology. Based on the EPG data, the relation between the tip/blade and the dorsum was evaluated by comparing the tongue-palate contact patterns in the anterior region (rows 1-4) with those in the posterior two regions (rows 5-7 and row 8). Various articulatory parameters were extracted and their correlations were considered. Our summary covers these three points: (i) configurational characteristics, (ii) coarticulatory characteristics, and (iii) articulatory parameters that are systematically related to linguistically contrastive events.

The consonants [t, d, n] are essentially apico-laminal on the tongue and dentalveolar on the palate. The length of apico-laminal closure for [t, d] varies with the voicing features and that for [n] depends on the contextual vowel. The consonant [n] is considered as an
alveolopalatal nasal, the quality of which ranges from [n^1] to [n]. This consonant is articulated with the pre-dorsum and the posterior half of the blade pressing against the alveolar region on the palate. This primary constriction is supported by the involvement of the tongue medio- and post-dorsum. There are language-specific differences in the constriction place: the Japanese [n] was formed slightly further forward (rows 1-4) compared to the Catalan [n] (rows 2-5). Furthermore, this frontal constriction exhibited idiosyncratic differences: the focal point of the closure is more advanced for speaker TM (rows 2-3) than for speaker MN (rows 3-4). It was found that these language-specific and speaker-specific differences were closely related to V-to-C coarticulatory effects. The consonant [r] involves both the tip and the very anterior part of the blade in making a central closure. The flicking movement and transient movement of the tongue tip/blade are interchangeably used by the same speaker. The presence of a complete articulatory closure along the alveolar ridge, as suggested by Laver (1994), is a well-justified characteristic of the coronal stop articulations.

It is reasonable to conclude that the target consonants [t, d, n, j, r] have consonant-specific control strategies for the tip/blade and the tongue dorsum. The EPG evidence demonstrated that: (i) the tongue dorsum in the production of [t] readily accommodates the height requirements specific to the adjacent vowels: (ii) the same is true of [n], for which the overall tongue body efficiently reflects the location and shape of the particular vowel to modify the nasal resonance: (iii) the vocalic gesture for [r] is minimised in degree so that it does not interfere with the flicking (or transient) movement of the tongue tip/blade; and (iv) [n] essentially involves the tongue dorsum for the constriction formation. Because the tongue tip/blade component is the primary articulator for coronal stop articulations, the different manifestations of the dorsum activity during articulatory closure are considered as self-maintained, in order to realise the tip/blade activity successfully.

This dependency relation between the two components of the tongue is generally referred to as the ‘coupling effect’ and this has been commonly described as ‘loose’ or ‘strong’ (e.g. Farnetani, et al. 1985; Kent & Moll, 1972; Recasens, et al. 1993). It appears that the estimation of the degree of coupling tends to be misleading when referred to as the property of only a single consonantal articulation. Furthermore, the term is ambiguous because it often refers to the extent to which the given consonantal articulation involves the contact between the tongue dorsum and the palate. Accordingly, what is strong in the coupling effects is the large amount of tongue-palate contact. This is wrong for the consonants [d, r]: these consonants exhibit a lesser degree of tongue-palate contact in the posterior or overall palate. Nevertheless, the movement of the tongue dorsum was highly constrained in a systematic way, in order to maintain the voicing characteristic and the tip/blade gesture.

The consonant-specific control strategies for the dorsum movement imply the extent to which the V-to-C coarticulatory effects are allowed, or resisted, in the given consonantal
articulation. The EPG data suggests that consonantal and vocalic influences differ in degree across the three regions, front, central, and back. In general, the effects of the consonant type are much more influential in the anterior region (rows 1-4), and both the consonant type and the contextual vowel affect the contact pattern in the posterior regions (rows 5-7 and row 8). The following patterns were commonly obtained\(^1\): \([\text{n}]>[\text{t}]>[\text{d}]>[\text{r}]\) in the full-context comparison; and \([\text{n}]>[\text{t}]>[\text{n}]>[\text{d}]>[\text{r}]\) in the \(V_1CV_2\) sequences, the \(V_2\) of which is limited to /a/ and /u/. For the vocalic gestures, the magnitude of displacement decreases from /i/ through /u/ to /a/. There are idiosyncratic differences in these hierarchical patterns. In the articulation of [r] one speaker (TM) indicates the tendency of incomplete closure in the context of the high vowels /i/ and /u/. In this connection, the situation is the same for the presence and absence of the full-tip closure in the [n] articulation. This, however, can also be considered as the permitted variability in the articulatory nature of [n]: the tip closure is readily tied in with the raising of the tongue pre- and medio-dorsum for making a primary constriction.

The anticipatory and carryover effects were examined by comparing the contact amount at the MAX point and by detecting the changes in a series of EPG palatograms. We showed that the carryover effects in the central and back regions are more systematic and more extensive than the anticipatory effects. This agrees with previous studies (e.g. Farnetani, et al., 1985; Parush et al., 1983; Recasens, 1984). The contextual vowels affect the positioning of the tongue body during an articulatory closure, more than the tongue tip/blade, or the length of the front constriction (e.g. Farnetani et al., 1985; Gay, 1977). High vowels, either in the \(V_1\) or \(V_2\) position, influence the duration of consonantal articulation and closure; the longest duration is obtained when \(V_2\) is a high-front vowel.

A series of EPG palatograms demonstrated the continuous activation of the central and back regions during the front closure. As Ohman (1966) suggests, the consonant-related movement of the tongue tip/blade is superimposed on the vowel-related movement of the tongue dorsum that continues throughout all the consonantal gesture. In the /a/ vowel context the tongue body movement is predominantly determined by the required position for consonantal lateral closure at both sides. On the other hand, the /i/ vowel raises the tongue body higher than necessary for the given consonantal constriction, so that the degree of elevation varies with the consonant type. The raising of the post-dorsum for /u/ appears to be less influential for the tip/blade gesture than that of medio- and post-dorsum for /i/. Rather, the post-dorsum gesture often provides good support, or stability, for the tip/blade activity to approach the front closure.

Furthermore, the EPG palatograms at the three temporal intervals imply that the coarticulatory effects extend beyond the consonantal articulation up to the acoustic onset

\(^1\) Note in passing that the tongue body is much more constrained in the production of a trill [r] relative to a tap (see Recasens (1991) and Recasens & Pallarès (1999) for detail).
(TR1) or onset (TR2) of the contextual vowels. It seems that the increase (or decrease) of dorsum movements gives shape to the target vowel gesture ($V_2$) during articulatory closure and continues to adjust the location after articulatory release of a given consonant. This characteristic agrees with the results reported by Butcher & Weiher (1976) for German, Fametani et al. (1985) for Italian, and Recasens (1984) for Catalan; but disagrees with Gay (1977) for American English.

Both anticipatory and carryover coarticulation appear spatiotemporally during the given coronal articulation. This contrasts with the claim made by Parush et al. (1983) on /k/ and /g/ in American English: the anticipatory effects are primarily temporal but the carryover effects are primarily spatial. Direct evidence supporting our view was found in the result of tap articulation. We have seen that in the symmetric context of /i/ the degree of dorsum elevation is reduced, or relaxed, before the tongue tip/blade starts its flicking (or transient) activity against the palate. This relaxation (or suppression) continues up to the acoustic onset of $V_2$, where the tip/blade gesture is already completed (see Figures 4-26 and 4-27). The same characteristic was also found in the asymmetric contexts. Furthermore, these results are not in complete agreement with the common assumption that carryover coarticulation, as opposed to anticipatory coarticulation, is mechanical: the articulator automatically accommodates the features of a foregoing phonetic segment essentially in the spatial domain (Gay, 1977; Kent & Moll, 1972; Recasens, 1984). Rather, the EPG evidence reveals that carryover coarticulation partly results from a co-production strategy deliberately regulated by a speaker.

The results of the regression analysis suggest that the parameter values are not only the distinguishing property of the individual consonants but also the shared property of all the coronal stop consonants in question. The three correlations between articulatory parameters were examined: (a) the duration of closure against that of consonantal articulation, (b) the closure duration against the front contact degree, and (c) the front contact degree against the central contact degree. It was found that each consonant has its own distinctive values for each parameter and the $r^2$ value greatly varies consonant-specifically. However, when the consonants are considered as a set, the contrasting values of the given parameters generally correlate to form a continuum. This view agrees with that proposed by Lindau (1985), who analysed r-sounds such as [ɾ, r, r, χ, k, j, ɾ] acoustically and attempted to relate them under the cover term ‘rhotics’. She suggests that ‘there is no physical property that constitutes the essence of all rhotics. Instead, the relation between members of the class of rhotics are more of a family resemblance (166)’. If we assume that the linear correlations above contribute to the underlying unity of the category, differing degrees of the $r^2$ value help us to infer the factors underlying the unity of coronal stop articulations.

As Lindau (1985) observed for the rhotics, it was found for the coronal stop consonants examined that: each member of the class of the coronal consonants allows a
resemblance to some other members with respect to some articulatory parameters. The two plosives and an alveolar nasal show similarities in the closure-to-consonantal duration, but differ in lingual configuration and intergestural coordination. A tap, an alveolopalatal nasal, and a voiceless plosive are alike in the pattern of intergestural coordination, but the control parameter varies with each consonant. A voiced plosive resembles a tap in front contact degree and lingual configuration, but the two consonants are different from each other in terms of the closure-to-consonantal duration and the intergestural coordination pattern. Beyond the properties of lingual parameters, the consonants \([t, d, n, r]\) are categorised by voicing and nasality parameters and by general articulation characteristics, obstruent and resonant.

These results have important consequences for the feature specification of the consonants in question. Although the consonants \([t, d, n, r]\) are specified by the articulator showing tongue-palate contact, they are precisely described in terms of the distinctive patterns of intergestural coordination. The consonant \([r]\), in particular, is better characterised by referring to the relative inactivation of the tongue dorsum. It is plausible to assume that this characteristic is a crucial prerequisite to generate the 'momentariness' of the tip/blade articulation. This constraint on the dorsum activity enables us to find out the similarities and differences between \([r]\) and \([d]\) or \([t]\).

Finally, some considerations may be made for \([r]\) in American English, based on what we have observed about the Japanese \([r]\). Stone & Hamlet (1982), using various sentences replacing all the syllables with /da/ (i.e. reiterant productions), examined the pattern of linguopalatal contact and the movement of the lower jaw. The various realisations of /d/ were classified into four categories: (type 1) voiceless/partially voiced /d/; (type 2) normal /d/; (type 3) short /d/ (voiced or voiceless); and (type 4) flap-like /d/. For the tongue-palate contact patterns in particular, Stone & Hamlet found that the area of tongue contact during unstressed CVs decreased in the order type 1>type 2> type 3> type 4; and that the tongue was retracted, to some degree, for the production of short /d/ and flap-like /d/ (i.e. the types 3 and 4). These results, as Stone & Hamlet imply, suggest that the flap in American English constitutes a continuum to voiced and voiceless alveolar stops. This view is clearly an analogy from the flapping rule in the unstressed intervocalic position. If we consider the flap in American English as produced by the same mechanism as \([r]\) in Japanese, then it is reasonable to expect that the depression of the tongue dorsum may occur. This activity may vary between the two languages since there are differences in the aspects of prosodic categories (e.g. stress vs. mora, syllable structure). Alternatively, the two taps/flaps might be produced by completely different mechanisms. The ideas remain to be proved.

It is reasonable to conclude that our findings reflect a language-specific property of how the control strategies are organised in the production of coronal stop articulations in
Japanese; and that the examination of systematic coarticulatory effects gives us important insights into the linguistically contrastive events in the vocal tract. The different degrees of interdependency between the two components of the tongue suggest that there are other kinds of distinctiveness at the physical level: the tongue region that does not make contact with the palate actively contributes to the significant activities of the tongue region that forms a contact with the palate. The important consequences of this view will be presented in chapter 7 with the example of the gestural analysis of /r/ variability in child and adult phonology.
Chapter 5

Lingual Articulation and Coarticulation of Voiceless Fricatives [s, ç, ç] and Affricates [ts, tç]

5.1 Introduction
This chapter deals with the articulatory gestures involved in the two different, but interconnected, sound categories: lingual fricatives [s, ç, ç]; and affricates [ts, tç] in Japanese. These consonants also share certain commonalities with stop articulations that we discussed in the previous chapters. The principal aim of this chapter is to provide an articulatory and coarticulatory characterisation of the above consonants as a function of vowel contexts.

Section 5.2 reviews previous descriptive and experimental research on [s, ç, ç] and [ts, tç] in Japanese and in other languages. The production requirements, or speakers' control strategies, for lingual activities are discussed and some hypotheses to be examined in the EPG experiments are formulated. Section 5.3 describes analysis procedures specific to the fricatives and the affricates. The experimental results are presented separately for each group of the consonants: the fricative articulations are discussed in section 5.4; and the affricate articulations are in section 5.5. In section 5.6 we attempt to characterise the two consonantal articulations in terms of the pattern of intergestural coordination; the control strategies, by which the two components of the tongue are organised. Section 5.7 concludes the chapter with a brief summary of our findings.

5.2 Previous Studies
5.2.1 Fricative Articulations
It is generally agreed that the production of fricatives involves complex motor control. Contrasting fricative articulations with stop articulations, Ladefoged & Maddieson (1996; 137) state that:

...A stop closure will produce more or less the same sound as long as it is complete, irrespective of whether there is firm or light contact. But in a fricative a variation of one millimetre in the position of the target for the crucial part of the vocal tract makes a great deal of difference. There has to be a very precisely shaped channel for a turbulent airstream to be produced...[I]n many fricatives, particularly sibilants, an exactly defined shape of the vocal tract has to be held for a noticeable period of time.

As far as the lingual gestures are concerned, 'the precision' is required for adjusting a typical positioning of the (whole) tongue and a formation of a narrow channel. By finely controlling these articulatory parameters, a particular configuration specific to a given lingual fricative is
generated. The importance of the transverse shape of the tongue is reflected in dichotomous expressions such as narrow and wide, grooved and slit, strident and mellow (e.g. Catford, 1977; Jakobson, Fant, & Halle, 1952). The close approximation causes the turbulent airflow regardless of voicing characteristics (Catford, 1977; Laver, 1994).

The above articulatory and acoustic properties specific to the target fricatives \([s, ç, ç]\) in our experiment are defined by the feature specifications in (1) below. These are orthodox representations (e.g. Keating, 1988a,b, 1991; Recasens, 1990) and irrelevant details are omitted.

\[
\begin{align*}
& \text{(1a) \quad [s]} \\
& \quad \text{Coronal} \\
& \quad \quad \text{[+anterior] \ [+strident]} \\
& \quad \quad \text{[-anterior]} \\
& \quad \quad \text{[+distributed]} \\
& \quad \quad \text{[+strident]} \\
& \quad \quad \text{[+high]} \\
& \quad \quad \text{[+strident]} \\
& \quad \quad \text{[-back]} \\
& \text{(1b) \quad [ç]} \\
& \quad \text{Coronal} \\
& \quad \quad \text{[+anterior] \ [+distributed]} \\
& \quad \quad \text{[+strident]} \\
& \quad \quad \text{[+high]} \\
& \quad \quad \text{[+strident]} \\
& \quad \quad \text{[+high]} \\
& \quad \quad \text{[+strident]} \\
& \quad \quad \text{[+strident]} \\
& \quad \quad \text{[+high]} \\
& \quad \quad \text{[+strident]} \\
& \quad \quad \text{[+back]} \\
& \text{(1c) \quad [ç]} \\
& \quad \text{Coronal} \\
& \quad \quad \text{[+anterior] \ [+distributed]} \\
& \quad \quad \text{[+strident]} \\
& \quad \quad \text{[+high]} \\
& \quad \quad \text{[+strident]} \\
& \quad \quad \text{[+high]} \\
& \quad \quad \text{[+strident]} \\
& \quad \quad \text{[+high]} \\
& \quad \quad \text{[+strident]} \\
& \quad \quad \text{[+back]} \\
\end{align*}
\]

The consonants \([s]\) and \([ç]\) are sibilants with the feature \([+strident]\) and belong to the class of coronal consonants. They contrast with each other in anteriority and in the presence (or absence) of the dorsal component. The consonant \([ç]\) is a non-sibilant, contrasting with \([s, ç]\) in stridency (i.e. \([-strident]\)), and is one of the dorsal consonants. Thus, the three consonants systematically differ from each other in terms of the place features that are defined by the active articulator, namely the tongue. The articulatory correlates of these phonological distinctions are the area and configuration of tongue-palate contact, the parameters that require the greater precision in their controls.

These observations support two hypotheses: (i) narrow channel width and constriction location can be considered as a combined articulatory parameter for \([s, ç, ç]\); and (ii) the lingual fricatives strongly resist the coarticulatory effects of the surrounding vowels. However, to my knowledge, there has not been any systematic investigation of the articulatory dimensions of the fricative production in Japanese. The lack of empirical evidence bedevils articulatory descriptions of Japanese \([s, ç, ç]\). We shall discuss the above two hypotheses based on the results of previous studies in other languages.

---

1) The voiceless postalveolar fricative \([ç]\) such as in English is specified by the same coronal component as in \([ç]\) without the dorsal component (Keating, 1988). This implicitly suggests that the alveolopalatal \([ç]\) is a 'palatalised' version of the postalveolar \([ç]\).
The extent to which the channel width is distinguished between the contrasting fricatives is dependent upon the constriction place for a given fricative. This hypothesis was examined by Fletcher & Newman (1991). Using the EPG technique they analysed the production of [s] and [j] spoken by two speakers of American English. It was found that there were regularly recurring channel widths and positions for the sounds: the [s] had a 6-8mm groove formed at the anterior region of the alveolar ridge by one speaker and at the posterior region of the ridge by the other; the [j] had a 10-12mm groove at around the corner of the alveolar ridge by both speakers. In the perceptual experiment varying the groove widths and places, Fletcher & Newman found that, while the turbulent noise generated at the anterior edge of the alveolar ridge was judged as [s], that formed at the posterior edge of the ridge was heard as [j]; and that, while noise created through the 6mm width was judged as [s], that from the 10-12mm was perceived as [j]. Accordingly, Fletcher & Newman (1991; 857) concluded that the groove widths were useful as targets for contrasting [s] with [j], even though there was an interactive effect between the width and place. Dagenais et al. (1994) report similar results in their EPG study: the groove width of [s] is consistently narrower than that of [j]. Dixit (1995) also shows the same tendency for Hindi [s] and [j]. Daganais & Critz-Crosby (1991) found that groove widths of [s] did not vary with different vowel contexts. These results for [s] and [j] generally support the hypothesis that the groove widths and places in lingual fricatives correlate linearly: the width becomes broader as the place shifts backwards.

However, the assumption of a linear correlation between the two parameters is not always proved. Lindblad & Lundqvist (1995), using EPG, examined the dental /s/, alveolopalatal /ç/ and retroflex [ç] in Swedish. It is reported that the groove width of /s/ is generally narrower than those of /ç/ and [ç]; and that the grooves tend to be similar between /ç/ and [ç]. This tendency, as Lindblad & Lundqvist mention, is related to vowel contexts: the variation in [ç] is greater than that in /s/ and /ç/. Furthermore, /ç/ and [ç] were found to have the same anterior end position. This suggests that the groove place of /ç/ and [ç] may not be a primary distinguishing articulatory feature. Ladefoged & Wu (1984) report similar results for the alveolar [s], retroflex [ç], and alveolopalatal [ç] in Pekingese. Their palatograms indicate that: (i) [s] has a narrower width at the backs of the upper teeth; (ii) the width and place of [ç] are broader and slightly further back than [s] but the width varies considerably; and (iii) the groove width of [ç] fluctuates between that of [s] and [ç], and the groove place moves noticeably along the alveolar ridge in the anterior-posterior dimension. In contrast to the variability in the two parameters, what can be seen from their lateral x-ray tracings is that there is a hollowing of the tongue in the [s] articulation; in [ç] the tip/blade component of the tongue is flat, rather than curled back, and is more bunched up. In [ç] the tongue medio-dorsum is raised and this effectively makes the pharyngeal cavity wider than [s] and [ç].
These conflicting results do not necessarily negate the hypothesis above: the groove width and place correlate linearly. It must be pointed out that, when a single speaker is considered, the groove width and place are fairly constant in the data of Lindblad & Lundqvist (1995) and Ladefoged & Wu (1984). Also, there is a difference between the inventory of coronal sibilants in (American) English and those in Swedish and Pekingese: while the sibilants in Swedish and Pekingese are crowded together along the alveolar ridge, those in English are less crowded. The intra-speaker consistency and the language-specific diversity suggest that speakers of a language selectively control the two parameters, groove width and place, in order to produce naturally sounding sibilants (Fletcher & Newman, 1991). Furthermore, the possible relationship between the sibilant and non-sibilant productions has not been examined. We thus extend the above hypothesis of the linear correlation to a non-sibilant lingual fricative and assume that the groove widths become broader in the order $[s]<[ç]<[ç]$ in Japanese.

The specification of sibilants (and lingual fricatives in general) needs to incorporate the overall shape of the tongue. As mentioned above, the characteristic hollow was found on the front of the tongue during the $[s]$ production: Ladefoged & Wu (1984) for Pekingese. The similar results were reported for British English (Ladefoged, 1957) and for American English (Stone, 1990). Stone (1991) examined the EPG and x-ray data of $[s]$ and $[ʃ]$ in American English, and found that: (i) during the production of $[s]$ the overall tongue body was grooved midsagittally up to the narrow channel; and (ii) during $[ʃ]$, in contrast, the posterior region of the tongue was grooved but the anterior region was ungrooved. Based on their EPG results of a bite-block experiment, Flege et al. (1988) suggest that, while a groove is critical for $[s]$ production, a complete closure is not critical for $[t]$. Stone et al. (1992) further examined the two consonantal articulations and found that the variations of cross-sectional tongue shapes were derived from the contact location specific to a given fricative. These results explicitly demonstrate that the tongue shape is a distinctive articulatory feature of the lingual fricatives. This has an important implication for our hypothesis above. Under the linear correlation hypothesis, the groove width is considered as a dependent variable of the groove place. The tongue shape can also be considered as a dependent parameter of the constriction location. Given this implicational relationship, it can be asked how these three parameters are organised systematically in the production of $[s]$, $[ç]$, and $[ç]$ in Japanese.

The requirements of the precise positioning result in a higher steadiness of the tongue shape during the production of the (lingual) fricatives (Ladefoged & Maddieson, 1996). The consequence of this is that the coarticulatory effects of the surrounding vowels exert less influence on the consonants. For the grooved fricative $[s]$ in particular there is articulatory evidence to show that, while the tip/blade position is relatively invariant, the
tongue dorsum is relatively free to coarticulate with the adjacent vowels (Carney & Moll, 1971; Dagenais & Critz-Crosby, 1991 for American English; Hardcastle & Clark, 1981 for British English; Hoole et al. 1993 for British English, French and German). Although there is little empirical evidence for the production of [ç] and [ç], it is supposed that, since they actively involve the tongue dorsum in their constriction formation, the coarticulatory effects are more constrained spatially and configurationally.

The relationship between the precise positioning of the tongue and susceptibility to the coarticulatory effects is especially noteworthy in the case of [s]. Previous articulatory descriptions of the voiceless fricative [s] in Japanese differ in the specification of the place of articulation. It is considered as lamino-alveolar by Akamatsu (1997), Kawakami (1977), Okada (1999) and Sakuma (1929): the tongue blade articulates against the alveolar ridge with the tongue tip either kept against the lower teeth, or raised towards the rims of the upper teeth. In addition to the lamino-alveolar place, Amanuma et al. (1978) list the apico-dental place, in which the rim of the tongue makes contact with the upper teeth. This discrepancy in the place specification has not been considered seriously. We must distinguish the apical or laminal point on the tongue on the one hand, from the dental or alveolar place on the palate on the other. These are independently controllable parameters. Nevertheless, it is in reality hard to decide just based on introspection whether the [s] articulation is consistently laminal with the tip static at the back of the lower teeth, or is constantly apical with the tip raised towards the backs of the upper teeth. Although Akamatsu (1997; 93f.) asserts that the positioning of the tongue tip is not crucial in the [s] articulation, there is a close link between the raising gesture of the tip and the advancement of the constriction place towards the dental region. Based on the empirical articulatory data, we have to inquire whether the variability in the constriction place is derived from contextual, or idiosyncratic, factor(s).

5.2.2 Affricate Articulations
Affricate articulation consists of a stop closure followed by a frictional release (e.g. Fletcher, 1992; Ladefoged & Maddieson, 1996). This dynamic characteristic has commonly been described by using two phonetic symbols and also by specifying two values of the manner feature, [continuant]. The representations of [ts] and [tc] in Japanese are given below.

\[
\begin{align*}
(2a) \quad & [ts] \\
& [-continuant] [+continuant] \\
& Coronal \\
& [+anterior] [+strident]
\end{align*}
\]

\[
\begin{align*}
(2b) \quad & [tc] \\
& [-continuant] [continuant] \\
& Coronal \quad Dorsal \\
& [-anterior] [+distributed] [+strident] [+high] [-back]
\end{align*}
\]
Stevens (1993; 39) recognises three acoustic and aerodynamic stages in the release of the affricate [tʃ], after the release of complete articulatory occlusion in the alveolar region. These are: (i) there is an initial brief transient with a spectral peak in the F4 region; (ii) about 40-50ms after the articulatory release the constriction has widened; and (iii) the frication noise increases in amplitude. Because these modifications are organised systematically, Stevens argues that the affricate cannot be considered as the fricative proceeded by the stop. The articulatory coordination between a stop gesture and a fricative gesture is 'closely knitted' and the friction in affricates has a shorter duration than that in fricatives (Gimson, 1980; 172). In addition, the friction of affricates has a more rapid rise time than that of fricatives (Howell & Rosen, 1983).

A potential articulatory correlate of the 'close-knit' character is attributed to the constriction location and the tongue shape. In the production of [tʃ], for instance, the tongue

---

2) Two points are worth mentioning in passing. There is a debate on the nature of affricates in phonological analysis, namely whether they are a single segment or a compound. The details can be summarised by two points (Hill, 1958; 36): (i) Bloomfield attempted to settle down the issue by showing the single-compound contrast between white shoes and why choose; and (ii) this was analysed by Bloch & Trager as /-t+s/- and /-+ ts/- respectively. After this there had been little agreement on the matter. Hill (ibid.) further states:

The controversy has at times seemed to be settled, but has each time broken out afresh. At an early period of phonetic study, the International Phonetic Association devised an alphabet for the writing of sounds, familiar to many students under the name “IPA”. The background of this alphabet was largely the sounds of French, and since in French a sound like that of chump is indeed a sequence, the IPA wrote the English sounds as a sequence. For long this analysis was accepted universally, and many older books on phonetics present it without discussion.

Hill himself lists five contrasts: white shoes /-t+s/-; courtship /-ts/-; why choose /-+ c/-; catch in /-c+-/; and kitchen /-c/- (Hill, op. cit. 37). These 'juncture' analyses do not help us to decide the matter, because either interpretation, single or compound, depends on phonetic contexts (Haugen & Twaddell, 1942). Crucial in the issue is the balance between the representation and physical reality (see Firth, 1948; Browman & Goldstein, 1990).

Another important implication for phonetic analysis is the difference between affricated stops and affricates. All stop consonants, in principle, can be released with audible friction. Thus, in British English, we find affrication of stops in words such as tea [tʰai], call [kʰou], Betty [betʰai] (vs. Betsy [bɛtsai]), Dick [dʰikʰ], and bad [bədʰ] (Cockney); snake [snaikʰ], short [ʃɔtʰ], daughter [dɔtʰa], back [bækʰ-bax] (Merseyside) (Wells, 1982; 322f. and 371f). Gimson (1980; 160) suggests that the main difference between the two lies in the 'brevity of the friction associated with the affricated plosives'.

In standard Japanese, it seems that affrication of stops is very rare. Yet, in the regional accent of the Kyushu area, a similar phenomenon to that found in English is observed: [mʰite]-[mʰiʃe] 'look (Imp.)'; [dʰeta]-[dʰeθa] 'came out'; [sʊ δe]-[sʊ dʰe] 'sleeve' (examples are from Kamimura (1975); phonetic transcriptions are mine). Notice here that the segment in question is palatalised as well as affricated. One possible explanation for this palatalisation with affrication is that in this accent, the tongue position of the /e/ vowel is closer to the palate than that in the other regional accents. At the time of releasing the complete occlusion, the tip/blade activity is coordinated already with the 'unusual' positioning of the tongue dorsum for /e/. Given this anticipatory activity, the higher position of the dorsum is maintained after the withdrawal of the tongue tip/blade from the alveolar ridge. Since a relatively close approximation remains in the oral cavity, the voiced, or voiceless, friction can be generated as the airflow runs through the constriction after the stop release.
tip/blade is raised and is pressed to form a complete occlusion along the alveolar ridge, 'a closed position in which the main part of the tongue is shaped nearly as for \( \text{ʃ} \)' (Jones, 1960; 160, §601). Consequently, as Jones' palatogram illustration reveals, the constriction location is slightly further back than that of an independent [t]: Jones notes that the involvement of the tip varies with speakers (Jones, op. cit. 161, §603). The same applies to [ts] (Jones, op. cit. 164, §618). In this case, however, there is no substantial change in the place of complete closure. Judging from x-ray tracings, Ladefoged & Wu (1984) state that at the point of maximum constriction, the essential tongue shapes of the affricates [ts, ðs, ðc] in Pekingese are very similar to those of the corresponding fricatives [s, ð, ç]. It appears that in these homorganic affricates the articulatory gesture for the second element, namely the fricative phase, is already planned: the closing location and the fricative release are united. Therefore, a seemingly contradictory characterisation can be derived: the fricative gesture is a prerequisite for the affricate articulation. This is clearly demonstrated by Hardcastle et al. (1995). They compared the stop and fricative phases of /tʃ/ with independent /t/ and /ʃ/, in the utterance produced by the speakers who have normal /t/ but abnormal /ʃ/ and /tʃ/. Hardcastle et al. showed that the constriction place of /tʃ/ was predictable from that of the fricative /ʃ/, and suggest that the main concern in therapeutic training should be the establishment of correct /t/ and /ʃ/.

The strict requirement that the fricative phase must be homorganic to the stop phase allows us to assume that the coarticulatory activities in the affricate articulation are similar to those in the fricative articulation. The stop phase of the target affricates [ts, ðc] in Japanese should differ from each other in the amount of contact in the posterior regions. This is because the second element [ç] involves the raising gesture of the tongue dorsum (see the diagram in (2b)). By the same token, the coarticulatory influences may be resisted strongly in the fricative phase of [tç]. Thus, we attempt to characterise the characteristic of 'close-knit' during the affricates by examining the two articulatory stages detected by EPG.

5.3 Data Collection and Analysis

5.3.1 Speech Items
The speech items consisted of VCV disyllabic (two-mora) words in which the consonants [s, ç, ç, ts, ðc] and the vowels /i, a, u/ were in all possible combinations. They were mostly nonsense words.

There are two points to note. First, the sequence of /si/ in this experiment corresponds to [s'ii]. This form is not found in native Japanese words but is common in the pronunciation of loanwords such as ['s'ii] 'sea', ['s'ิงฤ'] 'single', [i's'i'd'ii] 'CD' (see chapter 2 for the variability). We thus use the phonemicisation /sji/ for [çi] when necessary. Second, some combinations with [ts] rarely occur and can be considered as peripheral
phonetic forms. As mentioned in chapter 2, the affricate [ts] is analysed as an allophone of /t/ before /u/: /tuki→[tsuɪ,kʰi] 'moon'; /katu→[ˈkatsu̯] 'win'. Actually the consonant does occur before other vowels in the pronunciation of loanwords: [ˈtottsii] 'Tootsie', [fiˈrentse] 'Firenze', [ˈpi̯ittsa(ː)pʰ(d)za] 'pizza', and [kanˈtsoone] 'canzone'. In addition, there are some examples from native Japanese words: [oˈtoosan]→[oˈtottsan] 'father/dad', [goˈtçisoo]→[gotˈtsoo] 'feast' (Hattori, 1960). These pronunciations are considered as non-standard forms. Furthermore, we find an example in connected speech: [naˈtsu o (ˈmatsu)]→[naˈtsuu (ˈmatsu)] 'summer + object maker ‘o’ + wait (I will wait until summer comes)' (Kawakami, 1977). The form [tso] is derived from one of common connected speech processes, namely vowel coalescence).

5.3.2 Measurements
The configurational characteristics were studied at the point of maximum linguopalatal constriction or narrowing. For the fricatives [s, ç, ç] the EPG frame corresponding to the mid-point of the continuous noise energy, specified with acoustic recordings simultaneously made, was regarded as the MAX frame. For the affricates [ts, tç] the contact information was measured at three temporal intervals. First, the MAX point was the frame showing the maximum number of on-electrodes during the complete occlusion for the stop portion of the affricate. Second, the release point was the first frame in which the complete closure was broken up. Third, the frame corresponding to the onset of the friction was specified by the point in the acoustic recordings: after the release burst, the rapid rise of the friction noise began on the spectrogram. Contact configurations measured at the above temporal points were averaged over five repetitions and were presented in the form of EPG prototypical palatograms.

The width of the fricative channel for [s, ç, ç] was measured in millimetres. Although a common method is to count the actual number of off-electrodes at the place of major constriction (e.g. Engstrand, 1989), this measure is not an appropriate indicator of the constriction width. Because the spacing between the electrodes on the artificial palate (for the Reading EPG system) varies with reference to anterior and posterior regions, the same number of un-contacted electrodes on different rows does not imply the same width (Pandeli, 1993). We, therefore, used the measure in millimetres, a method similar to that in Pandeli

3) Another possible interpretation concerns speaker's knowledge of vowel devoicing and gestural overlap in connected speech. It is commonly supposed that the high vowels in an accented syllable are not usually susceptible to vowel devoicing. However, it is possible to assume that a speaker analogously applies the rules of vowel devoicing to the last vowel of [naˈtsui]. The consequence of this is 'gestural re-organisation' (Browman & Goldstein, 1991): the vowel gesture for [ə] is timed with the affricate [ts] and masks the gesture for the potentially devoiced [ui]. The advantage of this account is to capture the connected speech phenomenon as a gradient process, rather than categorical.
(1993). The measurement took two steps. First, after identifying the MAX frame for each of the five repetitions produced by a speaker, the minimum number of off-electrodes and its position in any one row were inspected. This inspection also involved the specification of the midpoint of the major constriction of a given consonant. Then, the width of the fricative channel was estimated for each token by measuring the distance between the adjacent electrodes on the row. Because of the permissible 'articulatory asymmetry' (Hamlet et al. 1986), the point of the maximum narrowing for [s] in particular tended to skew to the left for both speakers (who are right-handed). In such a case the distance between the adjacent electrode and the edge of the EPG palate was measured. This measurement can be illustrated by the example of the hypothesised MAX frame of the consonant [s] given in Figure 5-1, where on-electrodes are indicated by ‘0’ and off-electrodes by a dot.

\[
\begin{array}{cccccc}
1 & 12345678 & 00 \cdots 00 & 00 \cdots 00 & 00 \cdots 00 & 00 \cdots 00 \\
2 & 0 \cdots 00 & \text{dental-velar} & \text{alveolar} & \text{postalvelar} & \text{postalvelar} \\
3 & 0 \cdots 00 & \text{post-velar} & \text{palatal} & \text{post-palatal} & \text{postalvelar} \\
4 & 0 \cdots 00 & \text{postalvelar} & \text{palatal} & \text{post-palatal} & \text{postalvelar} \\
5 & 0 \cdots 00 & \text{post-palatal} & \text{palatal} & \text{post-velar} & \text{postalvelar} \\
6 & 0 \cdots 00 & \text{post-palatal} & \text{palatal} & \text{post-velar} & \text{postalvelar} \\
7 & 0 \cdots 00 & \text{postalvelar} & \text{palatal} & \text{post-velar} & \text{postalvelar} \\
8 & 0 \cdots 00 & \text{post-palatal} & \text{palatal} & \text{post-velar} & \text{postalvelar} \\
\end{array}
\]

\text{FRONT} \quad \text{CENTRAL} \quad \text{BACK}

\text{FIGURE 5-1: ARRANGEMENT OF ELECTRODES AND THREE REGIONS}

In this particular case the major constriction is formed at row 1 and there are two off-electrodes on that row. The constriction width is estimated by measuring the distance between the midpoint of the electrodes in column 3 and column 6, not by simply counting the number of off-electrodes. The measurement of the constriction width was made for the symmetrical sequences.

The degree of V-to-C coarticulatory effects was examined in the same way that was used in the previous experiment (see chapter 4, section 4.2.2). As shown in Figure 5-1 above, the surface of the artificial palate was divided into three regions, front, central and back. The percentage values of on-electrodes were computed separately for those three regions. For the fricatives [s, c, ç] the V₁ fixed sequences were used for quantifying the anticipatory effects of the V₂ upon the consonantal articulation, and the V₂ fixed sequences were used for analysing the carryover effects of the V₁ upon the given target consonant. This method, however, is not appropriate for quantifying the contextual effects upon the affricates [ts, te]. This is because the consonants involve the two articulatory stages, namely the stop element followed by the fricative element. Thus, the effects of the preceding vowel (V₁) was measured at the MAX point during the complete occlusion of the stop portion and those of the following vowel (V₂) were measured at the frame corresponding to the onset of the fricative portion.
5.4 Results of Experiment 1: Fricative Articulations

5.4.1 Configurational Characteristics of [s, ç, ç]

5.4.1.1 [s]

The EPG palatograms of [s] in the vowel symmetrical contexts for the two speakers are given in Figure 5-2. The consonant involves dental articulation in which the rim of the tongue touches the inner surface of the upper central incisors. A narrow constriction is formed in the dentalveolar region with accompanying side contact. This lateral seal sometimes becomes incomplete because the contact is made between the side of the tongue and the upper teeth ([asa] tokens produced by speaker MN). There are noticeable differences in the constriction formation between the speakers: apical realisation for speaker MN; and apicolaminal realisation for speaker TM. The involvement of the tongue tip/blade for the narrow constriction is substantially larger for speaker TM than for speaker MN. In addition, the two realisations differ in the amount of lateral contact. It appears that differences between apical and apicolaminal realisations are reflected, not only in the contact pattern of the major constriction, but also in that of the tongue overall.

(a) Speaker MN

(b) Speaker TM

Figure 5-2: Linguopalatal Configurations At The Max Point For [s] In The Vowel Symmetrical Contexts: ‘black’ 100-80%; ‘grey’ 79-60%; ‘striped’ 59-40%; ‘dotted’ 39-20%; ‘white’ 19-0%.

Apical and apicolaminal realisations manifest different coarticulatory effects from the surrounding vowel gestures. The EPG palatograms show that the apicolaminal realisation of [s] is more stable than the apical realisation. In Figure 5-2 (a), apical realisation of [s], the width of the narrow constriction at row 1 increases in the order /a/</u/<i/. Because the
tongue dorsum is raised and fronted for the vowel /i/, the tongue tip (blade) is consequently advanced and pressed against the upper incisors: the narrow constriction is actually made between the tongue tip and the upper incisor, rather than the hard palate. A similar effect is also observed in apicolaminal realisation of [isʰi] in Figure 5-2(b), but it is less extensive. These results are compatible with the EPG data of Catalan and Italian [n] reported by Recasens et al. (1993): apical realisation is more susceptible to coarticulatory effects. In addition, the characteristic matches with the view that the tongue tip is more movable than the tongue blade (Blandon & Nolan, 1977).

Although Amanuma et al. (1978) propose the two distinctive places of articulation, dental and alveolar, such a natural division on the palate may not help us to interpret the variability of the constriction place in the EPG data above. The tongue shape, apicality or laminality, and its susceptibility to the coarticulatory effects are more important. For both speakers in this experiment the narrow constriction is formed by raising the tongue tip with the rim of the tongue touching the back of the upper teeth. While the apicality, or laminality, is speaker-dependent, the apical realisation of [s] is more susceptible to the contextual effects. Consequently, the narrow constriction for speaker MN is fronted, or more 'dentalised', in the high vowel context and the groove width becomes broader. Therefore, as far as our EPG data is concerned, the variability of the constriction place is considered as contextual.

In spite of such realisational differences, the posterior contact generally reflects the height and backness of the tongue dorsum specific to the contextual vowels. The amount of contact increases in the order /a/ < /u/ < /i/). It must be noted that the central area on the palate is kept un-contacted in all the vowel contexts. This contact pattern, in turn, suggests that there is a hollow on the tongue, a characteristic frequently reported in previous descriptive and experimental studies (Ladefoged, 1957; Ladefoged & Wu, 1984; Stone, 1990, 1991; Stone et al. 1992).

5.4.1.2 [ç]
Figure 5-3 shows the EPG palatograms of [ç]. The narrowing constriction is primarily made in the postalveolar (rows 3-4) and pre-palatal (row 5) regions. This is accompanied by a larger amount of contact in the posterior regions. This suggests that both tongue blade and (medio-)dorsum are involved in the constriction formation, with the tongue tip always directed downwards. Therefore, the consonant is characterised as a voiceless lamino-dorso alveolopalatal fricative. Realisation of the constriction length is different for each speaker. It is consistently realised at rows 3-5 for speaker MN but varies within the range of rows 2-5 for speaker TM. However, it must be noted that the constriction shape in the anterior regions is essentially the same for each of the speakers.
(a) Speaker MN

(b) Speaker TM

FIGURE 5-3: LINGUOPALATAL CONFIGURATIONS AT THE MAX POINT FOR [ç] IN THE VOWEL SYMMETRICAL CONTEXTS: ‘black’ 100-80%; ‘grey’ 79-60%; ‘striped’ 59-40%; ‘dotted’ 39-20%; ‘white’ 19-0%.

The coarticulatory effects of the surrounding vowels are largely indistinct. This is true particularly for speaker TM. This characteristic comes from the articulatory nature of the consonant [ç]: it involves the tongue dorsum for the formation of the constriction. For speaker MN the contact in the posterior regions decreases in the order /i/>/u/>/a/. However, this variability, as seen in the EPG palatograms, does not affect the essential contact pattern realised by the tongue dorsum. Therefore, it is possible to assume that the Japanese [ç] can be considered as a complex segment, in which both the tongue blade and dorsum are actively involved in the formation. The extent to which the two components of the tongue are controlled is a matter of controversy (Keating, 1988, 1993; Recasens, 1990; Recasens et al. 1993, 1994). The issue of articulatory complexity will be discussed in chapter 6.

5.4.1.3 [ç]

Figure 5-4 displays the EPG palatograms of [ç]. It involves palatal articulation, in which the tongue medio- and post- dorsum is raised towards the central and back regions. The greatest narrowing constriction is typically formed at rows 7-8, with the lateral contact extending up to row 2, across the three vowel contexts. The overall tongue body is supposed to be convex to the palate and the tongue tip is always directed downwards. Although the contact pattern and configuration is similar to those of [i] and [j], the three gestures typically differ in the degree of coarticulatory resistance. We will return to this issue in section 5.4.2.3 later.

The EPG palatograms show that the effects of the contextual vowels are readily reflected in the consonantal articulation: the amount of tongue-palate contact decreases in the
order /i>//u>//a/. Because the consonant [ç] involves the raising gesture of the tongue dorsum that is certainly used for the vocalic gesture, the height and backness specific to the vowels are directly related to the contact pattern of [ç]. Therefore, the amount of contact is larger in the sequence [içi] where, in principle, the continuous raising of the tongue body is required, than in the sequence [aça] where the lower positioning is required. The fact that the amount of contact in [aça] for speaker TM is the smallest of the three contexts can be explained by the earlier lowering, or the undershoot of the raising gesture, of the tongue dorsum, a positioning required by the /a/ vowel. The consequence of this accommodative activity is a widening of the typical side-to-side width at rows 7-8.

![Diagram of tongue positions](image)

(a) Speaker MN

(b) Speaker TM

**Figure 5-4**: Linguopalatal Configurations at the Max Point for [ç] in the Vowel Symmetrical Contexts: ‘black’ 100-80%; ‘grey’ 79-60%; ‘striped’ 59-40%; ‘dotted’ 39-20%; ‘white’ 19-0%.

### 5.4.1.4 Constriction Widths of [s, c, ç]

Turning now to the measurements of the constriction width, we examine how it correlates to the constriction places specific to each of the target consonants. Table 5-1 summarises the mean constriction widths in millimetres and standard deviations. It reveals that the groove width is greater in [ç] than in [s] and [ç]: the two coronal consonants exhibit almost the same width. The results of one-way ANOVA, comparing differences in means of the measured widths, show a significant effect of the consonant type: [F(2,42)=59.40, p<0.0001] for speaker MN and [F(2,42)=103.00, p<0.0001] for speaker TM. Pair-wise comparisons between the particular consonants were conducted by t-test, yielding the relation [ç]>[s]=[ç] for both speakers: in the tokens of speaker MN, [t=7.24, p<0.0001, df=28] for [ç] vs. [s]; [t=1.64, p=0.055, df=28] for [s] vs. [ç]; in the tokens of speaker TM, [t=10.40, p<0.0001, df=28] for
[ç] vs. [s]; [t=0.10, p=0.45, df=28] for [s] vs. [ç]. In contrast, as the values of standard deviation suggest, the fricative channel width is more stable in [ç] than in [s] and [ç].

<table>
<thead>
<tr>
<th>SPEAKER MN</th>
<th>CONSONANTS AND CONSTRICION WIDTH (in mm.)</th>
<th>[s]</th>
<th>[ç]</th>
<th>[ç]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>/i_i/</td>
<td>15.85</td>
<td>1.64</td>
<td>11.50</td>
<td>0.00</td>
</tr>
<tr>
<td>/a_a/</td>
<td>11.00</td>
<td>0.00</td>
<td>11.50</td>
<td>0.00</td>
</tr>
<tr>
<td>/u_u/</td>
<td>11.00</td>
<td>1.37</td>
<td>11.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Pooled-Vowel Context</td>
<td>12.62</td>
<td>2.62</td>
<td>11.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Range</td>
<td>9.50-17.50</td>
<td>11.50</td>
<td>15.00-25.00</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPEAKER TM</th>
<th>[s]</th>
<th>[ç]</th>
<th>[ç]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
</tr>
<tr>
<td>/i_i/</td>
<td>9.50</td>
<td>2.21</td>
<td>7.50</td>
</tr>
<tr>
<td>/a_a/</td>
<td>8.50</td>
<td>1.54</td>
<td>9.90</td>
</tr>
<tr>
<td>/u_u/</td>
<td>8.30</td>
<td>1.92</td>
<td>8.70</td>
</tr>
<tr>
<td>Pooled-Vowel Context</td>
<td>8.77</td>
<td>1.85</td>
<td>8.70</td>
</tr>
<tr>
<td>Range</td>
<td>6.00-12.50</td>
<td>7.50-10.50</td>
<td>18.00-35.00</td>
</tr>
</tbody>
</table>

Table 5-1 Means and Standard Deviations of Constriction Width for [s], [ç], and [ç].

Figure 5-5 presents the interdependency between the constriction width and the constriction place, with the data of the two speakers pooled. For [s] and [ç] the similar thin clustering in the vertical dimension occurs, while its range and variability are significantly different: the greatest value of [s] overlaps the smallest value of [ç]. In contrast, the plots of [ç] exhibit a tight clustering, all the plots of which are within the range of variations of [s]. These distinctive patterns of clustering suggest that the different degrees of the precision are required for the shaping activities of the tongue. The significant $r^2$ value suggests that about 67% of the variation between the major narrowing widths is due to their location on the palate, in the sense that the width tends to increase linearly as the location shifts backwards.

![Figure 5-5: Regression Analysis Of The Constriction Place and Width Across Speakers](image)

FIGURE 5-5: REGRESSION ANALYSIS OF THE CONSTRICION PLACE AND WIDTH ACROSS SPEAKERS

X-axis=EPG rows; Y-axis=measures of the constriction width in millimetre.

211
Although a tolerable $r^2$ value is obtained, it is questionable whether a simple linear correlation can be assumed for the way in which the three consonants differ in their places of articulation. Whereas a higher degree of constancy was observed for the channel width within each consonant group, a greater degree of variability was also observed, particularly for [s] and [ç], across speakers and vowel contexts. This variability uncovers the fact that an identical, or nearly identical, size of channel width can be attained by the activities of the functionally different regions of the tongue, namely: the tip/(blade) for [s]; the blade for [ç]; and the dorsum for [ç]. This suggests that the precise distinction of the groove width is not important for contrasting the fricatives in question.

In contrast, what are consistently maintained are the constriction location and the overall shape of the tongue. As observed in the EPG palatograms, the anterior space of the palate expands as the constriction location shifts backwards: [s]<[ç]<[ç]. This widening is associated with the tongue shape. The tip is directed upwards in the [s] production and a typical hollow is formed along the midsagittal line on the tongue. A wider space of the anterior region is obtained for [ç], in which the tip is constantly directed downwards and the anterior of the tongue is flat. Because of raising and fronting of the blade and dorsum, the pharyngeal cavity is also expanded (Ladefoged & Wu, 1984). Further widening is observed for [ç], in which the dorsum is raised upwards and the anterior portion of the tongue is moved backwards. Thus, the expansion of the anterior space on the palate is associated with the backward shift of the constriction place.

This is compatible with the results reported by Heinz & Stevens (1961). In their perceptual experiment, in which the synthetic stimuli with various resonant frequencies within the range 2500-8000Hz were used, it was found that listeners’ responses shifted consistently in the order [ʃ]>[s]>[ʃ], as the resonant frequency increased. The modulation of resonance frequency can be associated with the changes of the front ‘cavity’ size that are manifested by the activities of the tongue and lip gestures. The spectrograms in Figures 5-6 and 5-7 illustrate a single production of each of the three fricatives [s, ç, ç] in the /a_a/ context. In these particular utterances it can be observed that: in [a,sa], the main energy of the resonant frequency is concentrated around the level of 6000Hz; in [a,ça], the range of frequencies is broader (approximately 4000-7000Hz: the strongest frequency location is around 5000Hz) and the amplitude is stronger; and in [a,ça], the frequency resonance is in a lower region, around 3000Hz. It is reasonable to suppose that the distribution of the frequency resonance is inversely related to the EPG contact pattern in the anterior region: the higher fricative noise is generated as the constriction moves forwards. These observations agree with those of Ladefoged & Maddieson (1996): lingual fricatives have a greater constancy in the tongue shape.
The sampling rate is 16kHz; frequency and amplitude at the midpoint of the friction noise were measured by a single spectral slice; (a) [a,sa] 5656Hz and 24dB; (b) [a,ça] 4656Hz and 42 dB; [aça] 3062Hz and 47dB.
The sampling rate is 16kHz; frequency and amplitude at the midpoint of the friction noise were measured by a single spectral slice; (a) [a,sa] 5750Hz and 20dB; [a,ca] 4906Hz and 38 dB; [a,ça] 2906Hz and 38dB.
5.4.2 Coarticulatory Characteristics of [s, ç, ç]

5.4.2.1 [s]

The EPG palatograms of [s] in the vowel asymmetrical contexts are given in Figure 5-8 below. The individual characteristics that were described in the previous section are observed for the groove width and the overall pattern of tongue-palate contact. Whereas the channel width varies greatly across the phonetic contexts in MN’s productions, that in TM’s is fairly stable and exhibits the larger amount of contact. The articulation is apical and apico-laminal respectively.

(a) Speaker MN

(b) Speaker TM

![Palatograms for [s] in different contexts](image_url)

**Figure 5-8:** Linguopalatal Configurations at the Max Point for [s] in the Asymmetrical Contexts: ‘black’ 100-80%; ‘grey’ 79-60%; ‘striped’ 59-40%; ‘dotted’ 39-20%; ‘white’ 19-0%.
Figure 5-9 shows mean percentage values of tongue-palate contact in the three regions for the two speakers. Differences between apical and apico-laminal realisations are clearly observed. In the apical realisation of [s] (MN), the amount of contact in the front region is smaller than in the apico-laminal realisation (TM).

(a) Speaker MN

(b) Speaker TM

FIGURE 5-9: MEAN PERCENTAGE VALUES OF LINGUOPALATAL CONTACT FOR [s] IN THE ASYMMETRICAL CONTEXTS: (a) speaker MN; (b) speaker TM; X-axis=the V/s/V sequences; Y-axis=measures of the contact amount (%)
It is worth noting that, when the two speakers are considered independently, the percentage values in the front region are not substantially affected by the contextual vowels. For the apico-laminal [s] (TM) the posterior two regions exhibit very similar values across the vowel contexts, but the values vary considerably in the apical [s] (MN).

A one-way ANOVA was conducted to test for differences in the amount of contact as a function of the flanking vowels in the anticipatory and carryover contexts. Pair-wise comparisons by t-test were also made to examine the relation between the particular vowels. The results are summarised in Table 5-2. It can be seen that, although significant differences reflect the realisational idiosyncrasies, carryover effects are larger than anticipatory effects. For speaker TM in all contexts anticipatory effects are not significant, except for the back region of the /iCV/ sequences, and there are no significant differences in the amount of contact across the vowel contexts. For speaker MN, in contrast, the effects are observed primarily in the front and the central region, yielding the relation /i/>/a=/u/. On the other hand, carryover effects were found for both speakers. In the front region, the contact amount significantly increases, when the preceding vowel is either /i/ or /u/ (MN). Yet, there are no significant differences in TM’s data. In the posterior two regions, the contact degree is larger in the high vowel contexts than in the low vowel context. This implies that the height of the tongue dorsum during the narrow constriction is associated with that of the preceding vowel. Because larger and systematic effects of the flanking V₁ are observed for both apical and apico-laminal realisations, it appears that carryover effects have clear predominance in the [s] production.

### Table 5-2: One-way ANOVA Results for Anticipatory and Carryover Effects on [s]

<table>
<thead>
<tr>
<th>Anticipatory Effects</th>
<th>Regions and F Values†‡</th>
<th>Carryover Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Anticipatory Effects</strong></td>
<td><strong>Regions and F Values†‡</strong></td>
<td><strong>Carryover Effects</strong></td>
</tr>
<tr>
<td><strong>Front</strong></td>
<td><strong>Central</strong></td>
<td><strong>Back</strong></td>
</tr>
<tr>
<td>/iCV/</td>
<td>i&gt;a=u</td>
<td>28.00**</td>
</tr>
<tr>
<td>/uCV/</td>
<td>i=a&gt;v</td>
<td>6.00*</td>
</tr>
<tr>
<td>/aCV/</td>
<td>i&gt;a&gt;u</td>
<td>26.00**</td>
</tr>
<tr>
<td><strong>Back</strong></td>
<td><strong>Central</strong></td>
<td><strong>Back</strong></td>
</tr>
<tr>
<td>/iCV/</td>
<td>i=a&gt;u</td>
<td>2.21</td>
</tr>
<tr>
<td>/uCV/</td>
<td>u&gt;i&gt;a</td>
<td>1.20</td>
</tr>
<tr>
<td>/aCV/</td>
<td>u&gt;=i &gt;u</td>
<td>0.11</td>
</tr>
</tbody>
</table>

†‡=p<0.05, **=p<0.01, unmarked = no significant

†The relation between the vowel pairs was assessed by unpaired t-test; the criterion for the significance was 0.01. The symbol > indicates a significant difference but the symbol = indicates no significant difference.
The EPG palatograms of [ç] in the asymmetrical contexts are shown in Figure 5-10 below. The frontmost row (row 1) is always uncontacted and the characteristic empty space is maintained in the front region. The contact pattern is quite consistent across the different vowel contexts when a single speaker is considered. The stability and relatively large amount of contact become clear if we examine the mean percentage values which are presented graphically in Figure 5-11.

(a) Speaker MN

(b) Speaker TM

Figure 5-10: Linguopalatal Configurations at the Max Point for [ç] in the Asymmetrical Contexts: 'black' 100-80%; 'grey' 79-60%; 'striped' 59-40%; 'dotted' 39-20%; 'white' 19-0%.

(a) Speaker MN; (b) Speaker TM
Figure 5-11: Mean Percentage Values of LingualPalatal Contact for \[\text{c}\] in The Asymmetrical Contexts: (a) speaker MN; (b) speaker TM.

X-axis=the V/sj/V sequences; Y-axis=measures of the contact amount (%)
The one-way ANOVA results in Table 5-3 reveal that far fewer vowel-dependent coarticulatory effects reach significance in the three regions. This suggests that the alveolopalatal [ç] requires the active involvement of both the tongue tip/blade and the dorsum.

As we discussed in chapter 4, similar results were obtained for the alveolopalatal [n]. This consonant also involves the two components of the tongue in its constriction formation. However, variability of the tip participation was observed, the quality ranging from [n] to [p] according to speakers and vowel contexts (see Figures 4-9-1 and 4-9-2). Furthermore, carryover effects systematically affect the positioning of the tongue body during the complete occlusion: contact in the posterior two regions was larger in the high vowel contexts than in the low vowel contexts (see Table 4-8). For [ç] there are a few instances that can be considered as height-dependent variability: the central region of the /aCV/ and /iCV/ sequences (TM), the central and back regions of the /VCa/ sequences (MN). However, these effects are not as systematic and extensive as those in the consonant [n]. In addition, the consonant [ç] exhibits a fairly constant pattern of linguopalatal contact across vowel contexts. It is therefore reasonable to suppose that, although the same place label specifies the two consonants, the articulatory gesture of [ç] more strongly resists coproduction with the preceding and the following vowel gestures than [n]. This is due to the difference in the manner of articulation, namely nasal and fricative.

### Table 5-3: One-Way ANOVA Results for Anticipatory and Carryover Effects on [ç]

<table>
<thead>
<tr>
<th>Anticipatory Effects</th>
<th>Regions and F Values†‡</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPEAKER MN</strong></td>
<td>FRONT (F(2,12))</td>
</tr>
<tr>
<td>/iCV/</td>
<td>i=u=a</td>
</tr>
<tr>
<td>/uCV/</td>
<td>u=a=i</td>
</tr>
<tr>
<td>/aCV/</td>
<td>i=u=a</td>
</tr>
<tr>
<td><strong>SPEAKER TM</strong></td>
<td>FRONT (F(2,12))</td>
</tr>
<tr>
<td>/iCV/</td>
<td>i=a=u</td>
</tr>
<tr>
<td>/uCV/</td>
<td>i=a=u</td>
</tr>
<tr>
<td>/aCV/</td>
<td>i=u=a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Carryover Effects</th>
<th>Regions and F Values†‡</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPEAKER MN</strong></td>
<td>FRONT (F(2,12))</td>
</tr>
<tr>
<td>/VCi/</td>
<td>a=i=u</td>
</tr>
<tr>
<td>/VCu/</td>
<td>a=u=i</td>
</tr>
<tr>
<td>/VCa/</td>
<td>a=u=i</td>
</tr>
<tr>
<td><strong>SPEAKER TM</strong></td>
<td>FRONT (F(2,12))</td>
</tr>
<tr>
<td>/VCi/</td>
<td>u=i=a</td>
</tr>
<tr>
<td>/VCu/</td>
<td>u=a=i</td>
</tr>
<tr>
<td>/VCa/</td>
<td>a=u=i</td>
</tr>
</tbody>
</table>

†*=p<0.05, **=p<0.01, unmarked = non-significant
‡The relation between the vowel pairs was assessed by unpaired t-test; the criterion for the significance was 0.01. The symbol > indicates a significant difference but the symbol = indicates no significant difference.
5.4.2.3  [ç]

The EPG palatograms of dorso-palatal [ç] were shown in Figure 5-12 below. The front region exhibits a typical empty space, the size of which is wider than the alveolopalatal [ç]. There are idiosyncratic differences in the overall amount of contact between the two speakers: the stable area (black region) is larger in speaker MN than in speaker TM. Yet, the essential contact pattern, or the tongue shape, is maintained.

Figure 5-13 (next page) presents mean percentage values of contact in the three regions for the two speakers. It can be observed that, while the contact amount in the posterior

(a) Speaker MN

(b) Speaker TM

Figure 5-12: Linguopalatal Configurations At The Max Point For [ç] In The Asymmetrical Contexts: ‘black’ 100-80%; ‘grey’ 79-60%; ‘striped’ 59-40%; ‘dotted’ 39-20%; ‘white’ 19-0%.
two regions is almost the same as, or even greater than, the consonant [ç], that in the front
region is noticeably smaller and varies, to some extent, with vowel contexts.

(a) Speaker MN

(b) Speaker TM

FIGURE 5-13: MEAN PERCENTAGE VALUES OF LINGUOPALATAL CONTACT FOR [ç] IN THE
ASYMMETRICAL CONTEXTS: (a) speaker MN; (b) speaker TM; X-axis=the V/s/V sequences;
Y-axis=measures of the contact amount (%)
Table 5-4 summarises the results of a one-way ANOVA that examined anticipatory and carryover effects of the contextual vowels upon the consonantal articulation. Few anticipatory effects reach significance in the three regions and the percentage values are largely indistinct across the three flanking vowels. In contrast, carryover effects are observed in the front and the central regions. The significant relation between the particular vowel pairs in these two regions differs between the speakers. This is primarily due to the idiosyncratic differences between the tongue contact patterns.

The forward extension (i.e. the contact in the front region) was observed for both speakers (see palatograms in Figure 5-12). This gesture is maintained when the V1 is the low vowel: the relation tends to become /a/>/i=/u/. Given the interdependency of the two components of the tongue, it could be supposed that the forward extension provides support for the raising of the dorsum, but it also establishes common ground for the articulatory gesture of [ç]. This can explain the frequently encountered observation that the distinction between [ç] and [ç] tends to be lost particularly in the Tokyo accent (e.g. Edwards, 1903; Kawakami, 1977): [çj, to] ‘human being’–> *[çito], *[çibatej], [çj,tçi] ‘seven’–> *[çi,tçi] ‘pawnshop’ (* = non-standard). In contrast, the overall amount of contact is larger in speaker MN than in speaker TM. This characteristic manifests different relationships between the particular vowels in the central region: the carryover effects of /a/ are weaker in TM but stronger in MN. These results suggest that the preceding vowel more systematically affects the activities of the tip/blade and medio-dorsum in the production of [ç].

<table>
<thead>
<tr>
<th>ANTICIPATORY EFFECTS</th>
<th>REGIONS AND F VALUES† ‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPEAKER MN</td>
<td>FRONT</td>
</tr>
<tr>
<td>/iCV/</td>
<td>u=a=i</td>
</tr>
<tr>
<td>/uCV/</td>
<td>u=i&gt;a</td>
</tr>
<tr>
<td>/aCV/</td>
<td>a=i=u</td>
</tr>
<tr>
<td>SPEAKER TM</td>
<td>FRONT</td>
</tr>
<tr>
<td>/iCV/</td>
<td>i=u=a</td>
</tr>
<tr>
<td>/uCV/</td>
<td>u=a=i</td>
</tr>
<tr>
<td>/aCV/</td>
<td>u&gt;i=a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CARRYOVER EFFECTS</th>
<th>REGIONS AND F VALUES† ‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPEAKER MN</td>
<td>FRONT</td>
</tr>
<tr>
<td>/VCi/</td>
<td>a&gt;i=u</td>
</tr>
<tr>
<td>/VCu/</td>
<td>a=u=i</td>
</tr>
<tr>
<td>/VCa/</td>
<td>a&gt;i=u</td>
</tr>
<tr>
<td>SPEAKER TM</td>
<td>FRONT</td>
</tr>
<tr>
<td>/VCi/</td>
<td>i=a=u</td>
</tr>
<tr>
<td>/VCu/</td>
<td>a&gt;i=u</td>
</tr>
<tr>
<td>/VCa/</td>
<td>i=u=a</td>
</tr>
</tbody>
</table>

†*p<0.05, **p<0.01, unmarked = non-significant
‡ The relation between the vowel pairs was assessed by unpaired t-test; the criterion for the significance was 0.01. The symbol > indicates a significant difference but the symbol = indicates no significant difference.
A few remarks should be made concerning the relationship between the palatal fricative [ç], palatal glide [j], and front vowel [i]. The segments [ç], [j] and [i] can be specified by the same constriction place on both the palate and the tongue, namely dorso-palatal. If we compare the EPG palatograms in Figure 5-12 above with those in Figure 6-2 in the next chapter, we find that the three contrasting segments exhibit essentially the same contact pattern at the moment of maximum constriction. The identification of their distinctiveness must incorporate other articulatory properties. Catford (1977; 120) distinguishes the ‘fricative stricture’ from the ‘approximant stricture’ on the ground that ‘fricatives have turbulent flow through the channel, whether they are voiceless or voiced, whereas approximants have turbulent flow only when voiceless’. In this sense, the high front vowel [i], as Catford suggests, is of the stricture-type approximant, and is transcribed as [j] when devoiced. In contrast, the palatal glide [j] is often transcribed as [ç] when devoiced. For instance, Gimson (1980: 212) exemplifies the consonant as an allophonic realisation of the sequence /h+/j/ in British English: ‘[t]he sequence of /h+/j/ as in hue /hju:/ [hçu:] may coalesce into [ç]...The number of words...is, however, restricted (Hugh, hew, hue, human, humour, etc.)...’ These observations suggest that constriction degrees, as well as the voicing distinction of the laryngeal gesture, play an important role in the characterisation of [ç], [j], and [i].

The palatal fricative [ç] incorporates some aspects of the articulatory nature of the palatal glide [j] and the high front vowel [i]. The two vocoids are similar to each other in terms of constriction degree but are different from each other in terms of gestural duration and coarticulatory resistance: the articulatory gesture of the vowel [i] is longer in duration and is more resistant to the coarticulatory effects of the surrounding vowels; and the glide [j] shows coarticulatory effects particularly in the front region (we will discuss these points in detail in the next chapter on palatalisation). On the one hand, the degree of constriction of the palatal fricative [ç] is often greater than that of the vowel [i]: the consonant does not sound like [i] even if it is voiced. On the other hand, coarticulatory effects on [ç] were found in the front and the central regions. These similarities and differences demonstrate that the articulatory distinction between vowel and consonant is not clear-cut, suggesting that the systematic articulatory controls underlying the two categories and their manifested patterns form a continuum.

5.5 Results of Experiment 2: Affricate Articulations

5.5.1 Movement Characteristics of [ts] and [tc]

We shall begin by inspecting the overall articulatory process of the affricate production. EPG full-printouts of [ts] and [tc] in the /a_a/ context are given in Figures 5-14 and 5-15 respectively. Several landmarks of the articulatory stages are also indicated.
(a) speaker MN

(b) speaker TM

It can be seen that the contact increases postero-anteriorly and the complete occlusion is formed at the alveolar region (rows 1-3). During the latter half of the closure the tip/blade component of the tongue begins to shift its shape preparing for the constriction of the fricative portion: the amount of contact in the alveolar region decreases postero-anteriorly. After breaking the complete occlusion there is a time interval (0-3 frames) where the activity of the tongue freezes. This is also observed in the production of [tʃ] in Figure 5-15. Fleethar (1989) reports this 'slowed-down', or 'even stopped', movement during the production of [tʃ].
in American English. At the time of the beginning of the friction the constriction in the alveolar region is further reduced in degree and progressively expands the empty space. It is supposed that the overall tongue forms a hollow midsagitally. The duration of the [s] in the affricate is shorter, so that the narrow constriction often becomes fairly wide at the temporal midpoint of the friction.

Figure 5-15 shows the EPG printout of [tç] produced by two speakers. While the approach phase is similar to that of [ts], the contact at the frontmost row (row 1) tends to be incomplete. The complete occlusion is formed at rows 2-4 (MN) and at rows 1-3 (TM). This variation is compatible with the contact pattern of an independent [e] (see Figure 5-10 above). At the time of approximately two-third progression of the closure, the contact pattern

(a) speaker MN

(b) speaker TM

![Figure 5-15: EPG Patterns of Japanese [tç] Produced by the Two Speakers: (a) Speaker MN; (b) Speaker TM. The points indicated are (a) the onset of complete closure; (b) MAX; (c) the offset of complete closure; (d) release; (e) the beginning of friction; (f) the temporal midpoint of friction; and (g) the beginning of the second vowel.](image-url)
shifts to the formation of the fricative constriction. In contrast to [ts], the contact decreases antero-posteriorly and the space in the front region is increasingly broadened. This preparatory activity is more clearly observed in MN’s utterance than TM’s. The directionality differences of contact withdrawal are due to the individual variation of the constriction location of the fricative. At the temporal midpoint of the friction the constriction shape (already) declines to anticipate the following vowel. Thus, the fricative duration of [tc], as with that of [ts], is shorter than that of a corresponding individual fricative.

For examining the configurational pattern in the vowel symmetrical contexts, three particular EPG frames were chosen: MAX of the complete closure; the closure offset (the last frame of the complete occlusion); and the fricative onset. The EPG palatograms, averaged over five repetitions, were made and are shown in Figure 5-16 for [ts] and in Figure 5-17 for [tc].

For the affricate [ts] the stop element is produced with a primary alveolar place of articulation. The length of the complete occlusion is formed at rows 1-3. This tends to be reduced to rows 1-2 and the frontmost row often shows incomplete closure in the sequence of [itsi] produced by MN. The stop constriction readily reflects effects of the preceding vowel (V1): the contact in the posterior two regions increases in the order [a]<[u]<[i]. On the other hand, the contact pattern at the point of the fricative onset generally reflects idiosyncratic characteristics, namely apical and apico-laminal, the realisational differences that we observed for the production of an independent [s] by the two speakers. As the black areas on the EPG palatograms suggest, the amount of contact varies with the contextual vowels in the apical realisation (MN): [a]<[u]<[i]. In contrast, the apico-laminal realisation (TM) exhibits less variation in the contact degree across the vowel contexts.

For the stop element of the affricate [tc] there is a substantial increase in the amount of contact and the retraction of the constriction location. Both speakers show that the contact at row 1 tends to be withdrawn in the /i_i/ sequence. The length of the central closure extends from row 1 (or 2) to row 4 (or 3). It appears that the configurational pattern of the stop occlusion is very similar to that of [n] (see Figures 4-4-1, 4-4-2, 4-9-1, and 4-9-2 in chapter 4). Yet, the stop element of [tc] reflects the coarticulatory effects of the surrounding vowels: the contact amount increases in the order [a]<[u]<[i]. The fricative element of [tc], on the other hand, reveals little variation of contact pattern across the vowel contexts. It must be noted that the constriction location extends slightly further forward in the /i/ and /a/ contexts than in the /u/ context (TM). A similar tendency was also observed for the same speaker in the production of an independent [ç] (see Figure 5-3 above). This slight advancement of the anterior edge of the narrow constriction does not affect the phonetic quality of either the independent fricative or the fricative element: they are certainly accepted as normal. We will return to this point later.
Figure 5-16: Linguopalatal Configurations of [ts] at the point of MAX, Closure Offset, and Fricative Onset (Symmetrical Contexts): (a) speaker MN; (b) speaker TM
Figure 5-17: Linguopalatal Configurations of [te] at the Point of MAX, Closure Offset, and Fricative Onset (Symmetrical Contexts): (a) speaker MN; (b) speaker TM
5.5.2 Coarticulatory Characteristics of [ts] and [tc]

The coarticulatory effects of the preceding (V₁) and the following (V₂) vowels were analysed for the three regions, front, central, and back. The effects of the preceding vowels upon the stop element of the affricates were examined at the MAX point of complete occlusion. The effects of the following vowels upon the fricative element were analysed at the frame corresponding to the onset of friction. In the following discussion we use the symbol [t'] for the stop element of the affricates [ts, tc], and [s'] and [ç'] for the fricative elements of the affricates [ts, tc] respectively.

Figure 5-18 presents the mean percentage values of the contextual effects upon the [ts] articulation. The EPG palatograms are derived from the three temporal points, MAX, closure offset, and fricative onset. They are given in Figures 5-19 and 5-20. It can be seen from the summary measures below that coarticulatory trends are similar between the two speakers. For [t'] the amounts of contact in the front region are fairly stable, while those in the posterior two regions appear to vary with the preceding vowel type. For [s'] the contact in the front region tends to vary together with that in the central and back regions.

![Figure 5-18: Effects of the preceding and following vowels on [ts]: Speaker MN (top panels) and Speaker TM (bottom panels); X-axis=the V/s/V sequences; Y-axis=measures of the contact amount (%)](image)
Figure 5-19: Linguopalatal Configurations of [ts] at the point of MAX, Closure Offset, and Fricative Onset (Asymmetrical Contexts): speaker MN
Figure 5.20: Lingual-Palatal configurations of [ts] at the point of MAX, Closure Offset, and Fricative Onset (asymmetrical contexts): speaker TM.
The results of a one-way ANOVA are summarised in Table 5-5 above. For the [t'] of [ts] significant differences of the preceding vowels were found for all the regions. One exception is the front region (MN) but the different amount of contact derived from the flanking V₁ turns out to be statistically not significant. The pattern of significant differences is very similar to that in an independent [t] (see Table 4-7 in chapter 4). In contrast, the data for speaker TM shows that the contact amount is greater when the V₁ is /a/, as compared with that when the V₁ is either /i/ or /u/. This is due to the anticipation of the fricative element, rather than the contextual effects. Because the effects of V₁ were not found for TM’s data of an independent [t] (carryover effects, Table 4-7), the variation could be viewed as a difference in the advance (motor) planning between [t'] and [t].

The posterior two regions of [t'] evidently reflect the articulatory nature of the preceding vowels. The relation between the contextual vowels in the central region is different between the two speakers: /i/>=u=/a/ (MN) and /i/>=u>/a/ (TM). This admits of two interpretations. First, as we discussed in chapter 4, there are idiosyncratic differences in the production of [ui]: the involvement of the medio-dorsum is larger for TM and the vowel [ui] has a more centralised quality. Because the positioning of the tongue dorsum during the narrow constriction depends primarily on the height of the preceding vowel, the raising of the medio-dorsum for the V₁ [ui] is carried over to the stop element [t']. Second, if we assume that the fricative element [s'] is already planned and its motor execution begins at the MAX point, then the higher amount of contact in the [ui] context (TM) can be considered as a preparatory activity for apico-laminal articulation of [s'].

On the other hand, the fricative element [s'] of [ts] at the friction onset shows significant differences in all the regions. The front region shows opposite results in the relation between the vowel pairs: the three flanking vowels significantly affect the contact degrees for MN but not for TM. This disparity can be accounted for by the realisational differences of an independent [s] between the two speakers, namely apical and apico-laminal: the former is more susceptible to the coarticulatory effects than the latter. Given this, we can
expect that the posterior two regions also exhibit the pattern related to the two types of [s] realisation. However, as Table 5-5 indicates, such an expectation turns out to be wrong. The pattern of significant differences is similar to that of [t'], rather than an independent [s] (see Table 5-2 above for anticipatory effects). These results suggest that the (individual) characteristics, or realisational differences, of an independent [s] are maintained only in the pattern of contact in the front region; and that the dorsum activities are interwoven in the production of the affricates [ts]. This accurately represents articulatory aspects of the ‘close-knit’ (Gimson, 1980) nature of the consonantal gesture.

Turning now to the affricate [tç], Figure 5-21 below shows the changes of mean percentage values of [t'] and [ç'] as a function of vowel contexts. The EPG palatograms are given in Figures 5-22 and 5-23. Similarly to the affricate [ts], the general trend of coarticulatory effects is similar between the two speakers. For the stop element [t'] the front contact is greater when the V1 is either /a/ or /u/: the smaller contact in the /i/ context is due to the fact that the contact in row 1 tends to be withdrawn. The contact amount in the posterior two regions, in contrast, decreases in the order /i/>/u/>/a/. However, these coarticulatory trends are not obtained in the fricative element [ç'].
Figure 5-22: Lingual Palatal Configurations of [tc] at the point of MAX, Closure Offset, and Fricative Onset (Asymmetrical Contexts): Speaker MN
Figure 5-23: Linguopalatal configurations of [tc] at the point of MAX, Closure Offset, and Fricative Onset (asymmetrical contexts): speaker TM
Table 5-6: One-way ANOVA Results for the Effects of V1 on [tʰ] and Those of V2 on [ç']

<table>
<thead>
<tr>
<th>Effects Of V1 On [tʰ]</th>
<th>Regions And F Values†‡</th>
<th>Effects Of V2 On [ç']</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FRONT</strong></td>
<td><strong>CENTRAL</strong></td>
<td><strong>BACK</strong></td>
</tr>
<tr>
<td><strong>F(2,42)</strong></td>
<td><strong>F(2,42)</strong></td>
<td><strong>F(2,42)</strong></td>
</tr>
<tr>
<td><strong>SPEAKER MN</strong></td>
<td><strong>u=a=i</strong></td>
<td><strong>i=u&gt;a</strong></td>
</tr>
<tr>
<td></td>
<td><strong>27.90</strong></td>
<td><strong>35.40</strong></td>
</tr>
<tr>
<td></td>
<td><strong>25.43</strong></td>
<td><strong>118.85</strong></td>
</tr>
<tr>
<td><strong>SPEAKER TM</strong></td>
<td><strong>a =u=i</strong></td>
<td><strong>i=u&gt;a</strong></td>
</tr>
<tr>
<td></td>
<td><strong>28.84</strong></td>
<td><strong>22.65</strong></td>
</tr>
</tbody>
</table>

**FRONT**             | **CENTRAL**             | **BACK**              |
| **F(2,42)**           | **F(2,42)**             | **F(2,42)**           |
| **SPEAKER MN**        | **u=a=i**               | **u=i=a**             |
|                      | **1.82**                | **2.02**              |
| **SPEAKER TM**        | **i=a=u**               | **i=ü=a**             |
|                      | **6.00**                | **7.00**              |

†=p<0.05, ‡=p<0.01, unmarked = non-significant

The relation between the vowel pairs was assessed by unpaired t-test; the criterion for the significance was 0.01. The symbol > indicates a significant difference but the symbol = indicates no significant difference.

Table 5-6 above presents the results of a one-way ANOVA to test for the effects of the preceding and the following vowels for the two elements of the affricate [tç]. The stop element [tʰ] indicates significant effects in all the regions. Different relations between the vowel pairs in the front region are due to the involvement of the tip activity: the tongue tip tends to be directed downwards across the vowel contexts in MN’s utterances, but tends to participate in the complete closure in TM’s production. In contrast, the posterior two regions reveal the general tendency that the amount of contact is larger in the high vowel contexts than in the low vowel context. Since the height of the preceding vowel determines the positioning of the tongue dorsum during the closure, the results demonstrate that the stop element [tʰ] is sensitive to the height of the V₁, rather than the backness. This characteristic, height-dependent coarticulation, was also observed for the alveolopalatal nasal [n] (see section 4.3.3.3 in chapter 4). It appears that the significant pattern is, to some extent, similar to that found for [n] (see Table 4-8 for carryover effects).

The fricative element [ç'] of [tç] exhibits far fewer significant effects in the three regions and there is a strong tendency that the effects of the flanking V2 are held off. There is one exception in the back region (TM), showing the significant relation /i//=u/>/a/. This can be explained by the height-dependent coarticulation mentioned above: the anticipation of the following low vowel causes the earlier lowering of the backmost contact. As observed in sections 5.4.1.2 and 5.4.2.2 above, the stronger degree of coarticulatory resistance is the characteristic of the alveolopalatal fricative [ç]. This articulatory nature is retained in the fricative element of the affricate [tç].

The two characteristics, height-dependent coarticulation and stronger coarticulatory resistance, are common properties that we found for the alveolopalatal nasal [n] and fricative [ç]. It is reasonable that the two elements of the alveolopalatal affricate [tç] have those two properties. It follows from this that the presence of the shared properties characterises the articulatory nature of ‘close-knit’ in the alveolopalatal affricate.
5.6 Intergestural Coordination

In the previous sections we have examined the articulatory characteristics of the constriction shapes, side-to-side widths, anticipatory and carryover coarticulation, specific to the fricatives [s, ç, ç] and affricates [ts, te]. In this section, we investigate intergestural coordination between the two components of the tongue during the production of the fricatives and the fricative elements of the affricates. The data of contact in the front region and the back region was studied for dependency between the tongue tip/blade and the dorsum. A regression analysis was first carried out for the group of the fricatives and the fricative elements [s, ç, s', ç']. Then the five consonants including [ç] were analysed separately. Two points need to be made. First, the measurement point was different between [s, ç] and [s', ç']. For the former, the percentage values in the two regions were measured at the MAX point, but for the latter, the values were calculated at the frame corresponding to the friction onset of the affricates [ts, te]. However, this difference, as the examination of the contact data suggested, might not substantially affect our analysis. Second, the regression of the two variables differed when each of the five consonants was analysed separately. For the coronal consonants the front region contact was regressed against the central region contact. For the dorsal consonant, on the other hand, the inverted regression was made. This is because the segments in question involve diverse regions on the tongue for making the given major constriction.

Figure 5-24 presents the regression plots for the group of fricatives [s, ç] and the fricative elements [s, ç] across the speakers and the vowel contexts. The plots exhibit a loose clustering and a very weak $r^2$ value was derived: approximately 19% of the variation of the front contact is accounted for by the variation of the central region contact.

![Figure 5-24: Regression Analysis of The Fricatives [s, ç] and The Fricative Elements [s', ç'] of The Affricates [ts, te]: Plots For Front Contact Degree Regressed Against Central Contact Degree (Pooled Across Speakers And Vowel Contexts): n=360](image_url)
Figure 5-25 below demonstrates the results of the regression analyses performed separately for each consonant. It is evident that the five consonants fall into two groups in terms of the $r^2$ values and the types of clustering: [s, ç, s', ç'] on the one hand; and [ç] on the other.

**Figure 5-25: Regression Analyses Of The Fricatives [s, ç, ç] And The Fricative Elements [s', ç'] Of The Affricates [ts, te]:** for [s, ç, s', ç'] the degree of front contact is regressed against that of central contact; for [ç] the central contact degree is regressed against that of front contact: n=90 for each of the consonants.

![Regression plots for consonants](image)
In the fricative [s] and the fricative element [s'], although the front region contact varies with the phonetic contexts, the variation of the central region contact is mostly restricted to the range 20-50%. This suggests that not only the activities of the tongue tip/blade but also those of the dorsum are constrained. This is compatible with our observations of the tongue shape and those of previous studies (e.g. Stone et al. 1992): a length-wise hollow is typically formed during the narrow constriction of [s] and [s']. Furthermore, coarticulatory effects are generally suppressed: the posterior region does not coarticulate freely with the surrounding vowels. Although the apicality, or laminality, differs idiosyncratically, only the lateral margins are accommodated to the vowel gestures.

The fricative [ç] and the fricative element [ç'] exhibit characteristics similar to those described in a different way. The two consonants are similar in the sense that they both reveal a tight clustering and their $r^2$ values are weaker than the other three consonants. The tight clustering explicitly suggests that the alveolopalatals actively involve both the tip/blade and dorsum components of the tongue. The consequence of this constraint is, as discussed before, a stronger degree of coarticulatory resistance to the adjacent vowels.

The palatal fricative [ç], on the other hand, indicates an acceptable $r^2$ value: approximately 65% of the variation in the front region is explained by the variation in the central region. The feature specification also predicts the variation of the involvement of the tip/blade component: the articulatory feature of the coronal node is unspecified. As seen in the analysis of coarticulatory effects, the amount of contact in the front region and in the central region tends to vary with the vowel contexts. Both tongue tip/blade and pre-(/medio-) dorsum, to some extent, are retracted and are supposed to facilitate the raising gesture of the tongue body. This active movement, however, does not function to increase the degree of coarticulatory resistance.

5.7 Summary and Conclusion
This chapter investigated the articulatory and coarticulatory characteristics of the voiceless fricatives [s, ç, ç] and affricates [ts, tç]. In the first set of experiments, the data for the lingual gestures of the fricatives were presented and that for the affricates were described in the second experiment.

The fricative [s] involves the tip (or rim) in the constriction formation. It was found that the realisation was considerably different for both speakers: apical and apico-laminal. These two patterns reflected not only the contact shape of the tongue tip/blade but also the overall shape of the tongue. The apical realisation is more susceptible to the coarticulatory effects of the surrounding vowels: the narrow constriction was fronted to touch the upper central incisors; consequently, the broader groove width was generated. This agrees with the data reported by Bladon & Nolan (1977) and Recasens et al. (1993). The consonant [ç] is an
alveolopalatal: the tongue blade and medio-dorsum are actively involved in the narrowing constriction. The consonant [ç] is a palatal articulation, in which the tip/blade component of the tongue is retracted and the dorsum is raised towards the posterior portion of the medio-palatal and post palatal regions on the palate.

Different trends of coarticulatory effects were observed for the three regions of the fricatives in question. For [s] it was found that there was a clear predominance of the carryover effects. Few coarticulatory effects were found for [ç]. For [ç] the preceding vowel influenced the amount of contact in the posterior two regions and the extension of anterior contact. This consonant revealed the height-dependent coarticulatory effects: the amount of contact generally increased in the /i/ context but decreased in the /a/ context. Nevertheless, the basic contact pattern was maintained across the vowel context.

The assumption that the constriction locations and the groove widths correlate linearly was tested for [s, ç, ç] in the vowel symmetrical contexts. The result of regression analysis reveals a tolerable $r^2$ value (0.669): the linear correlation is generally supported. However, although the three consonants have distinctive places of articulation, the side-to-side widths of the two coronal fricatives tend not to be statistically significant: a similar size of the groove width can be formed at different places of articulation. Therefore, it could be assumed that the groove width is not an important articulatory feature in distinguishing the contrasting fricatives. This agrees with a number of studies (e.g. Ladefoged & Maddieson, 1996; Ladefoged & Wu, 1984). In addition, the present results support the idea that a small number of spatial parameters may be enough for generating the acoustic effects of the sibilants ([s, š] in English) (Nguyen et al. 1994).

The overall tongue shape, rather than the width, is a more consistent and reliable parameter for identifying the fricatives. For [s] the middle of the tongue is not contacted. This is related to the self-maintenance of a typical hollowing formed midsagittally. For [ç] the whole body of the tongue is effectively fronted and raised upwards. Consequently the tongue becomes convex to the palate. Different constriction locations for [s, ç, ç] correspond to different sizes of the front cavity: the front contact progressively decreases as the constriction location shifts backwards.

Ladefoged & Maddieson (1996) proposed the idea that the alveolopalatal[s, z] can be characterised as palatalised post-alveolars. This characterisation is an attempt to incorporate the tongue shape features [grooved], [flat], [domed] and [palatalised] for the description of the fricatives: the retroflexes [s, z] are then identified as laminal flat post-alveolar; the palato-alveolar [ʃ, ʒ] as laminal domed post-alveolar; and the alveolopalatal[s, z] as laminal palatalised post-alveolar. Ladefoged & Maddieson (1997; 148, 153) mention that ‘the part of the tongue immediately behind the constriction for [ʃ] is ‘raised (or domed)’, as opposed to being hollowed for [s]’. They also mention that the
major difference between [J] and [ç] is 'in the degree of raising of the front of the tongue'. These specifications of the tongue shape correspond to the phonological representations described in (1). Halle & Stevens (1997) argue for the above proposal by comparing the Polish [s, z] with [ç, ï]. The reason for this is that the crucial difference between the two classes lies in the positioning of the tongue dorsum: there is an [i]-like shape in [ç] but not in [s]. The results of our EPG experiments for the Japanese [ç] raise a question concerning the opposition between the feature [domed] and [palatalised]. Also, the term ‘palatalised’, as we will discuss in the next section, can imply dynamic movement, as well as static posture. Since our EPG data also suggests that the front region (the blade posterior and pre-/medio-dorsum) is ‘raised’ for [ç], the two shape features in question are confusing. Stone et al. (1992), using Ultrasound imaging and the EPG technique, showed that during the production of the English [J], the anterior and dorsal regions of the tongue are raised, while the groove in the posterior region is deeper. Therefore, it is supposed that the difference in the ‘overall’ tongue shape between [ç] and [J] is better characterised by calling them convex and concave respectively.

The second set of experiments demonstrated that the ‘close-knit’ characteristics of the affricates manifested three properties: (i) the stop and fricative elements had to be homorganic, this is reflected in the approach and the release phases; (ii) the preparatory activity for the following fricative element started approximately two-thirds of the way through the stop closure; and (iii) the vowel-dependent coarticulatory effects were found to be similar between the two elements, rather than between the corresponding two elements, the independent stop and the fricative(s).

Overall, the degrees of dependency between the two components of the tongue are specific to each of the five fricatives [s, ç, s', c', ç']. The weaker r^2 values of [s, ç, s', c'] were related to the fact that the degree of contact in the two regions exhibited little variance: a relatively tight clustering of the plots was observed. It is supposed that the looser clustering of the fricative elements of the affricates was due to the carryover effects of the stop element; or possibly to the measurement point, the friction onset. The tight clustering of the plots implies that the tip/blade and the dorsum were tightly controlled by the production requirements for forming the relevant shape. In contrast, the r^2 value of [ç] suggests that there is a stronger correlation between the two components of the tongue. As observed in the coarticulatory characteristics, the extension of the front contact varied with the vowel context. Also, the varying degrees of the posterior contact were derived from the height-dependent coarticulatory effects. These two factors are clearly represented in the values of linear correlation.
Chapter 6

Gestural Coordination and Timing in Japanese Palatalisation

6.1 Introduction

Research on coarticulation, the phenomenon whereby 'the movements of articulatory organs overlap in time with those of different articulatory organs or different parts of the same organ' (Farnetani et al., 1989), explicitly demonstrates that, while the vocal tract in speech production is constrained bio-mechanically, it is organised phonologically. This leads us to another way of understanding the relation between language-specific allophonic rules in the phonological component and seemingly automatic coarticulation in the traditional speech production model. It is possible to conceptually incorporate the notion of coarticulation with linguistically specified assimilation, characterising it as phonologised coarticulation, or 'soft coarticulation' (Fujimura, 1990). This view assumes that the articulatory mechanism underlying a given allophonic variation is language-dependent, and avoids the arbitrary decision that the variation is not a learned behaviour. The basic issue then, is to examine and to parameterise the use of articulators specialised for a variation in question.

This chapter approaches the issue through a detailed analysis of lingual articulation of /n/, /s/, /r/, and /k/ in Japanese. Particular focus is placed on one common secondary articulation of consonantal segments, namely palatalisation. Palatalisation of lingual consonants is a typical case of articulatory antagonism, where competing and conflicting demands are assigned to the same articulator, the tongue. In the experimental phonetic field, surprisingly few studies have so far been made on the issue of tongue control during the articulation of palatal(ised) consonants in Japanese. Previous studies in other languages more often restrict their analyses to a palatal nasal [n] than the other palatal(ised) consonants. In the phonological field, although various proposals have been put forward to characterise the phenomenon in a unified way (e.g. Hume, 1994; Mester & Itô, 1989; Lahiri & Evers, 1991), there is considerable uncertainty as to the proper characterisation. The controversies and diverse proposals exist because of a basic lack of empirical support.

We address the question of intergestural coordination and timing between the primary articulatory gesture and the (secondary) dorsal gesture during the production of Japanese [n], [c], [r] and [k]. Their spatio-temporal characteristics are assessed qualitatively and quantitatively by the use of EPG. This will lead us further into a consideration of palatalisation as a language-specific coarticulatory strategy in Japanese. Furthermore, the

1) Some preliminary results were presented in Nakamura (1999, 2000).
exploration of interarticulatory coordination involves the problem of the articulatory complexity, or more commonly segment complexity, of palatal(ised) consonants. These two different, but interrelated, problems are discussed in the next section. We will attempt to develop the concept of the ‘correlation of palatalisation’ (Trubetzkoy, 1958), relating it to the idea of ‘articulatory settings’ (Honikman, 1964), namely that suggests that systematic articulatory manoeuvres are language-specifically decided. Reviewing previous works on Japanese and other languages, the experimental hypotheses that will be examined in section 6.3 are formulated concerning intergestural coordination and timing between the primary articulatory gesture and the (secondary) dorsal gesture. Combining the EPG evidence with that gained from the study of phonological structure of Japanese, we will propose that Japanese palatalisation is a language-specific coarticulatory strategy for the use of the tongue dorsum.

6.2 The Phonology and Tongue Dynamics of Japanese Palatalisation

6.2.1 Palatalisation as a Language-specific Coarticulatory Strategy

In Japanese, as we saw in Chapter 2 (section 2.3.1), palatal(ised) consonants appear before the high front vowel /i/ (Ci syllables) and before the nonfront vowels /a, o, u/. The latter cases are phonologically contrastive, being analysed as /Cj/ (CyV syllables). The general forms of the characteristic patterning are given below.

\[(1a) \ C \rightarrow \text{palatalised C} \quad /\text{_____/i/}\]
\[\text{e.g.} /n/ \rightarrow [n], /s/ \rightarrow [ç], /r/ \rightarrow [r¹], /k/ \rightarrow [k²]\]

\[(1b) \ C+j/ \rightarrow \text{palatalised C} \quad /\text{_____/a, o, u/}\]
\[\text{e.g.} /nj/ \rightarrow [n], /sj/ \rightarrow [ç], /rj/ \rightarrow [r¹], /kj/ \rightarrow [k²]\]

The rules say that, while the consonant acquires the phonetic feature of /i/ in (1a), the preceding consonant and /j/ coalesce into one in (1b). These processes at the higher pre-motor level are often referred to as ‘palatalisation’. This generates an identical palatal(ised) consonant, the feature specification of which serves as a surface phonetic representation that is converted to a plan for vocal tract activities. The description in (1) raises the question of whether there is any difference in what is characterised as the same palatal(ised) segment, and what are characterised as different palatalising environments. To analyse these questions it is crucial to study the phenomenon as the articulatory events in the vocal tract, rather than as the abstract allophonic processes in the phonological component.

The palatalisation phenomenon is characterised either statically or dynamically: ‘the addition of a high front tongue position, like that in [i], to another articulation’ and ‘a process in which the primary articulation is changed so that it becomes more palatal’ (Ladefoged, 1993; 230). A palatalised bilabial nasal [m³] implies that the [i]-like posture of the tongue...
body overlies the production of [m]. In a phrase such as I miss you the voiceless alveolar fricative [s] assimilates or may assimilate the feature of the following palatal glide [j] and gradually becomes more palatal. The final example is the shift of the constriction place for a velar stop as a function of the contextual vowels. Before /i/ the closure is made at a more central area, which is described as velar fronting. These are just three examples to show that the two conventional interpretations, static and dynamic, seemingly go well both with the description of the allophonic rules in the phonological component such as in (1) and with that of the movements in the oral cavity.

However, there are two difficulties in maintaining the term palatalisation with the static and dynamic sense mentioned above. First, the term is used to draw our attention to a ‘secondary’ posture, or gesture, of the tongue in addition to that of the primary posture or gesture. Accordingly, alveolopalatals and palatals are technically excluded, even though they involve an [i]-like shape of the tongue body in their articulation, as there are no palatalised alveolopalatals or palatalised palatals. Second, the two interpretations appear to draw a parallel between the articulatory activities in the vocal tract and the significance in the phonology. Regressive palatalisation in Ci syllables (1a), the phonologically non-contrastive case, is often considered as spontaneous assimilation, or the chaining effect of successive phonetic segments. This unreasonably involves a phoneme-allophone relationship in describing the articulatory events in the vocal tract. Hattori (1967; 542) explicitly suggests:

...the [k] in the Japanese [ki] is palatalised and the point of articulation is in the extreme front when compared with the [k] in [ka]. However, this does not mean that the former is palatalised by casual assimilation due to the influence of the succeeding [i] when aiming at the [k] of [ka]. Even in the most careful pronunciation of [ki], articulation is aimed primarily at the palatalised [k].

It has been disputed whether regressive palatalisation is phonologically motivated or not (e.g. Coleman, 1998). The distinction between pre-programmed and chaining is rather arbitrary. A good example is allophonic variation in the English /l/. Two distinct allophones, clear and dark /l/, are ‘conditional’ allophones: clear /l/ occurs before vowels and /j/ as in leap, block, silly, will you, while dark /l/ occurs elsewhere. There is a fairly general agreement that dark /l/ is specified by a phonological rule, velarisation, which assigns the feature [+back] to /l/ in syllable rhyme position (e.g. Halle & Mohanan, 1985). This implicitly proposes that clear /l/ is ‘predictable’ from universal phonetics. There is no reason to assume that clear /l/ is automatically implemented without any specification for such a resonance. Similarly Japanese palatalisation in the Ci syllables (1a) is ‘phonologically conditioned’ but is often said to be ‘predictable’. This essentially implies that the phenomenon requires no special phonemic representations, not that it is an involuntary action in the vocal tract.

In order to avoid these technical and practical dilemmas, we define palatalisation as
'a specialised use of a raising gesture of the tongue dorsum'. This working definition gives particular attention to the tongue dorsum as an essential articulator and effectively covers the articulation of alveolopalatals and palatals. By 'a specialised use' two characteristics are meant: (i) the function of the dorsum raising gesture is specific to the given consonantal segment; it depends on the primary articulator whether the dorsum raising gesture is regarded as 'secondary' or not; and (ii) the pattern of the spatio-temporal control may differ from language to language. This presupposes that the concept of 'articulatory settings', particularly 'tongue settings' (Honikman, 1964) is important in the analysis of the lingual gesture. We must now return to the point of the relationship between the feature specification and its consequence in a generative model of speech production.

The surface representations of the palatal(ised) consonants in $C_i$ and $CyV$ syllables in Japanese are indistinguishable. The posited representations converted to a plan for vocal tract activities are illustrated below. These are orthodox representations and irrelevant details are omitted (Keating, 1991, 1993).

(2a) **PALATALISED LABIALS**

<table>
<thead>
<tr>
<th>Place</th>
<th>Labial</th>
<th>Dorsal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[+high]</td>
<td>[-back]</td>
</tr>
</tbody>
</table>

(2b) **PALATALISED ANTERIOR CORONALS**

<table>
<thead>
<tr>
<th>Place</th>
<th>Coronal</th>
<th>Dorsal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[+anterior]</td>
<td>[+high]</td>
</tr>
</tbody>
</table>

(2c) **PALATALISED NON-ANTERIOR CORONALS**

<table>
<thead>
<tr>
<th>Place</th>
<th>Coronal</th>
<th>Dorsal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[-anterior]</td>
<td>[+distributed]</td>
</tr>
</tbody>
</table>

(2d) **PALATALISED VELARS**

<table>
<thead>
<tr>
<th>Place</th>
<th>Dorsal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[+high]</td>
</tr>
</tbody>
</table>

The executing articulator of palatalisation is the tongue dorsum that consists of the two features [+high] and [-back]. By adding the dorsal node labials and anterior coronals turn into complex segments as in (2a-c), whereas velars do not as in (2d). Since velars are already

---

2) Mester & Itô (1989) claim that palatalisation is realised by raising of the tongue blade, rather than the dorsum. Thus, palatalisation is coronalisation, by which a coronal component is added to the primary articulation: if the component is already present, it is further enhanced by [-anterior]. There is no clear indication, or empirical support, for such a proposal. Apart from the details of the specification theory that the authors assume, the proposal may come from the difference in what is considered as the blade. Catford (1977) and Keating (1991) point out that, while the blade is 'the part that lies opposite the teeth and alveolar ridge when the tongue is at rest (Catford, 1977; 143)' in British phonetics, that part is called the tip in American phonetics. Thus, what Mester & Itô count as the blade appears to include both the blade and front of the tongue in British tradition. Even if we admit such a terminological gap, it is wrong to suppose that the tongue dorsum is irrelevant for palatalisation.
specified as dorsal [+high], it incorporates a paired task, namely the feature [-back], to effect palatalisation. If we use these diagrams to describe the target palatal(ised) consonants in Japanese, we apply the representation in (2b) for [H], that in (2c) for [n] and [ç], and that in (2d) for [k^]3). It is tacitly believed that the consonants are implemented, or timed, in the same way regardless of their distinct origins. However, there is no conclusive proof of how the (secondary) dorsal gesture interacts with the primary articulatory gesture in the palatal(ised) consonants in question.

A previous EPG study by Miyawaki et al. (1974a) examined /n/ in the VCV and VC/j/V sequences, where the vowels were possible combinations of /i/ and /u/. It was found that there are two spatial effects: an overall increase in contact and a backward shift of the constriction, in both /Vni/ syllables and /Vnju/ syllables at the closure mid-point. Their data indicates that there is no substantial divergence in the contact amount between both /n/s. In contrast, the articulatory gestures for /i/ and /j/ in the vowel /u/ symmetrical context were characteristically different from each other: there was a typical ‘bulging of the tongue towards the mid-anterior portion of the hard palate for /j/’, while this was not found for /i/ (Miyawaki et al., 1974a, 52f). This observation of the two vocoids was partially verified using electromyographic techniques (EMG). Kakita et al. (1976) obtained EMG recordings from the genioglossus muscle (GG) and found that the activities corresponding to /j/ at maximum value were reduced especially in the posterior fibres of the GG, compared to those to /i/. After comparing the static display of X-ray data, Kakita et al. suggest that such different GG activities were reflected in differences in the surface shape of the tongue, the root of which is not pulled forward during the production of /j/. Yet, as the authors acknowledge, the temporal, as well as spatial, aspects of tongue activity are not negligible in the analysis of /j/.

In their subsequent study Miyawaki et al. (1974b) asserted that the articulatory nature of Japanese palatalisation lies in the [j]-element rather than the [i]-element. This relies heavily on the ‘characteristic bulging of the tongue’ for the palatal glide /j/ and it is taken as evidence to draw a distinction between the two palatalising environments, namely the articulatory gesture for /i/ and that for /j/. However, the claim apparently contradicts the fact that a palatal(ised) consonant appears before a high front vowel. Since both the high front vowel /i/ and the palatal glide /j/ are typified as approximant articulation (Catford, 1977), the presence and absence of the tongue ‘bulging’ can be considered as a (possibly idiosyncratic) difference in degree, rather than in kind, of the tongue dorsum raising. Furthermore, it is not at all clear how the tongue dorsum is spatio-temporally coordinated with the tongue tip/blade during the production of the given palatal(ised) consonant. It remains an unsettled question

3) Although languages may differ in the coronal features for non-anterior coronals (2c), this is not crucial here. The point is that the executing articulator is the tongue dorsum for all the palatal(ised) consonant5.
whether the palatal(ised) consonants in \( Ci \) and \( CyV \) syllables are articulated in the same way.

The findings of previous EPG and EMG experiments, though they are very limited, suggest that the constriction shape alone may not help to differentiate two versions of the palatal(ised) consonants. When we consider phonological patterns in (1), it is reasonable to hypothesise that the palatal(ised) consonants involve the two types of the (secondary) dorsal raising gesture, namely \([i]\)-like and \([j]\)-like gestures. Support for this hypothesis comes from Recasens (1984a,b,c). In Catalan the dorsum raising gesture for the sequence \([nj]\) is, surprisingly, smaller in degree and shorter in duration than that for \([n]\) or the sequence \([ni]\). This can be explained by the fact that the consonant \([n]\) has a phonemic status in the language. Recasens' EPG data shows that, while the linguopalatal patterns at maximum constriction are almost identical, a three-way distinction is made in terms of the magnitude and timing of the dorsum raising gesture. Further support is found in Maddieson & Emmorey (1985). In the cross-language acoustic study of Amharic, Yoruba and Zuni, the authors demonstrate that, although the same phonetic symbol is used to represent the semivowels, each language has distinctive acoustic targets for \(/j/\) and \(/w/\). This suggests, from the articulatory point of view, that palatal(ised) segments differ from language to language in the nature and extent of the involvement of the tongue dorsum raising gesture.

The hypothesis that the palatal(ised) consonants in Japanese involve two types of temporal coproduction with the dorsum raising gesture, is of direct relevance in phonological representation. As exemplified in (2), all the three classes of the palatal(ised) consonants, namely palatalised labials, (non-)anterior coronals and velars, contain the dorsal node of \([+\text{high}, -\text{back}]\) without any 'temporal' specification. The above hypothesis, if supported, makes it more obvious that the time element must be treated explicitly at the abstract phonological level. In addition, such intergestural timing would be a language-specific part of phonetic knowledge. In traditional phonetics there is an idea of capturing the phonetic individuality of a language. This is called 'articulatory settings' (Honikman, 1964), the essential idea of which is that systematic articulatory manoeuvres are language-specifically decided. This point has a connection with the general observation that the vocal tract in speech production is organised phonologically within the biomechanical limitations. If we assume that interarticular coordination is the minimum domain of 'articulatory settings' (Laver, 1994; 396), then it seems reasonable to suppose that the raising gesture of the tongue dorsum could be generalised as one language-specific coarticulatory strategy to effect palatalisation.

6.2.2 The 'Complexity' Debate
The extent to which the larger amount of tongue-palate contact, observed in palatal consonants, is 'actively' controlled is controversial. Although in the previous discussion we
have assumed the view proposed by Keating (1988, 1993), that alveolopalatal and palatal consonants are complex segments, as in the representations in (2b,c), we have to inquire, to some extent, into the articulatory complexity of those consonants.

Keating (1988) describes the palatals as involving more than one class of articulators on the grounds that these consonants demonstrate a longer constriction along the palate than the others. The articulators in question are 'the very back of the blade and the large front part of the dorsum' that correspond to the 'complex' specification of the coronal and dorsal components. There is no distinction in the tongue body features between alveolopalata and palatal consonants: they are defined by corona [-anterior, +distributed] and dorsal [+high, -back] as illustrated in (2c)⁴. Thus, it is assumed that two distinctive articulators are simultaneously executed in those consonantal articulations. This view concurs with what is called 'contiguous articulations' by Catford (1977* 194): 'they involve adjacent articulatory zones or sub-zones' and the articulators are 'so close to each other that they cannot be moved completely independently of each other'. It must be noted, as Catford states and Keating implicitly acknowledges, that the natural divisions both on the palate and on the tongue are presupposed for their descriptions.

The argument against such a complex-segment analysis has come from Recasens and his colleagues (1990, 1993, 1994a,b, 1995, 1997). Based on the visual inspection of lateral x-ray tracings and the results of the EPG experiments, Recasens (1990) rejects the two traditional articulatory zones on the palate, namely alveolar and palatal, and gives evidence to show that the two classes of the palatal consonants involve independent control parameters: alveolopalata ([p], [c], [k], [t], and [tç]) are primarily predorso-postalveolo-prepalatal (the involvement of the lamina and medio/postdorsum is barely acknowledged); and true palatals, [j] and [ç] are mostly predorso-prepalatal and mediodorso-mediopalatal respectively. Whereas the active articulator is primarily the tongue pre- and medio-dorsum, the constriction place ranges from the postalveolar zone to the mediopalatal zone. It is implicitly suggested that the correlation between the active and passive articulators is oversimplified in the conventional assumption: the part of the tongue making contact against the palate is that which lies directly under that location, except for retroflex articulation. Hence, the complex-segment view is cancelled. The point of a simplex-segment proposal is that alveolopalatal consonants are articulated with a much more specific tongue region, that of pushing the tongue predorsum against the intersection between the alveolar ridge and the front arch of the hard palate.

The robustness in the complex-segment concept is that the tongue blade and dorsum are controlled independently. Given this, it is possible to hypothesise that the amount of dorsal contact for an alveolopalatal is the same as, or larger than, that for a palatal glide.

⁴ A laminal palato-alveolar [j] and an apical retroflex [s] are simple [-anterior] coronals without the dorsal component of the palatal specification (2c) (Keating, 1993).
Alternatively, if an alveolopalatal is a simplex segment, then there are few time lags between
the activity of the tongue front and that of the tongue dorsum for making a constriction on
the hard palate. Recasens tested these hypotheses as to the spatial and temporal aspects of the
palatal articulation. Using the EPG technique (the Reading system), the comparison was made
for [n] and [j] produced by five Catalan speakers. Recasens et al. (1994) found that the
contact amount in medio- and post-palatal zones was less extensive in [n] than in [j]. The data
indicates that the contact for [n] is first made at the row 4 or 5 (the back of the alveolar zone
and the front of the prepalatal zone); and that the contacts anterior and posterior to those rows
gradually increase during the complete closure. Recasens and Romero (1997), using an
electromagnetic midsagittal articulometer (or EMMA), measured time intervals between the
tongue front and the tongue dorsum at maximum constriction for Catalan /ɲ/ and /nj/ and for
Russian /ɲ/ and /nj/ ([n̪j]). It was found that time lags were shorter in Catalan /ɲ/ than in
Russian /ɲ/; and that the sequence of /nj/ consistently showed a longer time lag than the other
consonants in both languages. From these results it is asserted that an alveolopalatal [n] is not
a complex segment, in which the two components of the tongue are activated independently.
In contrast, a palatalised alveolar [n̪] and a sequence of [nj] are complex segments.

There are reasons to doubt the above account of a simplex segment. First, the
'accompanying' laminal and dorsal activities for [n], which effectively produce the extensive
contact along the palate, are considered as mechanical coupling effects: 'an [probable]
increase in the force with which the primary articulator [the tongue predorsum] presses the
palatal surface' (Recasens et al. 1994; 88; Recasens & Romero, 1997; 58). This claim,
however, is unfounded. Second, the shortening of intergestural overlap does not properly
support the proposition that the consonant is produced by a single gesture, but simply
demonstrates the fact that the two independent components of the tongue are activated
without the temporal loss. The cluster of the findings in the spatio-temporal domains makes it
obvious that the timing of intergestural overlap, or the degree of articulatory blending, is
conditioned phonologically and is controlled consciously by a speaker of a language. Although Recasens conflates the issue of articulatory complexity with that of various degrees
d of gestural blending, this is undesirable.

Furthermore, such a conflation of the issues leads to a very distorted view of
palatalisation: 'palatalisation simply requires the loss of temporal lag between the two
gestures of the cluster [nj]; the outcome of this process is a new consonant produced with one
articulator at a single place of articulation' (Recasens et al. 1993; 230). Recasens et al. (1995)
propose the process of 'gestural blending' (Browman & Goldstein, 1989), by which the
output becomes an intermediate articulation between the two original gestures, as opposed to
'segmental merging', in which the two lingual gestures are blended completely as in Catalan
/ɲ/. In short, the distinction is made whether the palatal(ised) consonants are 'derived' by
allophonic processes or not. Referring to Miyawaki et al. (1974) and to cases like /sj/>[ʃʃ] in the English phrase bless you, it is assumed that ‘palatalisation as gestural blending’ may vary stylistically and idiosyncratically in its degree. It is obvious that the phoneme-allophone relationship underlies such one-stage and two-stage views of ‘palatalisation’: that is, the Japanese palatal(ised) consonants that occur before /i/ or /j/ are the results of unintended assimilation. It would be a mistake to assume that the different degrees of gestural overlap are indicative of the presence or absence of palatalisation.

It follows from what has been discussed that the issue of articulatory complexity is not a question of whether the palatal articulation is realised in a complex way or in a simplex way. What is at issue is how and why various degrees of intergestural blending between the two components of the tongue occur in the palatal(ised) segments. Results of previous studies suggest that the language-specific factors are involved in the patterning of blending. In addition, it is plausible to assume that different segments show different blending patterns. These problems are effectively conceptualised by our working definition of palatalisation as ‘a specialised use of a raising gesture of the tongue dorsum’. This parallels our formulation of the hypothesis in the previous section: the palatal(ised) consonants in Japanese involve two types of temporal co-production with the dorsum raising gesture.

In the following EPG experiments we shall examine the above hypothesis concerning the systematic articulatory patterning underlying the palatal(ised) consonants in Japanese. Strangely, previous studies limit their analysis to the articulation of an (alveolo)palatal nasal; we investigate the two palatalising environments [i] and [j], and the four target consonants that differ in places and manners of articulation and in the phonological status; [n], [g], [ɾ], and [k] in Ci and CyV syllables. Their spatio-temporal characteristics are quantified by various EPG parameters. We will argue that: (i) the (secondary) dorsal gestures for palatal(ised) consonants in Japanese incorporate certain aspects of the articulatory nature of the high front vowel /i/ and the approximant /j/; (ii) there are two ways of resolving antagonistic gestures by the tongue, articulatory blending and sequencing, the choice of which depends on the primary articulation; and (iii) speakers employ two contrastive timing strategies for the (secondary) dorsal gesture to effect palatalisation.

6.3 Data Collection and Analysis

Two experiments were conducted. The first was an attempt to define the canonical articulatory gesture for /i/ and /j/. The second was to examine how the (secondary) dorsal gesture interacts with the primary articulatory gesture in the target consonants.

6.3.1 Speech Items

The selected speech items consisted of (i) /ViV/ and /VjV/ symmetrical vowel contexts,
where the vowels were non-front vowels /a, o, u/; and (ii) /V₁CV₂/ and /V₁CjV₂/ sequences, where the V₁ was a fixed vowel /a/ and the V₂ was a changing vowel of /i, a, u/ and the consonants were /n, s, r, k/. All the target strings were disyllabic, or two-mora, words. It must be noted that the sequence of /asi/ is phonetically realised as [aci].

6.3.2 Measurements

The general articulatory properties were studied on the MAX contact pattern of the two vowels [i, j] and target consonants [n, ç, r̊, k̊]. The contact configurations were averaged over five repetitions and were represented by an EPG prototypical palatogram. For [ç] tokens, the frame corresponding to the fricative mid-point, specified with acoustic recordings, was regarded as the MAX point. The degree of the tongue dorsum involvement was examined both in the spatial and in the temporal domains.

For the spatial characteristics, the dorsum activities were inferred from the amount of contact in the central region at the Max point (see Figure 6-1). The comparison was made between the two vowels and between the four target consonants. Although the phonetic segments in question constitute different sound classes, vocoids and conoids, the tongue dorsum is certainly involved in both articulations. It can be asked whether the amount of central contact in [i, j] is different from that in [n, ç, r̊, k̊]. This question, as we discussed above, is considered in relation to the issue of articulatory complexity by Recasens: the same or larger degree of contact between the consonant ([n] in particular) and [i] or [j] is regarded as evidence for articulatory complexity. The problem increases in our experiments, since the palatal(ised) consonants in Japanese have two possible phonological origins, namely Ci and CyV syllables. If this phonological characteristic reflects the pattern of articulatory gestures, it is reasonable to suppose that there are some detectable traces of an [i]-like gesture or a [j]-like gesture during the production of the consonants in the two contrastive syllables. While it can be expected from the previous studies that there were few differences in the spatial contact patterns, particularly those of [n], further examination will be made for the other palatal(ised) consonants [ç, r̊, k̊]. This comparison includes the assessment of vowel-dependent coarticulatory effects on the tongue dorsum gesture.

For an overall characterisation of a contact pattern, the ‘centre of gravity (COG)’
was calculated for each of the target consonants \([n, ð, ɹ, k]\) both in Ci and CyV syllables. The COG is a measure to specify the location of the main concentration of activated electrodes across the whole palate, and the formula used in this experiment is as follows (Hardcastle \textit{et al.} 1991, 259).

\[
(3) \text{COG} = \frac{(0.5R_8 + 1.5R_7 + 2.5R_6 + 3.5R_5 + 4.5R_4 + 5.5R_3 + 6.5R_2 + 7.5R_1)}{\text{total number of contacts}}
\]

N.B. ‘R8, R7, etc’ = number of activated electrodes in Row 8, Row 7, etc.

Each row has a specific weight that increases towards more front rows. In this experiment the total number of contacts was measured as a percentage. The calculation was made from the first frame to the last frame of linguopalatal contact: a COG is specified for every single frame. Then, the values were averaged over five repetitions. Thus, the changing centre of electrode activation was plotted against the percentage of the overall contact: the peak of the COG corresponds to the main concentration of electrodes in a given consonant. Since the palatal(ised) consonants typically involve the tongue palate contact in the central region, it is reasonable to assume that they show the COG peaks somewhere around 3.5 (row 5) to 1.5 (row 7). We will examine the variability of the COG peak values of the four target consonants in terms of different places and manner of articulation and different syllable types.

The temporal characteristics were investigated in the articulatory trajectories that were computed from the raw EPG data varying time, corresponding to the regions, front, central, and back (see Figure 6-1 above). The pattern of the central trajectory was examined for the analysis of [i] and [j]. For the target palatal(ised) consonants, in contrast, intergestural coordination and timing were measured in terms of the interaction between two particular trajectories relevant for the given consonant: the front and central trajectories for \([n, ð, ɹ]\); and the central and back ones for \([k]\).

The diagram in (4) below schematically illustrates the central trajectory of [i] or [j]. The x-axis indicates time in frames and the y-axis indicates the percentage of on-electrodes in the region. The measurement points to quantify temporal characteristics are also shown.
The consistent cut-off point was set at 20% of tongue-palate contact in the central region. This value was determined by the visual inspection of the computation and illustrated a near-minimum activation value, or a minimally deviated starting-point, of the contact changes towards the target segment. Because the formation of a high back vowel [ui] involves the contact only at both sides of row 7 and 8, which are also used in the gestures corresponding to [i] or [j], the cut-off point was set to delimit the distinctive articulatory movement for the two vocoids. Such a 20% cut-off point was also applied to the temporal measurements of the palatal(ised) consonants.

As schematised in (4), the durational characteristics of [i] and [j] were quantified by several temporal points on the central trajectories. The following four parameters were obtained above the 20% cut-off point. First, REGION DURATION \( \langle a \rangle \) is the total duration of tongue-palate contact in the central region. Second, APPROACH DURATION \( \langle b \rangle \) is measured, in principle, from the frame showing the systematic contact change (or increase) towards the target segment to the first frame of the MAX constriction. Third, HOLD DURATION \( \langle c \rangle \) is the duration of the maximum constriction, or articulatory plateau. Lastly, RELEASE DURATION \( \langle d \rangle \) is calculated from the last frame of the MAX constriction to the frame showing the cessation of the contact decrease.

The measurement points for the four target consonants are illustrated in diagram (5), with the schematic representation of a palatalised alveolar consonant.

For this type of articulatory gesture we select the five parameters, the first four of which were proposed by Byrd (1994; 28f.), were selected with some modifications. First, SEQUENCE DURATION \( \langle a_1 \rangle \) is the total duration of linguopalatal contact for the two regions. Secondly, REGION DURATION is the duration of a particular region: the intervals \( \langle a_2 \rangle \) and \( \langle a_3 \rangle \) in (5) are the duration of the front region and that of the central region respectively. Both
sequence and region durations are measured above the 20% cut-off point consistent with the measurement of [i, j]. Secondly, \textit{TIME BETWEEN PEAKS} \(<b>\) indicates the time between the peak contact in one region and that in the other. The peak point is calculated as the temporal centre of the plateau of maximum contact. This illustrates the timing of the secondary palatal gesture. Thirdly, \textit{SEQUENCE OVERLAP} \(<c>\) is the percent of the total sequence duration during which contact occurred in both regions. It shows the degree of co-production in the front and central and in the central and back region respectively. The simple measurement, however, did not work properly in our particular case. Because of the continuous nature of tongue-palate contact, it was difficult to isolate co-production characteristics unique to a given primary gesture. To solve this problem, a \textit{cut-off point}\(^{5}\) was set at 50% of linguopalatal contact. This point was derived from visual inspection of the computation, and demonstrated a near-maximum activation value in the central region during the production of /j/ and the target consonants. Thus, the results were calculated above that point. Finally, \textit{PERCENT ANTICIPATION} is the percentage of the central contact at the left-edge (the first frame of MAX or the plateau) of the front contact as the value relative to the maximum percentage contact in the temporal centre of the plateau in the central region: the anticipation value is computed for the contact percentage at the point \(d1\) against that at the point \(d2\). This demonstrates to what extent the articulatory gesture for the (secondary) dorsal articulation has been anticipated. This is regarded as being indicative of the magnitude of palatalisation.

6.4 Results and Discussion

6.4.1 Experiment 1: the canonical EPG patterns of /i/ and /j/

We first consider the configurational and spatial characteristics of the two vocoids and then move on to their temporal characteristics. These results will later be compared to those of the palatal(ised) consonants in section 6.4.2.2. EPG prototypical palatograms of /i/ and /j/ at the MAX point are given in Figure 6-2. The electrodes activation was averaged over five repetitions and was represented by the five different shades depending on the activation frequency.

The two vocoids have a palatal articulation, with a primary constriction along the rows 4-8. This suggests that the constriction place and stricture types are largely similar between /i/ and /j/. It should be noted that the front region (rows 1-4) is relatively inactive in /j/ except in the vowel /u/ context. This is due to the typical characteristics of the articulation of Japanese /u/. It shows contact just at both sides of row 7 and 8, suggesting that the constriction place is not actually back but central. It is supposed that the posterior portion of the tongue medio-dorsum is involved in the constrictive approximation. Such a vowel gesture

\(^{5}\) Katrina Hayward first suggested this idea to me.
influences the tongue gesture for /j/ and results in the increase of the anterior contact. We will return to this point later. The results generally support the idea that /i/ and /j/ have the same constriction location in the central region. The EPG palatograms also suggest that the contact amount is generally larger in speaker MN than in speaker TM.

**Figure 6-2:** Linguopalatal Configurations at the Max Point for /i/ and /j/ in the Vowel Symmetrical Contexts: ‘black’ 100-80%; ‘grey’ 79-60%; ‘striped’ 59-40%; ‘dotted’ 39-20%; ‘white’ 19-0%.
Differences in the amount of contact in the central region were evaluated within and across the two vocoids. Figure 6-3 above summarises the average contact amount across two speakers on the left-hand side (pooled-speaker context) and across two speakers and vowel contexts on the right-hand side (pooled-speaker-vowel context). For the latter context, a two-way ANOVA with the two types of the vocoids and the three vowel environments as factors was conducted. The significant main effect was found for the vowel contexts (F(2,54)=9.089, p=0.00039), but not in the vocoid type (F(1,54)=2.032, p=0.1596). This was confirmed by a one-tailed t-test: [df=58, t=1.2235, p=0.1130]. A significant interaction between the two factors was found at the 5% significance level (F(2,54)=3.290, p=0.0448).

We shall now look more carefully into the effects of the vowel contexts and the speakers.

For the pooled-speaker context, the average contact amount of which is given on the left-hand side of Figure 6-3, the variability of electrode activation in the central region was evaluated separately for the vowel contexts and for the two vocoids. The results are reported in Tables 6-1 and 6-2. In Table 6-1 below, the different effects of the two vocoids and the two speakers are found in each of the three vowel contexts. The two factors have significant effects in the /a/ and /o/ symmetrical contexts, but not in the /u/ context. It can be seen that, while the contact degree is higher in /i/ than in /j/ for the /a/ and /o/ contexts, the opposite relation is found for the /u/ vowel context. These differences turn out to be statistically significant only in the /o/ context.

**Table 6-1: Two-way ANOVA Results for the Effects of Vowels in Pooled-Speaker Context**

<table>
<thead>
<tr>
<th>FACTORS</th>
<th>/i/ and /j/</th>
<th>/a/ a/</th>
<th>/o/ o/</th>
<th>/u/ u/</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 vocoids</td>
<td>F(1,16)</td>
<td>4.806, p=0.043</td>
<td>33.080, p=0.0001</td>
<td>2.184, p=0.158</td>
</tr>
<tr>
<td>2 speakers</td>
<td>F(1,16)</td>
<td>56.320, p=0.0001</td>
<td>61.800, p=0.0001</td>
<td>2.184, p=0.158</td>
</tr>
<tr>
<td>Interaction</td>
<td>F(1,16)</td>
<td>1.763, p=0.202</td>
<td>46.327, p=0.0001</td>
<td>2.184, p=0.158</td>
</tr>
<tr>
<td>t-TEST (unpaired)</td>
<td>df</td>
<td>t</td>
<td>t</td>
<td>t</td>
</tr>
<tr>
<td>Relation</td>
<td>/i/&gt;/j/</td>
<td>1.080, p=0.147</td>
<td>2.190, p=0.020</td>
<td>-1.389, p=0.090</td>
</tr>
</tbody>
</table>
Table 6-2: One-way ANOVA Results For The Effects Of The Contextual Vowels on Linguopalatal Contact For /i/ And /j/ In Pooled-Speaker Context

<table>
<thead>
<tr>
<th>Vowel Context</th>
<th>/i/</th>
<th>/j/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/iai/-/oio/</td>
<td>F(2,27)=0.865, p=0.432</td>
<td>F(2,27)=20.879, p&lt;0.0001</td>
</tr>
<tr>
<td>/iai/-/uiu/</td>
<td>t value (df==18) 0.651 1.552 0.574</td>
<td>p=0.261 p=0.069 p&lt;0.286</td>
</tr>
<tr>
<td>/oio/-/uiu/</td>
<td>p=0.657 4.654 6.531</td>
<td>p=0.259 p=0.0001 p&lt;0.0001</td>
</tr>
<tr>
<td>/aja/-/ojo/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/aja/-/uju/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ojo/-/uju/</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6-2 above shows the results of the separate analysis that was carried out for /i/ and /j/ to see the effects of the flanking vowels. Evidently the vowel contexts have a significant influence upon the articulatory gesture for /j/ in the central region, but not upon that for /i/. The amount of contact is significantly larger in /uju/ than in the other two sequences. In contrast, the activation of electrodes for /i/ is rather stable (see Figure 6-3), and the differences across the vowel context are not statistically significant. These results clearly demonstrate that, although the two vocoids certainly involve the tongue dorsum for their formation, the vowel /i/ as compared with the approximant /j/ has a stronger resistance to coarticulatory effects of the adjacent vowels. These results have a link to the frequently observed fact that /i/ is the strongest influencing vowel upon the consonantal articulation (e.g. Butcher & Weiher, 1976; Farnetani et al., 1985; Dagenais et al., 1994).

We observed that the amount of MAX tongue-palate contact for the two gestures is not significantly different across the pooled speaker-and-vowel contexts (see Figure 6-3). However, as visualised in Figure 6-2, the EPG palatograms indicate that there are noticeable (quantitative) differences in the actual formation of a constrictive approximation between the two speakers: the overall contact for /i/ tends to be larger than that for /j/ in speaker MN, while this tendency is indistinct in speaker TM. A statistical analysis clearly reflects idiosyncratic differences in contact degree. The results are reported in Table 6-3.

Table 6-3: Cross-speaker Differences For The Degree Of Central Region Contact

<table>
<thead>
<tr>
<th>TWO VOCOIDS</th>
<th>DEGREE OF CENTRAL REGION CONTACT (SD)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Speaker MN</td>
<td>Speaker TM</td>
</tr>
<tr>
<td>/iai/</td>
<td>54.18 (2.93)</td>
<td>45.80 (1.58)</td>
</tr>
<tr>
<td>/oio/</td>
<td>57.48 (1.83)</td>
<td>50.86 (3.47)</td>
</tr>
<tr>
<td>/uiu/</td>
<td>55.02 (1.83)</td>
<td>50.86 (3.47)</td>
</tr>
<tr>
<td>Pooled Vowel Context</td>
<td>55.56 (2.54)</td>
<td>47.48 (3.22)</td>
</tr>
<tr>
<td>/aja/</td>
<td>50.84 (1.87)</td>
<td>44.98 (1.83)</td>
</tr>
<tr>
<td>/ojo/</td>
<td>47.48 (2.30)</td>
<td>46.64 (1.87)</td>
</tr>
<tr>
<td>/uju/</td>
<td>55.02 (1.83)</td>
<td>55.02 (4.56)</td>
</tr>
<tr>
<td>Pooled Vowel Context</td>
<td>51.11 (3.69)</td>
<td>48.88 (5.34)</td>
</tr>
</tbody>
</table>

Two-way ANOVA

| 2 Vocoids   | F(1,24) 148.29, p<0.0001 | 14.56, p=0.16 |
| 3 Vowel Contexts | F(2,24) 21.25, p<0.0001 | 171.41, p<0.0001 |
| Interaction  | F(2,24) 64.79, p<0.0001 | 16.07, p=0.12 |

†The relation between the particular pairs was assessed by unpaired t-test (one-tailed): *=p<0.05; **=p<0.01; unmarked = non-significant
Results for the vocoid type were opposite for the two speakers. This factor has a significant effect upon the variability of contact degree in the central region for speaker MN, but not for speaker TM. For both speakers the contextual vowels have a significant effect. These results reflect the relationships between specific pairs within and across the two vocoids. The degrees of contact of /aia/ and /oio/ in the data of speaker MN are significantly higher than those of /aja/ and /ojo/ respectively. Such results were not obtained for the data of speaker TM, where there are no significant differences in contact degree between the two vocoids. It is worth noting that the electrode activation for /uju/ significantly increases for both speakers. It shows the highest percentage of the three /j/ sequences: the value is even higher than that of /uiu/ in the data of speaker TM. Such an effect of the vowel upon /j/ paradoxically yields a statistically insignificant difference against the contact degree of /uiu/ in the data of speaker MN. The result again suggests that the articulatory gesture for /j/ is more susceptible to the coarticulatory influence from the surrounding vowels, particularly /u/. This can be accounted for by the fact that the vowel /u/ ([u]) is articulated with (the posterior portion of) the tongue medio-dorsum. This gesture facilitates the palatal articulation for both /i/ and /j/, yielding the greater tongue-palate contact and also creating extensive contact in the anterior regions of the palate (see Figure 6-2). When large inter-speaker variability and coarticulatory effects are considered, it is difficult to assert that the constrictive approximation for the two vocoids is implemented quantitatively in the same manner. Nevertheless, the essential stricture type of the approximant can be attained for the two distinct palatal gestures regardless of the contextual vowel effects.

Turning to the second point, the temporal characteristics of the two vocoids are examined by the articulatory trajectories in the central region. As explained in the previous section, durational measurements referred to the period of maximum constriction and the trajectories were divided into the three durational segments corresponding to the articulatory phases, approach, hold, and release. The total of the three durations was considered as (central) region duration. The measurements were made above the 20% cut-off point of tongue-palate contact and the results were compared for each of the three phases and for all the phases in total.

The trajectories of the target sequences are shown in Figure 6-4-1 for speaker MN and in Figure 6-4-2 for speaker TM. The electrode activation is plotted along the y-axis against the x-axis that represents the time by the number of frames. It can be seen from the actual trajectories in Figures 6-4-1 and 6-4-2 that the central trajectory of /i/ shows a longer contact than that of /j/. There are noticeable effects of the contextual vowels on the contact duration of the two vocoids.
Figure 6-4-1: Front and Central Trajectories of /i/ and /j/ (Speaker MN)

The upper three panels are the trajectories for /i/ and the lower three panels are those for /j/.

X-axis = time in frames; Y-axis = measures of the percentage of contact (%)
Figure 6-4-2: Front and Central Trajectories of /i/ and /j/ (Speaker TM)
The upper three panels are the trajectories for /i/ and the lower three panels are those for /j/.
X-axis = time in frames; Y-axis = measures of the percentage of contact (%)
We begin by the analysis of the total duration of the central region. Figure 6-5 above presents average contact duration for the two speakers. The general trend is that the tongue-palate contact lasts longer in the vowel than in the approximant. The ANOVA results demonstrate significant effects with respect to the vocoid type, \( F(1,24)=239.701, p<0.0001 \) (MN) and \( F(1,24)=113.463, p<0.0001 \) (TM); and the three flanking vowels, \( F(2,24)=212.464, p<0.0001 \) (MN) and \( F(2,24)=169.313, p<0.0001 \) (TM). The interaction between the two factors was not significant: \( F(2,24)=2.149, p=0.138 \) (MN) and \( F(2,24)=1.698, p=0.204 \) (TM). The examination of specific pairs is summarised in Table 6-4 below.

<table>
<thead>
<tr>
<th>Relation</th>
<th>Speaker MN</th>
<th>Speaker TM</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/&lt;/i/&gt;</td>
<td>t value (df=8)</td>
<td>t value (df=8)</td>
</tr>
<tr>
<td>/aia/-/aja/</td>
<td>12.283</td>
<td>16.018</td>
</tr>
<tr>
<td>/oio/-/ojo/</td>
<td>16.018</td>
<td>5.483</td>
</tr>
<tr>
<td>/uio/-/uju/</td>
<td>5.483</td>
<td>8.022</td>
</tr>
<tr>
<td>/aia/-/aja/</td>
<td>5.031</td>
<td>5.031</td>
</tr>
<tr>
<td>/oio/-/ojo/</td>
<td>5.031</td>
<td>15.719</td>
</tr>
<tr>
<td>/uio/-/uju/</td>
<td>8.022</td>
<td>5.031</td>
</tr>
</tbody>
</table>

Evidently the total contact duration in the central region is longer in /i/ than in /j/. It must be noted that, although the two vocoids typically show similar durational intervals in the contexts of /a/ and /o/, the contact duration is substantially longer in the /u/ vowel context. Since this tendency is observed for both vocoids and for both speakers, we must examine the effects of the vowel contexts, particularly /u/.

The results of ANOVA and t-test, making a separate comparison for /i/ and /j/, are shown in Table 6-5 (next page). Significant differences in the flanking vowels are observed for both vocoids, although there is a tendency to level out the variability in the non-high vowels. In accordance with the significant tendency, the /u/ vowel appears to cause extensive coarticulatory influence upon the two vocoids in the spatio-temporal domains. Both /i/ and /j/, in that context, reveal the increase in contact degree and in contact duration. Whereas the spatial characteristic tends to be indistinguishable between the two gestures, they are typically distinguished from each other in the temporal characteristic (see Table 6-3). This leads us to the question of which articulatory stage(s) is crucial for the distinction between /i/ and /j/.
Table 6-5: One-way ANOVA results for the effect of the vowel contexts on the region duration

<table>
<thead>
<tr>
<th>Speaker</th>
<th>/i/</th>
<th>/j/</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-way ANOVA</td>
<td>F(2,12)=77.745, p&lt;0.0001</td>
<td>F(2,12)=153.156, p&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t-TEST†</th>
<th>/ai/-/oi/</th>
<th>/ai/-/ui/</th>
<th>/oi/-/iu/</th>
<th>/aj/-/ojo/</th>
<th>/aj/-/uju/</th>
<th>/ojo-/uju/</th>
</tr>
</thead>
<tbody>
<tr>
<td>t value (df=8)</td>
<td>1.188</td>
<td>9.248</td>
<td>9.461</td>
<td>2.272</td>
<td>14.00</td>
<td>13.683</td>
</tr>
<tr>
<td>p</td>
<td>0.134</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>0.022</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relation</th>
<th>/a/=/o/&lt;/u/</th>
<th>/o/&lt;/a/=/u/</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Speaker</th>
<th>/i/</th>
<th>/j/</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-way ANOVA</td>
<td>F(2,12)=64.841, p&lt;0.0001</td>
<td>F(2,12)=163.114, p&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t-TEST†</th>
<th>/ai/-/oi/</th>
<th>/ai/-/ui/</th>
<th>/oi/-/iu/</th>
<th>/aj/-/ojo/</th>
<th>/aj/-/uju/</th>
<th>/ojo-/uju/</th>
</tr>
</thead>
<tbody>
<tr>
<td>t value (df=8)</td>
<td>0.862</td>
<td>9.718</td>
<td>9.205</td>
<td>1.767</td>
<td>16.111</td>
<td>13.716</td>
</tr>
<tr>
<td>p</td>
<td>0.206</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>0.057</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relation</th>
<th>/a/=/o/&lt;/u/</th>
<th>/a/=/o/&lt;/u/</th>
</tr>
</thead>
</table>

† unpaired and one-tailed t-test: the symbol =, no significant; the symbol <, significant at the p<0.01 level

Table 6-6 (next page) reports the ANOVA results comparing the effects of the vocoid type and the flanking vowels for the three distinctive durational segments, approach, hold, and release. No significant differences were observed for the approach duration. The hold duration revealed significant differences in two factors and the release duration revealed those in the flanking vowels. These results are consistent between the two speakers. Differences in duration within and across the two vocoids were examined in more detail for particular pairs. In general, the results show similar trends in the three durational segments for both /i/ and /j/ and for the two speakers: (i) the temporal evolutions are similar between the two palatal gestures in both approach and release phases, in which there is little influence from the contextual vowels; (ii) the coarticulatory effects in the symmetrical contexts are related to the hold and release phases, where a durational hierarchy /a/=/o/</u/ is found for both /i/ and /j/; and (iii) even though the contact duration considerably increases in the /u/ contexts, the characteristic durational property of /i/>/j/ is retained.
### Table 6-6: ANOVA and t-Test for the Durational Differences in the Three Articulatory Stages between /i/ and /j/

<table>
<thead>
<tr>
<th>SPEAKER MN</th>
<th>DURATION</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>/aia/</strong></td>
<td>Approach (SD)</td>
<td>Hold (SD)</td>
<td>Release (SD)</td>
</tr>
<tr>
<td><strong>/oio/</strong></td>
<td>46.34 (5.23)</td>
<td><strong>66.86 (13.11)</strong></td>
<td><strong>50.97 (8.63)</strong></td>
</tr>
<tr>
<td><strong>/uiu/</strong></td>
<td>50.97 (11.37)</td>
<td>102.61 (16.38)</td>
<td>70.83 (9.24)</td>
</tr>
<tr>
<td>Pooled Vowel Context</td>
<td>47.66 (7.48)</td>
<td>79.44 (20.59)</td>
<td>55.82 (13.86)</td>
</tr>
<tr>
<td><strong>/aja/</strong></td>
<td>43.03 (5.23)</td>
<td>29.12 (5.92)</td>
<td><strong>45.01 (2.96)</strong></td>
</tr>
<tr>
<td><strong>/ojo/</strong></td>
<td>41.70 (3.77)</td>
<td><strong>24.49 (6.01)</strong></td>
<td><strong>39.05 (3.62)</strong></td>
</tr>
<tr>
<td><strong>/uju/</strong></td>
<td>49.65 (11.22)</td>
<td>56.27 (8.43)</td>
<td><strong>76.79 (11.07)</strong></td>
</tr>
<tr>
<td>Pooled Vowel Context</td>
<td>44.79 (7.99)</td>
<td>36.63 (15.84)</td>
<td>53.62 (18.31)</td>
</tr>
</tbody>
</table>

**Two-way ANOVA**

| 2 Vowoids | F(1,24) | 133.46, p=0.0001 | 43.37, p<0.0001 |
| 3 Vowel context | F(2,24) | 33.63, p<0.0001 | 3.24, p=0.056 |
| Interaction | F(2,24) | 0.26, p=0.76 | 3.24, p=0.056 |

<table>
<thead>
<tr>
<th>t-TEST for /i/-/j/</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>/aia/-/aja/</strong></td>
<td>0.872, p=0.204</td>
<td>10.606, p&lt;0.0001</td>
<td>0.163, <strong>/i//=/j/</strong></td>
</tr>
<tr>
<td><strong>/oio/-/ojo/</strong></td>
<td>1.605, p=0.073</td>
<td>6.566, p&lt;0.0001</td>
<td>0.923, <strong>/i//=/j/</strong></td>
</tr>
<tr>
<td><strong>/uiu/-/uju/</strong></td>
<td>0.185, p=0.428</td>
<td>5.622, p&lt;0.0001</td>
<td>0.163, <strong>/i//=/j/</strong></td>
</tr>
</tbody>
</table>

### Table 6-6: ANOVA and t-Test for the Durational Differences in the Three Articulatory Stages between /i/ and /j/ (continued)

<table>
<thead>
<tr>
<th>SPEAKER TM</th>
<th>DURATION</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>/aia/</strong></td>
<td>Approach (SD)</td>
<td>Hold (SD)</td>
<td>Release (SD)</td>
</tr>
<tr>
<td><strong>/oio/</strong></td>
<td>41.76 (12.07)</td>
<td><strong>62.89 (13.24)</strong></td>
<td><strong>38.39 (8.94)</strong></td>
</tr>
<tr>
<td><strong>/uiu/</strong></td>
<td>48.98 (9.76)</td>
<td><strong>111.87 (17.57)</strong></td>
<td><strong>68.84 (8.56)</strong></td>
</tr>
<tr>
<td>Pooled Vowel Context</td>
<td>42.80 (10.27)</td>
<td>83.85 (24.40)</td>
<td>47.44 (17.72)</td>
</tr>
<tr>
<td><strong>/aja/</strong></td>
<td>39.05 (9.18)</td>
<td>32.43 (13.72)</td>
<td><strong>41.70 (4.40)</strong></td>
</tr>
<tr>
<td><strong>/ojo/</strong></td>
<td>31.77 (2.96)</td>
<td><strong>30.45 (10.30)</strong></td>
<td><strong>43.69 (5.92)</strong></td>
</tr>
<tr>
<td><strong>/uju/</strong></td>
<td>40.38 (7.54)</td>
<td>61.56 (16.97)</td>
<td><strong>76.12 (16.87)</strong></td>
</tr>
<tr>
<td>Pooled Vowel Context</td>
<td>37.07 (7.63)</td>
<td>41.48 (19.57)</td>
<td>53.84 (19.07)</td>
</tr>
</tbody>
</table>

**Two-way ANOVA**

| 2 Vowoids | F(1,24) | 246.87, p=0.0078 | 3.21, p=0.08 |
| 3 Vowel context | F(2,24) | 175.66, p=0.0111 | 25.10, p<0.0001 |
| Interaction | F(2,24) | 94.58, p=0.29 | 0.02, p=0.97 |

<table>
<thead>
<tr>
<th>t-TEST for /i/-/j/</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>/aia/-/aja/</strong></td>
<td>0.258, p=0.401</td>
<td>7.892, p&lt;0.0001</td>
<td>1.482, p=0.088</td>
</tr>
<tr>
<td><strong>/oio/-/ojo/</strong></td>
<td>1.786, p=0.055</td>
<td>4.322, p&lt;0.0001</td>
<td>1.104, p=0.055</td>
</tr>
<tr>
<td><strong>/uiu/-/uju/</strong></td>
<td>1.559, p=0.078</td>
<td>4.603, p&lt;0.0001</td>
<td>0.860, p=0.078</td>
</tr>
</tbody>
</table>

†The relation between the particular pairs was assessed by unpaired t-test (one-tailed): **=p<0.01; *-=p<0.05; unmarked = no significant
In summary, we have examined the spatio-temporal characteristics of /i/ and /j/ in the vowel symmetrical contexts of /a/, /o/, and /u/. We have found evidence that the two vocoids systematically differ in their spatial and temporal properties. For the spatial properties: (i) /i/ and /j/ are articulated with the tongue medio-dorsum pushing against the palatal region (rows 5-8); (ii) the amount of tongue-palate contact in the central region varies considerably with speakers and vowel contexts; and (iii) the articulatory gesture for /j/ is more susceptible to the coarticulatory effects from the surrounding vowels. Remarkable contextual effects were observed in the /j/-gesture in the /uju/ sequence: the contact extends towards the anterior regions of the palate and the contact substantially increases in the central region. Such coarticulatory effects appear to be related to the typical production characteristics of Japanese /u/, as well as the articulatory nature of /j/.

For the temporal properties: (i) the overall duration of the contact in the central region is characteristically longer in /i/ than in /j/ regardless of the vowel contexts; (ii) the vowel effects on the contact duration appear to be limited to the hold and release phases, there is little difference in the approach duration within and across the two vocoids; (iii) their hold and release durations are longer in the /u/ vowel context; and (iv) the durational difference of the hold phase is the distinctive temporal property of /i/ and /j/, the duration into and out of the hold phase is not significantly different between the two palatal gestures. The central trajectory of /i/ shows a more stable and longer contact (i.e. articulatory plateau) than that of /j/. In other words, /j/ is a temporally compressed version of /i/ in terms of the hold phase duration, or articulatory plateau.

The spatio-temporal characteristics of /i/ and /j/ that we found are directly related to our working definition that palatalisation is a specialised use of a raising gesture of the tongue dorsum. It seems reasonable to develop our hypothesis that palatalisation incorporates some aspects of the articulatory gesture for /i/ and /j/. In the next section we shall examine how the /i/-like and /j/-like gesture manifests during the articulation of the palatal(ised) consonants.

6.4.2 Experiment 2: the target consonants in the palatalising environments

6.4.2.1 Spatial Characteristics of Palatal(ised) Consonants in Ci and CyV Syllables

EPG prototypical palatograms averaged over five repetitions are presented in Figure 6-6 and 6-7. There is no substantial divergence in the configuration of the tongue-palate contact for the target consonants [p, ç, k]. A large contact area over the surface of the palate is achieved by the extension of the side contact towards the mid-sagittal line.

For the production of [p], the posterior contact (rows 4-8) increases and the frontmost contact tends to be withdrawn. The closure is made by pressing the anterodorsum of the tongue against postalveolar and palatal areas (rows 2-5), causing the tongue tip (and the anterior portion of the blade) to be directed downwards. Consequently, the place of primary
constriction shifts backwards as compared with [n]. These spatial configurations characterise the consonant as an alveopalatal nasal [n]. For the alveolopalatal fricative [c], the maximum narrowing is formed at the rows 2 or 3-5. Notice here that, although both [p] and [c] are described with the same place label, the main constriction of the latter is formed slightly further back: we could identify [p] as a palatalised alveolar; but [c] as an alveolarised palatal. Also, it is worth mentioning that there are differences in the length of the constricted region between the speakers. The anterior end is row 3 for speaker MN but is row 2 for speaker TM. This (idiosyncratic) anterior realisation must be related to the variability in the posterior end of the articulatory channel in speaker TM: the narrowest region is regularly formed by pressing the blade at rows 2-3 and this is accompanied by the pre- (and presumably medio-) dorsum constriction ranging from row 4 to 5. The palatalised alveolar tap [r'] is articulated with the tongue tip/blade against the alveolar area (rows 2-3). When adjacent to
the open vowel /a/, there is little contact in the posterior region. This suggests that the two components of the tongue have no conflict. The palatalising effects appear at the posterior region, the magnitude of which is less than that of the other consonants. This is particularly attributed to the manner requirements. The palatalised velar consonant [kʰ] shows another effect, namely fronting of the constriction place, where complete contact across the back of the palate (row 7-8) is made and the side contacts extend further forward. There are differences in the length of the complete occlusion: row 8 for speaker MN and rows 7-8 for speaker TM. This suggests that the constriction place is more fronted in the sense that the posterior portion of the tongue medio-dorsum, as well as the tongue post-dorsum, is effectively pressed against the back of the palate during the production. The overall configurational patterns are largely similar to each other within the limits of each speaker.
The first question to be discussed is the effects of the two speakers and the flanking vowels (V₂) upon the variability of the contact degree for the target consonants, in relation to the variation between the two types of syllables, namely C₁ and C₂YV. Mean percentage with standard deviation of the central contact at MAX in the pooled-speaker context is given in Figure 6-8. Evidently the percentage values of the consonants [p, c, k'] are more than twice as high as those of [r']. Among the three vowels the /u/ context reveals a higher activation value across all the consonants. This suggests that the raising gesture of the tongue body is constrained consonant-specifically and is susceptible to the anticipatory coarticulation of the following vowel. An ANOVA was conducted to test for differences in the speakers and in the vowel contexts among the mean activation values of the four target consonants. In the upper half of Table 6-7 we find that both factors have significant effects upon the variability of the central contact, except for the consonant [r'] where the contextual effects are not significant. Differences between the particular pairs of the syllables were further compared by t-test. The results are given in the lower half of Table 6-7. Significant differences were found between Cya and Cyu syllables with [p] and [k'] and between C₁ and Cyu syllables with [k']. The consonants [c] and [r] reveal no significant differences in any of the combinations.

<table>
<thead>
<tr>
<th>Two-way ANOVA</th>
<th>[p]</th>
<th>[c]</th>
<th>[r]</th>
<th>[k']</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 SPEAKERS</td>
<td>F(1.24)</td>
<td>6.405, p=0.018</td>
<td>92.95, p&lt;0.0001</td>
<td>4.820, p=0.038</td>
</tr>
<tr>
<td>3-SYLLABLE CONTEXTS</td>
<td>F(2.24)</td>
<td>4.297, p=0.025</td>
<td>3.769, p=0.037</td>
<td>1.577, p=0.227</td>
</tr>
<tr>
<td>INTERACTION</td>
<td>F(2.24)</td>
<td>1.233, p=0.303</td>
<td>3.316, p=0.053</td>
<td>0.534, p=0.592</td>
</tr>
<tr>
<td>t-TEST†</td>
<td>df</td>
<td>[n]</td>
<td>[c]</td>
<td>[r]</td>
</tr>
<tr>
<td>SYLLABLE TYPES</td>
<td>18</td>
<td>1.097</td>
<td>0.143</td>
<td>0.750</td>
</tr>
<tr>
<td>C₁ - Cya</td>
<td>18</td>
<td>1.599</td>
<td>0.063</td>
<td>0.557</td>
</tr>
<tr>
<td>C₁ - Cyu</td>
<td>18</td>
<td>2.746</td>
<td>0.006</td>
<td>1.182</td>
</tr>
</tbody>
</table>

† unpaired and one-tailed
Next, we examined each activation value in the central region separately for the two speakers, to study the inter-speaker variability. Table 6-8 below summarises means, standard deviations and the results of the analysis performed for each speaker. No significant differences in the vowel contexts across the consonants were obtained for speaker MN. Accordingly the comparisons of particular syllable-pairs show no significant differences either. In contrast, significant differences, except for the consonant \([r^3]\), were found for speaker TM. The relationships between particular syllable-pairs demonstrate that: (i) no significant differences were obtained for \([r^3]\) in any of the examined combinations; (ii) significant differences are found in the comparisons between Ci and Cya syllables with \([c]\) and between Ci and Cyu syllables with \([k^3]\); and (iii) Cya and Cyu syllables significantly differ in all the consonants except for \([r^3]\).

From the above examinations of the electrode activation values in the pooled-speaker and in the separate-speaker settings, three general points become clear. First, as we expected, the spatial manifestation of palatalisation in Ci and CyV syllables is largely similar to each other within the limit of each target consonant. The tongue medio-dorsum actively participates in the constriction formation for the two alveolopalatals \([p]\) and \([ç]\) and the velar \([k^3]\). Second, the inter-speaker disparities for the relationship between particular syllable-pairs

<table>
<thead>
<tr>
<th>TABLE 6-8: DIFFERENCES BETWEEN THE VOWEL CONTEXTS AND BETWEEN THE SYLLABLE PAIRS FOR THE CONTACT AMOUNT IN THE CENTRAL REGION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPEAKER MN</strong></td>
</tr>
<tr>
<td>([a_i])</td>
</tr>
<tr>
<td>([a_a])</td>
</tr>
<tr>
<td>([a_u])</td>
</tr>
<tr>
<td><strong>ONE-WAY ANOVA</strong></td>
</tr>
<tr>
<td><strong>3-VOWEL CONTEXTS</strong></td>
</tr>
<tr>
<td><strong>t-TEST</strong></td>
</tr>
<tr>
<td><strong>SYLLABLE TYPES</strong></td>
</tr>
<tr>
<td>Ci - Cya</td>
</tr>
<tr>
<td>Ci - Cyu</td>
</tr>
<tr>
<td>Cya - Cyu</td>
</tr>
<tr>
<td><strong>SPEAKER TM</strong></td>
</tr>
<tr>
<td>([a_i])</td>
</tr>
<tr>
<td>([a_a])</td>
</tr>
<tr>
<td>([a_u])</td>
</tr>
<tr>
<td><strong>ONE-WAY ANOVA</strong></td>
</tr>
<tr>
<td><strong>3-VOWEL CONTEXTS</strong></td>
</tr>
<tr>
<td><strong>t-TEST</strong></td>
</tr>
<tr>
<td><strong>SYLLABLE TYPES</strong></td>
</tr>
<tr>
<td>Ci - Cya</td>
</tr>
<tr>
<td>Ci - Cyu</td>
</tr>
<tr>
<td>Cya - Cyu</td>
</tr>
</tbody>
</table>

†The bracket reflects the significant difference obtained by unpaired t-test (one-tailed).
are related to the idiosyncratic strategies for V-to-C coarticulation. The smaller activation value in Cya syllables compared to that in Cyu syllables for speaker TM can be accounted for both by the anticipatory coarticulation (post-dorsum lowering) for the following low vowel and by the articulatory gesture typical to [tu]. The results for [ç] in particular demonstrate that the constriction location and the V-to-C coarticulation are closely related to each other. The narrowest articulatory channel is formed slightly further forward in speaker TM. This appears to give a greater degree of freedom to the dorsum activities. In contrast, the disparity in the [kʰ] results comes, not only from the spatial, but also from the temporal aspects of coarticulatory strategies. We shall return to this point later. These idiosyncratic differences are negligible but do not affect the first point for the similarities in configurational patterning between Ci and CyV syllables. Rather, individual variations clarify the fact that the tongue dorsum activity varies with the location of a primary constriction in the palatal(ised) consonants. The third point is the unusual characteristics in the [f-' œ] tokens. As discussed in Chapters 3 and 4, the raising gesture of the tongue body is tightly constrained in order for the tongue tip/blade to make a distinctive flicking or transient movement. In other words, the dorsum raising gesture to effect palatalisation is highly antagonistic to a tapping gesture. Thus, the central region revealed the extremely small values of electrode activation as compared to the other target consonants. This was consistent between the two speakers.

6.4.2.2 Testing Articulatory Complexity

Let us consider differences in the electrode activation between the articulatory gesture for /i, j/ and that for the palatal(ised) consonants. Recasens et al. (1994) found that in Catalan lingual contact at the medio and postpalatal zones for [n] is less extensive and more variable than that for [j]. Assuming that the contact degree in that region is indicative of articulatory complexity, the authors claim that the alveolopalatal [n] (in Catalan) is a simple segment (see section 6.2.2). We carried out such a comparison to seek differences between the different phonological origins (i.e. the two distinctive palatalising environments) for particular pairs across the target consonants. The vowel contexts were utilised as the basis for the multiple comparison: the sequences of /aCi/ and /aCjα/ were compared with those of /aia/ and /aja/, while the /aCju/ sequences were compared with the /uia/ and /uju/ sequences. The results are given in Table 6-9 (next page).

The ANOVA results for the segment type (i.e. /i/, /j/, and a palatal(ised) consonant) reflect the general tendency that the activation value in the central region is specific to each segment. A further comparison between a palatal(ised) consonant and /i/ or /j/ demonstrates that the contact values of palatal(ised) consonants are larger than, or the same as, those of /i, j/. This result, if we apply the above view of Recasens et al. (1994), serves as evidence that the four palatal(ised) consonants in Japanese are complex segments: the contact degree of the palatal(ised)
consonants always exceeds that of the two vocoids. Because of the production constraints discussed before, the consonant [r'], the activation value of which is smaller than those of /i/ and /j/, is not considered as a simplex segment. Also, it is further confirmed that the two versions of palatal(ised) consonants are not efficiently distinguished from each other either by the constriction shape or by the constriction degree in the central region.

**TABLE 6-9: COMPARISON OF THE CONTACT AMOUNT IN THE CENTRAL REGION BETWEEN /i, j/ AND THE PALATAL(ISED) CONSONANTS (speaker MN for the upper panel and speaker TM for the lower)**

<table>
<thead>
<tr>
<th>Speaker MN</th>
<th>/i/ and /j/</th>
<th>/aia/54.18%</th>
<th>/uiu/ 55.02%</th>
<th>/aja/ 50.84%</th>
<th>/uju/ 55.02%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CONSONANTS</strong></td>
<td>%</td>
<td><strong>ONE-WAY ANOVA</strong></td>
<td><strong>t-test (df = 8)</strong></td>
<td><strong>t-test (df = 8)</strong></td>
<td><strong>Relation</strong></td>
</tr>
<tr>
<td><strong>/n/</strong></td>
<td>[aii] 64.18</td>
<td>27.53**</td>
<td>4.690**</td>
<td>7.102**</td>
<td><strong>/n/&gt; /i, /j/</strong></td>
</tr>
<tr>
<td></td>
<td>[ai] 63.34</td>
<td>9.30**</td>
<td>2.559**</td>
<td>3.641**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[aiu] 66.66</td>
<td>20.67**</td>
<td>4.799**</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>/s/</strong></td>
<td>[aii] 59.98</td>
<td>12.22**</td>
<td>2.720*</td>
<td>4.866**</td>
<td><strong>/s/&gt; /i, /j/</strong></td>
</tr>
<tr>
<td></td>
<td>[ai] 60.82</td>
<td>22.20**</td>
<td>3.981**</td>
<td>7.514**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[aiu] 61.66</td>
<td>21.50**</td>
<td>5.656**</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>/r/</strong></td>
<td>[ar'i] 25.84</td>
<td>229.65**</td>
<td>18.187**</td>
<td>21.044**</td>
<td><strong>/r/&gt; /i, /j/</strong></td>
</tr>
<tr>
<td></td>
<td>[ar'a] 25.86</td>
<td>86.22**</td>
<td>10.253**</td>
<td>9.715**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[ar'u] 28.34</td>
<td>189.09**</td>
<td>15.172**</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>/k/</strong></td>
<td>[ak'i] 64.16</td>
<td>13.81**</td>
<td>3.199**</td>
<td>4.512**</td>
<td><strong>/k/&gt; /i, /j/</strong></td>
</tr>
<tr>
<td></td>
<td>[ak'a] 59.98</td>
<td>18.40**</td>
<td>3.478**</td>
<td>6.881**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[ak'u] 64.16</td>
<td>14.24**</td>
<td>4.011**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speaker TM</th>
<th>/i/ and /j/</th>
<th>/aia/54.18%</th>
<th>/uiu/ 55.02%</th>
<th>/aja/ 50.84%</th>
<th>/uju/ 55.02%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CONSONANTS</strong></td>
<td>%</td>
<td><strong>ONE-WAY ANOVA†</strong></td>
<td><strong>t-test (df = 8)‡</strong></td>
<td><strong>t-test (df = 8)‡</strong></td>
<td><strong>Relation</strong></td>
</tr>
<tr>
<td><strong>/n/</strong></td>
<td>[aii] 60.00</td>
<td>14.95**</td>
<td>3.845**</td>
<td>4.041**</td>
<td><strong>/n/&gt; /i, /j/</strong></td>
</tr>
<tr>
<td></td>
<td>[ai] 54.18</td>
<td>26.86**</td>
<td>5.621**</td>
<td>5.945**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[aiu] 65.02</td>
<td>20.80**</td>
<td>7.592**</td>
<td>4.378**</td>
<td></td>
</tr>
<tr>
<td><strong>/s/</strong></td>
<td>[aii] 52.52</td>
<td>23.01**</td>
<td>5.383**</td>
<td>5.731**</td>
<td><strong>/s/&gt; /i, /j/</strong></td>
</tr>
<tr>
<td></td>
<td>[ai] 47.48</td>
<td>2.184</td>
<td>1.345</td>
<td>1.900</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[aiu] 53.34</td>
<td>1.46</td>
<td>1.127</td>
<td>0.655</td>
<td></td>
</tr>
<tr>
<td><strong>/r/</strong></td>
<td>[ar'i] 25.00</td>
<td>355.04**</td>
<td>29.415**</td>
<td>24.365**</td>
<td><strong>/r/&gt; /i, /j/</strong></td>
</tr>
<tr>
<td></td>
<td>[ar'a] 22.50</td>
<td>133.23**</td>
<td>12.896**</td>
<td>12.126**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[ar'u] 25.00</td>
<td>120.64**</td>
<td>16.624**</td>
<td>14.721**</td>
<td></td>
</tr>
<tr>
<td><strong>/k/</strong></td>
<td>[ak'i] 63.34</td>
<td>88.60**</td>
<td>10.178**</td>
<td>10.357**</td>
<td><strong>/k/&gt; /i, /j/</strong></td>
</tr>
<tr>
<td></td>
<td>[ak'a] 64.18</td>
<td>158.49**</td>
<td>14.723**</td>
<td>14.594**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[ak'u] 73.34</td>
<td>38.66**</td>
<td>8.533**</td>
<td>6.218**</td>
<td></td>
</tr>
</tbody>
</table>

†**=p<0.01; *=p<0.05; unmarked = non-significant

‡The relation between the particular pairs was assessed by unpaired t-test (one-tailed): **=p<0.01; *=p<0.05; unmarked = no significant
6.4.2.3 The Centre of Gravity

This discovery, in turn, leads us to the question of what constitute the similarities among the articulatory gestures for the palatal(ised) consonants. This enquiry was examined by comparing a dynamic parameter of lingual contact, namely the centre of gravity (COG). As described before, the COG is an EPG parameter that specifies the location of the main concentration of activated electrodes across the whole palate (Hardcastle et al. 1991). The COG data in Figure 6-10 (next page) was calculated from the first to the last frame of linguopalatal contact and was averaged over five repetitions. The percentage contact of the entire palate was plotted on the y-axis against the changing centre of electrode activation on the x-axis. We examined the variability of the COG peak values that corresponds to the maximum activation of electrodes in a given consonant.

The peak COG value averaged across the two speakers is given in Figure 6-9, where the values of /ana/ and /ara/ are included for reference. The one-way ANOVA comparing differences between the COG peak values turned out to be significant: for the four target consonants, [F(12, 117) = 142.747, p<0.0001]; for [p, c, k'], [F(8, 81) = 109.655, p<0.0001]; and for [p, c], [F(5, 54) = 18.580, p<0.0001]. Differences between the speakers and the vowel contexts were further assessed for each consonant separately. The results are summarised in Table 6-10.

![Figure 6-9: COG Peak Values for the Four Palatal(ised) Consonants Averaged Across the Two Speakers](image)

**Figure 6-9: COG Peak Values for the Four Palatal(ised) Consonants Averaged Across the Two Speakers:** The relation between particular pairs was assessed by unpaired t-test (one-tailed), in which /ana/ and /ara/ are not included. A bracket indicates a significant difference, the criterion for which is p<0.01.

X-axis = the target palatal(ised) consonants: two non-palatalised consonants are included as a reference.
Y-axis = measures of the centre of the articulatory gravity
Figure 6-10: COG Contours Averaged Over Five Repetitions For The Four Palatal (I.S.E.D) Consonants: a rising arrow=approaching to the MAX; a falling arrow=releasing.
The results reveal the consonant-related patterns. Significant effects were found for the two speakers in all the consonants except for \([k^i]\); and for the flanking vowels in \([r^i]\] and \([k^i]\). Differences between the particular syllable-pairs turned out to be non-significant for \([p]\) and \([c]\], but significant for \([r^i]\] and \([k^i]\). It is reasonable to suppose that this consonant-related tendency comes from the difference in the manner and place of articulation. The conflict between the tongue tip/blade and the dorsum in the tap articulation affects the timing of the co-production, and may also result in the instability of the peak COG. For the velar, as observed in Figures 6-6 and 6-7, the anterior extension of contact characteristically varies with the vowel contexts. This is reflected by the variability of the COG peak values.

Table 6-11 below summarises the means of the COG peak values for each speaker in the upper panel and the comparisons between particular syllable-pairs in the lower panel.

<table>
<thead>
<tr>
<th>TABLE 6-10: EFFECTS OF THE SPEAKER AND THE VOWEL CONTEXTS ON THE COG PEAK VALUES (POOLED-SPEAKER CONTEXT)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Two-way ANOVA</strong></td>
</tr>
<tr>
<td><strong>FACTORS</strong></td>
</tr>
<tr>
<td>2 Speakers F(1,24)                                          52.290</td>
</tr>
<tr>
<td>3 Vowels F(2,24)                                           3.065</td>
</tr>
<tr>
<td>Interaction F(2,24)                                         1.346</td>
</tr>
<tr>
<td><strong>t-TEST</strong></td>
</tr>
<tr>
<td>(df=18)</td>
</tr>
<tr>
<td>Ci-Cya                                                      1.117</td>
</tr>
<tr>
<td>Ci-Cyu                                                      0.165</td>
</tr>
<tr>
<td>Cya-Cyu                                                     1.521</td>
</tr>
</tbody>
</table>

The results reveal the consonant-related patterns. Significant effects were found for the two speakers in all the consonants except for \([k^i]\); and for the flanking vowels in \([r^i]\] and \([k^i]\). Differences between the particular syllable-pairs turned out to be non-significant for \([p]\) and \([c]\], but significant for \([r^i]\] and \([k^i]\). It is reasonable to suppose that this consonant-related tendency comes from the difference in the manner and place of articulation. The conflict between the tongue tip/blade and the dorsum in the tap articulation affects the timing of the co-production, and may also result in the instability of the peak COG. For the velar, as observed in Figures 6-6 and 6-7, the anterior extension of contact characteristically varies with the vowel contexts. This is reflected by the variability of the COG peak values.

Table 6-11 below summarises the means of the COG peak values for each speaker in the upper panel and the comparisons between particular syllable-pairs in the lower panel.

<table>
<thead>
<tr>
<th>TABLE 6-11: MEANS AND STANDARD DEVIATIONS OF THE COG PEAK VALUES (INDIVIDUAL SPEAKERS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPEAKER</strong></td>
</tr>
<tr>
<td>MN</td>
</tr>
<tr>
<td>([a_i])</td>
</tr>
<tr>
<td>([a_a])</td>
</tr>
<tr>
<td>([a_u])</td>
</tr>
<tr>
<td><strong>AVERAGE</strong></td>
</tr>
<tr>
<td><strong>TM</strong></td>
</tr>
<tr>
<td>([a_i])</td>
</tr>
<tr>
<td>([a_a])</td>
</tr>
<tr>
<td>([a_u])</td>
</tr>
<tr>
<td><strong>AVERAGE</strong></td>
</tr>
</tbody>
</table>

| **SPEAKER**                                                  | \([n]\) | \([c]\) | \([r^i]\) | \([k^i]\) |
|---------------------------------------------------------------|
| **MN**                                                      |        |        |          |          |
| \([a_i]\)                                                   | 2.066  | 0.072  | 0.407    | 0.347    | 1.124    | 0.146    | 2.950   | 0.009    |
| \([a_a]\)                                                   | 0.467  | 0.326  | 0.313    | 0.380    | 4.022    | 0.001    | 1.988   | 0.041    |
| \([a_u]\)                                                   | 1.447  | 0.092  | 0.155    | 0.440    | 2.941    | 0.009    | 4.231   | 0.001    |
| **TM**                                                      |        |        |          |          |
| \([a_i]\)                                                   | 0.400  | 0.349  | 2.104    | 0.034    | 2.132    | 0.032    | 2.950   | 0.009    |
| \([a_a]\)                                                   | 1.774  | 0.056  | 1.269    | 0.119    | 1.451    | 0.092    | 1.988   | 0.041    |
| \([a_u]\)                                                   | 2.854  | 0.010  | 0.302    | 0.384    | 2.930    | 0.009    | 4.231   | 0.001    |

274
The pattern of significant differences is largely similar to that in the comparisons under the pooled-speaker setting (Table 6-10). It can be seen that there are some individual variations. This is not surprising. Since the COG was calculated as a function of the percentage contact of the whole palate, both the coarticulatory effects from the vowel contexts and idiosyncratic preferences of the constriction location influence the peak values of the COG. We would like to focus on the distributions of the COG values with reference to the palate. This is illustrated as follows:

The distributions of the COG peak values, when they are averaged across the vowel contexts, are very close and assemble around the central position of the palate, namely the intersection between row 4 and 5. The COGs of /ana/ and /ara/, for instance, are much higher and are allocated to around rows 2-3(4): 5.52 (s.d. 0.28: speaker MN) and 6.28 (s.d. 0.15: speaker TM) for /ana/; and 4.87 (s.d. 0.06: MN) and 5.16 (s.d. 0.01: TM) for /ara/. Since each consonant has its own distinctive interaction between the active and passive articulators, the COG peak values serve as a good indicator of the ‘palatalised’ articulation with reference to the overall palate. Thus, it is reasonable to conclude that the palatal(ised) consonants have their centre of articulatory gravity around the postalveolar, pre- and medio-palatal region on the palate and around the back of the blade, pre- and medio-dorsum region on the tongue.

6.4.2.4 Intergestural Coordination: analysis of five temporal parameters

Having examined the spatial and dynamic characteristics of the palatalised consonants, we now turn to the investigation of their temporal characteristics. Of particular interest are the questions of how the intergestural coordination is attained in the given palatal(ised) consonant; and whether there is a relationship between the two kinds of linguistic specification (i.e. Ci and CyV) and their patterns of articulatory implementation. Various EPG parameters were measured for the articulatory trajectories relevant for the given palatal(ised) consonant: the front and central trajectories for [n], [e], and [ʃ], and the central and back trajectories for [k']. The trajectories for the target syllables spoken by the two speakers are shown in Figures 6-11 through 6-14 and are presented in pairs for each consonant. Time in frames is on the x-axis and the percentage of electrode activation is on the y-axis. Each panel contains five tokens of the given target syllable.
Figure 6-11: Articulatory Trajectories of the Syllables with [ŋ]
X-axis = time in frames; Y-axis = measures of the percentage of contact (%)
Figure 6-12: Articulatory Trajectories of the Syllables with [ç]
X-axis = time in frames; Y-axis = measures of the percentage of contact (%)
FIGURE 6-13: ARTICULATORY TRAJECTORIES OF THE SYLLABLES WITH \([r']\)

X-axis=time in frames; Y-axis=measures of the percentage of contact (%)
Figure 6-14: Articulatory Trajectories of the Syllables With [kʰ]

(a) Speaker MN

(b) Speaker TM

Time (Frames)
The temporal aspects of contact pattern, as explained before, were quantified by the four indices: (i) sequence and region durations; (ii) time between the peaks of the two regions; (iii) sequence overlap; and (iv) percentage of the anticipatory raising gesture of the tongue dorsum. The measurements and the results of statistical analyses are presented in this order.

First, we consider the durations of the given two trajectories. **SEQUENCE DURATION** is the total duration of linguopalatal contact for the two regions. **REGION DURATION** is the duration of a particular region. Both sequence and region durations were measured above the 20% cut-off point consistent with the measurement of [i, j]. If the palatal(ised) consonants involve certain aspects of the [i] and [j] gestures, then the sequence and the central region durations reflect the typical temporal properties of the two gestures that we examined in section 6.4.1. In contrast, the movement for the primary constriction can be inferred from either the front or the back region duration relevant for the given palatal(ised) consonant.

Figure 6-15 below presents the mean sequence duration of the target consonants for each speaker. Evidently the total duration of the contact varies with the vowel contexts: the one tendency common to both speakers is that the duration increases in the order of /a/, /u/, /i/. The two-way ANOVA results demonstrate significant differences for the consonant type, [F(3,48)=226.886, p<0.0001] (MN) and [F(3,48)=216.333, p<0.0001] (TM); for the three flanking vowels, [F(2,48)=247.726, p<0.0001] (MN) and [F(3,48)=118.993, p<0.0001] (TM). Significant interactions for those two factors were also found: [F(6,48)=3.337, p<0.0001] (MN) and [F(6,48)=9.661, p<0.0001].

![Figure 6-15: Mean Sequence Durations of the Four Palatal(ised) Consonants (msec)](image)

X-axis = the target consonants; Y-axis = measures of the duration in milliseconds
Further examinations of the contextual effects were conducted for each of the four consonants, and the results were given in Table 6-12 above. Generally speaking, the sequence duration of all the consonants lasts substantially longer in the /i/ vowel context (i.e. Ci syllables). For the consonant [ç] the contact duration of the Ci and Cyu syllables tends to be indistinguishable (speaker TM). Also, the velar consonant exhibits idiosyncratic variation in duration. These consonant-related tendencies are due to the place of articulation: that is, the ways in which the active and passive articulators interact. We will return to this point later. Another point to note concerns the consonant [ɾ]. As we have shown when considering its spatial characteristics, the tongue dorsum is tightly constrained in the production of [ɾ] and this yields a smaller amount of tongue-palate contact in the central region. This is also evident in the temporal articulatory trajectories in Figure 6-13: the electrodes in the central region are not activated until the activation in the front region ceases. The results of the sequence duration, however, reveal a durational hierarchy similar to that of the other consonants. It appears that the coarticulatory effects of a tap [ɾ] upon the surrounding vowels (the following vowel in our case) lie, to a large extent, in the spatial domain, rather than the temporal domain.

Next, we consider the contact durations of the front (or back) and the central regions: the front region for the alveolo-palatal and the tap and the back region for the velar. The front (or back) region is involved in the main constriction and illustrates general durational patterns for the target consonants. The tongue dorsum activity is inferred from the durations of the central region contact and can be compared to those of the [i] and [j] gestures. The summary of the measurements for each parameter is presented in Figures 6-16 and 6-17 respectively.
N.B. The value of a palatalised velar represents the duration of the ‘back’ region.

(a) Speaker MN

(b) Speaker TM

**Figure 6-16: Mean Front Region Durations and Standard Deviations for the Four Palatal (ISID) Consonants (msec):** X-axis = the target consonants; Y-axis = measures of the duration in milliseconds.

**Figure 6-17: Mean Central Region Durations and Standard Deviations for the Four Palatal (ISID) Consonants (msec):** X-axis = the target consonants; Y-axis = measures of the duration in milliseconds.
Tables 6-13 and 6-14 below display the results of the statistical test for the effects of the vowel contexts upon the duration of each region. In general, the anticipatory effects are significant for the durational differences in both the front (or back) and the central region. For the consonants [n] and [ç], the durational differences show a similar trend in both regions (front and central): /a/</u/</i/. This suggests that the alveolopalatals are sensitive to the height of the tongue body for the following vowel. While the lower tongue position required for the vowel /a/ reduces the durations in the two regions, the higher elevation of the tongue for the vowel /i/ enhances the position of the tongue blade and dorsum for the two alveolopalatals, and effectively increases its contact period. This is compatible with the distribution of the peak COG values that we discussed before.

### Table 6-13: Effects of the Vowel Contexts on the Front (Back) Region Duration

<table>
<thead>
<tr>
<th>SPEAKER</th>
<th>[n]</th>
<th>[ç]</th>
<th>[r]</th>
<th>[k]</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ONE-WAY ANOVA†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(2,12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-TEST (df=8)‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ci-Cya</td>
<td>10.100**</td>
<td>8.913**</td>
<td>3.588**</td>
<td>10.031**</td>
</tr>
<tr>
<td>Ci-Cyu</td>
<td>2.558*</td>
<td>5.256**</td>
<td>3.310**</td>
<td>0.357</td>
</tr>
<tr>
<td>Cya-Cyu</td>
<td>9.644**</td>
<td>3.738**</td>
<td>0.218</td>
<td>6.366**</td>
</tr>
<tr>
<td>Relation</td>
<td>/i/&gt;/u/&gt;/a/</td>
<td>/i/&gt;/u/&gt;/a/</td>
<td>/i/&gt;/u/&gt;/a/</td>
<td>/i/&gt;/u/&gt;/a/</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPEAKER TM</th>
<th>[n]</th>
<th>[ç]</th>
<th>[r]</th>
<th>[k]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONE-WAY ANOVA†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(2,12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-TEST (df=8)‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ci-Cya</td>
<td>5.131**</td>
<td>13.773**</td>
<td>0.666</td>
<td>1.155</td>
</tr>
<tr>
<td>Ci-Cyu</td>
<td>3.806*</td>
<td>8.231**</td>
<td>0</td>
<td>3.639**</td>
</tr>
<tr>
<td>Cya-Cyu</td>
<td>1.639**</td>
<td>4.037**</td>
<td>0.624</td>
<td>3.687**</td>
</tr>
<tr>
<td>Relation</td>
<td>/i/&gt;/u/&gt;/a/</td>
<td>/i/&gt;/u/&gt;/a/</td>
<td>/i/&gt;/u/&gt;/a/</td>
<td>/i/&gt;/u/&gt;/a/</td>
</tr>
</tbody>
</table>

N.B. The value of a palatalised velar represents the duration of the ‘back’ region.

### Table 6-14: Effects of the Vowel Contexts on the Central Region Duration

<table>
<thead>
<tr>
<th>SPEAKER</th>
<th>[n]</th>
<th>[ç]</th>
<th>[r]</th>
<th>[k]</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ONE-WAY ANOVA†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(2,12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-TEST (df=8)‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ci-Cya</td>
<td>21.392**</td>
<td>7.904**</td>
<td>7.271**</td>
<td>11.013**</td>
</tr>
<tr>
<td>Ci-Cyu</td>
<td>5.875**</td>
<td>3.483**</td>
<td>3.007**</td>
<td>2.066**</td>
</tr>
<tr>
<td>Cya-Cyu</td>
<td>12.736**</td>
<td>4.901**</td>
<td>5.918**</td>
<td>10.751**</td>
</tr>
<tr>
<td>Relation</td>
<td>/i/&gt;/u/&gt;/a/</td>
<td>/i/&gt;/u/&gt;/a/</td>
<td>/i/&gt;/u/&gt;/a/</td>
<td>/i/&gt;/u/&gt;/a/</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPEAKER TM</th>
<th>[n]</th>
<th>[ç]</th>
<th>[r]</th>
<th>[k]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONE-WAY ANOVA†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(2,12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-TEST (df=8)‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ci-Cya</td>
<td>6.025**</td>
<td>12.750**</td>
<td>7.288**</td>
<td>4.218**</td>
</tr>
<tr>
<td>Ci-Cyu</td>
<td>2.803*</td>
<td>2.581*</td>
<td>1.643</td>
<td>1.679</td>
</tr>
<tr>
<td>Cya-Cyu</td>
<td>4.937**</td>
<td>18.812**</td>
<td>6.128**</td>
<td>5.846**</td>
</tr>
<tr>
<td>Relation</td>
<td>/i/&gt;/u/&gt;/a/</td>
<td>/i/&gt;/u/&gt;/a/</td>
<td>/i/&gt;/u/&gt;/a/</td>
<td>/i/&gt;/u/&gt;/a/</td>
</tr>
</tbody>
</table>

†**=p<0.01; *p<0.05; unmarked = no significant.
‡The relation between the particular pairs was assessed by unpaired t-test (one-tailed).
Such a relationship was not obtained for [r’]. The following vowels cause little influence on the duration of the front region. One exceptional instance was found for the vowel /i/ context in speaker MN. If we inspect the trajectories in Figure 6-13, it can be seen that the longer duration of the [ar'i] sequence is particularly due to the releasing activity of the consonant in question. However, the crucial characteristic of the shorter articulatory plateau is retained, since the distinctive flicking, or transient, movement is already finished. The duration in the central region, in contrast, varies greatly across the vowel contexts and the speakers. The most consistent trend is that the duration becomes longer in the /i/ vowel context and shorter in the /a/ vowel context. Such variability in the palatalising gesture is related to the fact that the production of the palatalised tap involves conflicting or competing control of the tongue gestures. Whereas the tip/blade activity is kept constant across the vowel contexts, the speakers may use different co-production strategies for the timing of the palatalising gesture. On the one hand, the trajectories of both speakers show noticeable lengthening of the release phase in the [ar'iui] sequences. On the other hand, the articulatory plateau (in the central region) of some tokens is shorter for speaker TM\(^6\). Consequently, the contact durations of [ar'i] and [ar'iui] for that speaker become indistinguishable. Another possibility is also related to an idiosyncratic property, namely the articulation of [ti]: the two speakers differ in the extent to which the back of the medio-dorsum and the post-dorsum are involved. More involvement of the medio-dorsum in speaker TM assists the palatalising gesture to keep the higher position of the tongue body. These results agree with the widely observed fact that a consonant that is highly resistant to coarticulatory effects is also a strong influencer to the surrounding phonetic segments (e.g. Recasens, 1984; Recasens et al. 1977).

The palatalised velar also reveals variation in the duration of the back and the central region. The raising gesture of the tongue body is already involved in the constriction formation of [k’]. This fact gives us a natural account that the duration in the back region becomes shorter with the anticipation for the following low vowel, but is longer in the high vowel contexts. This view also explains the durational variation in the central region contact. However, because of this fact, idiosyncratic differences in timing control are derived from the coordination between the tongue post-dorsum and medio-dorsum. It must be mentioned here that the ‘fronting’ of the velars commonly refers to the advancement of a velar closure before [i] or [j]. Such a description underestimates the articulation of palatalised (or fronted) velars. They are formed by the ‘fronted’ closure at around the post-palatal region with the secondary constriction of pressing the posterior portion of the medio-dorsum against the relevant region(s). These two gestures are, of course, a single consecutive movement of the tongue body. A similar observation is reported by Houde (1967), in which the velar [g] adjacent to

---

\(^6\) Although there is not much evidence to decide the matter in our experiments, such a shortening of the plateau in the /i/ gesture is a potential coarticulatory effect of [r] in the temporal domain.
the high vowels is decomposed and modelled by the target-directed vertical activity for fronted velar closure and by the non-target-directed horizontal activity for palatal closure. We do not agree with Houde (1967) in that the forward palatal gesture is regarded as (automatic) accommodative coarticulation (see 5.2.1. above). Thus, the palatalised velars, in principle, involve two articulatory gestures, one to make a closure and the other to effect palatalisation. The variability in the region durations (Figures 15-12 and 15-13) is partly due to idiosyncratic strategies for coordination timing between those two successive gestures. We will return to this point later when we discuss the parameter, PERCENT ANTICIPATION.

We have seen how the contact duration of the two particular regions varies in sequence and in isolation. It was found that all the palatal(ised) consonants have their distinctive movement patterns in the front region; and that the sequence and the central region duration vary as a function of the following vowels; the dominant trend is /a/</u/></i/. Recall that the articulatory gestures for /i/ and /j/ in the non-front vowel contexts show the same tendency of shortening, or lengthening, the tongue-palate contact in the central region (see the first experiment). This suggests that both the palatal vocoids and the palatal(ised) consonants are sensitive to the tongue body elevation of the following vowel. We have shown that the duration of the articulatory plateau (hold phase) distinguishes the /i/ gesture from the /j/ gesture. The former involves a longer articulatory plateau that becomes even longer when adjacent to /u/. The latter, /j/, has a shorter articulatory plateau but it becomes longer than that in /aia/ when adjacent to /u/. If we assume that these temporal properties are incorporated into the articulation of the given palatal(ised) consonants, the temporal co-ordination between the two components of the tongue may differ in extent and in timing. It seems reasonable to assume that the timing of the dorsum gesture is controlled in two ways: one is implemented with a stable and stretched articulatory plateau (i.e. Ci syllables) and the other is realised with a brief plateau (i.e. CyV syllables). We will further develop such a view and will examine the temporal interaction between the primary and the (secondary) dorsal gesture.

The second temporal parameter to be discussed is TIME BETWEEN THE PEAKS. It specifies the time between the peak contact in one region and that in the other. The peak point was calculated as the temporal centre of the articulatory plateau of maximum contact. Assuming that the absolute time and relative time are correlated to each other, the former is used to compare differences between the three flanking vowels; and the latter is computed relative to the sequence duration and is used to infer how a linguistic variable (i.e. syllable type, Ci and CyV) affects the peak intervals. This illustrates the timing of the (secondary) palatal gesture. The average peak intervals for each consonant in milliseconds are given in Figure 6-18. The ANOVA results comparing the absolute time for the central contact are summarised in Table 6-15.
**Figure 6-18:** Mean peak intervals and standard deviations for the four palatal (ised) consonants: X-axis = the target consonants; Y-axis = measures of the duration in milliseconds; () = s.d.

**Table 6-15:** Effects of the vowel contexts on peak interval

<table>
<thead>
<tr>
<th>Vowel Context</th>
<th>/a/</th>
<th>/u/</th>
<th>/i/</th>
<th>/u/</th>
<th>/a/</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speaker MN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/p/</td>
<td>(-)</td>
<td>119.817**</td>
<td>10.548**</td>
<td>49.776**</td>
<td>12.032**</td>
</tr>
<tr>
<td>t-Test (DF=8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ci-Cya</td>
<td>16.201**</td>
<td>3.914**</td>
<td>8.775**</td>
<td>4.258**</td>
<td></td>
</tr>
<tr>
<td>Ci-Cyu</td>
<td>10.028**</td>
<td>2.667*</td>
<td>7.345**</td>
<td>2.061*</td>
<td></td>
</tr>
<tr>
<td>Cya-Cyu</td>
<td>3.085**</td>
<td>2.372*</td>
<td>2.612**</td>
<td>3.273**</td>
<td></td>
</tr>
<tr>
<td>Relation</td>
<td>/i/&gt;u&gt;/a/</td>
<td>/i/&gt;u&gt;/a/</td>
<td>/i/&gt;u&gt;/a/</td>
<td>/i/&gt;u&gt;/a/</td>
<td></td>
</tr>
<tr>
<td><strong>Speaker TM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/p/</td>
<td>7.989**</td>
<td>12.739**</td>
<td>26.092**</td>
<td>60.681**</td>
<td></td>
</tr>
<tr>
<td>t-Test (DF=8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ci-Cya</td>
<td>3.124**</td>
<td>5.847**</td>
<td>6.909**</td>
<td>9.636**</td>
<td></td>
</tr>
<tr>
<td>Ci-Cyu</td>
<td>2.776*</td>
<td>2.457*</td>
<td>4.993**</td>
<td>7.076**</td>
<td></td>
</tr>
<tr>
<td>Cya-Cyu</td>
<td>0.566</td>
<td>2.287*</td>
<td>2.973**</td>
<td>2.151*</td>
<td></td>
</tr>
<tr>
<td>Relation</td>
<td>/i/&gt;u&gt;/a/</td>
<td>/i/&gt;u&gt;/a/</td>
<td>/i/&gt;u&gt;/a/</td>
<td>/i/&gt;u&gt;/a/</td>
<td></td>
</tr>
</tbody>
</table>

**Significance Levels:** **p < 0.01; *p < 0.05; unmarked = no significant**

The relation between the particular pairs was assessed by unpaired t-test (one-tailed).

Significant differences were observed for all the consonants and the peak interval increases in the order of /a/ < /u/ < /i/. This hierarchy can be inferred when the articulatory plateau in the central region gets longer, its midpoint, perforce, moves to the right in the articulatory trajectories: the longer the articulatory plateau, the greater the interval. Because the plateau of the front region is considered as more or less constant, the variability in the peak interval appears to come from the contextual shortening and lengthening of the contact
in the central region.

Next, we consider differences in the syllable type (i.e. Ci and CyV). The peak interval relative to the sequence duration was computed. The means of relative peak intervals and the results of the comparisons are shown in Table 6-16 below. Significant difference between the syllable type was observed for all the palatal(ised) consonants. The pair-wise comparison reveals that the relative peak interval of Ci syllables significantly differ from those of Cya and Cyu syllables. There is a temporal lag between the peak of the tip/blade (or post-dorsum) activity and that of the (pre- and) medio-dorsum activity in Ci syllables, that time interval is shortened in CyV syllables. This may become clear, to some extent, if we look at the articulatory trajectories in Figures 6-11 to 6-14. The alveolopalatals in Ci syllables show, on the one hand, that the tongue contact in the central region is gradually involved, during the maximum closure, in preparing for the following /i/, but those in CyV syllables show, on the other hand, that the peak contact in the central region synchronises nearly-completely with the maximum closure by the tongue (tip/blade). Similarly the palatalised velar shows the timing adjustment between the back and central regions, although idiosyncratic differences are negligible. In contrast, such a lengthening or shortening of the peak interval is less clear in the palatalised alveolar tap. In fact, the same pattern of relative peak interval observed for [r^] is essentially attained by the durational nature of the palatalising gesture. This is because the

<p>| Table 6-16: Mean Peak Intervals and Their Differences Between the Syllable Types |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Speaker</th>
<th>Mn</th>
<th>[p] % (s.d)</th>
<th>[ɛ] % (s.d)</th>
<th>[r^] % (s.d)</th>
<th>[k^] % (s.d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>aCi</td>
<td>31.85 (2.29)</td>
<td>20.44 (6.86)</td>
<td>54.65 (3.92)</td>
<td>11.01 (3.12)</td>
<td></td>
</tr>
<tr>
<td>aCya</td>
<td>17.49 (2.03)</td>
<td>10.11 (3.83)</td>
<td>40.87 (14.18)</td>
<td>4.90 (2.54)</td>
<td></td>
</tr>
<tr>
<td>aCyu</td>
<td>16.92 (3.64)</td>
<td>13.13 (3.17)</td>
<td>35.39 (5.08)</td>
<td>7.38 (1.58)</td>
<td></td>
</tr>
<tr>
<td>One-way ANOVA†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(2,12)</td>
<td>47.346**</td>
<td>5.883**</td>
<td>6.091**</td>
<td>7.546**</td>
<td></td>
</tr>
<tr>
<td>t-TEST (df=8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ci-Cya</td>
<td>10.462**</td>
<td>2.937**</td>
<td>2.093*</td>
<td>3.389**</td>
<td></td>
</tr>
<tr>
<td>Ci-Cyu</td>
<td>7.775**</td>
<td>2.159*</td>
<td>6.701**</td>
<td>2.105*</td>
<td></td>
</tr>
<tr>
<td>Cya-Cyu</td>
<td>0.305</td>
<td>1.359</td>
<td>0.813</td>
<td>1.849</td>
<td></td>
</tr>
<tr>
<td>Relation</td>
<td>Ci &gt; Cya =Cyu</td>
<td>Ci &gt; Cya =Cyu</td>
<td>Ci &gt; Cya =Cyu</td>
<td>Ci &gt; Cya =Cyu</td>
<td></td>
</tr>
<tr>
<td>Speaker TM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aCi</td>
<td>22.42 (7.24)</td>
<td>19.29 (6.55)</td>
<td>56.25 (3.59)</td>
<td>19.66 (1.36)</td>
<td></td>
</tr>
<tr>
<td>aCya</td>
<td>13.38 (2.54)</td>
<td>5.92 (2.49)</td>
<td>36.09 (18.48)</td>
<td>9.41 (2.09)</td>
<td></td>
</tr>
<tr>
<td>aCyu</td>
<td>6.49 (3.37)</td>
<td>10.63 (6.21)</td>
<td>40.34 (7.77)</td>
<td>9.69 (1.47)</td>
<td></td>
</tr>
<tr>
<td>One-way ANOVA†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(2,12)</td>
<td>6.876*</td>
<td>7.861**</td>
<td>4.085*</td>
<td>60.681**</td>
<td></td>
</tr>
<tr>
<td>t-TEST (df=8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ci-Cya</td>
<td>2.634*</td>
<td>4.262**</td>
<td>2.395*</td>
<td>9.161**</td>
<td></td>
</tr>
<tr>
<td>Ci-Cyu</td>
<td>2.930**</td>
<td>2.144*</td>
<td>4.152**</td>
<td>4.172**</td>
<td></td>
</tr>
<tr>
<td>Cya-Cyu</td>
<td>0.753</td>
<td>1.573</td>
<td>0.474</td>
<td>0.240</td>
<td></td>
</tr>
<tr>
<td>Relation</td>
<td>Ci &gt; Cya =Cyu</td>
<td>Ci &gt; Cya =Cyu</td>
<td>Ci &gt; Cya =Cyu</td>
<td>Ci &gt; Cya =Cyu</td>
<td></td>
</tr>
</tbody>
</table>

††**=*p<0.01; *p<0.05; unmarked = no significant
††The relation between the particular pairs was assessed by unpaired t-test (one-tailed).
shorter peak interval is not comparable with the synchronisation between the two components of the tongue: the tip/blade gesture is strongly resistant to the palatalising effects by the dorsum. These results, on the whole, suggest that the speakers make a meaningful distinction in timing control of the (secondary) dorsal raising gesture. More direct evidence supporting this view is found in the result of the third and fourth parameters, **SEQUENCE OVERLAP** and **PERCENT ANTICIPATION**.

For the discussion of those two parameters, we first summarise the measurements of the two parameters and then discuss consonant-specific characteristics on the basis of the statistical assessments. **SEQUENCE OVERLAP**, as explained before, is the percentage of the total sequence duration during which contact occurred in both regions. The value was computed relative to the total duration that was measured above 50% of linguopalatal contact. This measure specifies the degree of co-production in the front and central regions for [n, c, r] and in the central and back regions for [k']. For the latter, the area where the two regions overlap is actually measured for the duration of the central region above the 50% cut-off point. Figure 6-19 presents the average total and overlap durations in milliseconds.

The mean values of the sequence overlap are given in Figure 6-20. It can be seen that there are consonant-specific characteristics of co-production. The alveolopalatals reveal that the value of sequence overlap increases in the order of /i/</u/</a/. The larger the value of sequence overlap becomes, the more the two components are co-produced. The palatalised tap involves no overlapping gestures. Absolute time in milliseconds in Figure 6-19 represents only the front region activity above the 50% cut-off point. In contrast, the two regions overlap in the palatalised velar and the value of each context is very close.

The other parameter, **PERCENT ANTICIPATION**, demonstrates the extent to which the articulatory gesture for the (secondary) dorsal articulation has been anticipated at the left-edge (the first frame of MAX or the plateau of the front (or back) region). The percent anticipation of the tongue dorsum is a value relative to its maximum percentage contact in the temporal centre of the articulatory plateau. The mean values of percent anticipation for [n, c, r'] are given in Figure 6-21, in which the mean percentage values of tongue-palate contact at the left-edge and at the temporal mid-point of maximum constriction are also included. The anticipation value of [k'] was measured at two different temporal points: at the left-edge and the temporal mid-point of the articulatory plateau of the back region. Because of the articulatory nature of the consonant, as we will see, the usual measurement at the left-edge was unworkable. Yet, this value does indicate certain aspects of idiosyncratic strategies used to effect palatalisation. Figure 6-22 shows those values at the two temporal points. One common trend is that the anticipation value increases in the order of /i/</u/</a/.
Figure 6-19: Mean Total and Overlap Durations Above the 50% Cut-Off Point
X-axis = the target consonants; Y-axis = measures of the duration in milliseconds

(a) Speaker MN

(b) Speaker TM

Figure 6-20: Mean Sequence Overlap Above the Point of 50% Lingualpalatal Contact;
X-axis = the target consonants; Y-axis = measures of the overlap (%); standard deviations for speaker MN; /an'i/ (6.03), /anja/ (2.14), /anju/ (2.14), /asi/ (2.49), /asja/ (2.88), /asju/ (5.22); those for speaker TM; /an'i/ (2.17), /anja/ (1.84), /anju/ (3.09), /asi/ (3.43), /asja/ (12.05), /asju/ (5.25)
FIGURE 6-21: MEAN PERCENT-ANTICIPATION OF THE CENTRAL REGION FOR [p, c, ɛ]; only the anticipation value is shown; (a) speaker MN; (b) speaker TM

FIGURE 6-22: MEAN PERCENT-ANTICIPATION OF THE CENTRAL REGION FOR [k’]; (a) the values at the left-edge; (b) those at the temporal mid-point of the MAX plateau of the front region; only the anticipation value is shown
The results of an ANOVA and a pair-wise comparison are shown in Table 6-17 for the sequence overlap and Table 6-18 for the percent anticipation. It is evident that the pattern of significant differences generally indicates the distinction between the syllable types. The higher values in both sequence overlap and percent anticipation suggest that the two components of the tongue are more blended in CyV syllables than in Ci syllables.

**Table 6-17: Differences in Sequence Overlap Between the Ci and CyV Syllables**

<table>
<thead>
<tr>
<th>Speaker MN</th>
<th>F(2,12)</th>
<th>[p]</th>
<th>[c]</th>
<th>[k]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONE-WAY ANOVA</td>
<td>37.031, p&lt;0.0001</td>
<td>2.410, p=0.131</td>
<td>3.572, p=0.06</td>
<td></td>
</tr>
<tr>
<td>t-TEST (df=8)†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ci-Cya</td>
<td>7.275, p&lt;0.0001</td>
<td>2.805, p=0.111</td>
<td>1.695, p=0.064</td>
<td></td>
</tr>
<tr>
<td>Ci-Cyu</td>
<td>4.896, p&lt;0.0001</td>
<td>0.462, p=0.096</td>
<td>0.879, p=0.202</td>
<td></td>
</tr>
<tr>
<td>Cya-Cyu</td>
<td>4.854, p&lt;0.0001</td>
<td>1.623, p=0.071</td>
<td>2.682, p=0.012</td>
<td></td>
</tr>
<tr>
<td>Relation‡</td>
<td>Ci &lt; Cyu &lt; Cya</td>
<td>Ci = Cyu &lt; Cya</td>
<td>Ci = Cyu &lt; Cya</td>
<td></td>
</tr>
</tbody>
</table>

**Table 6-18: Differences in Percent-Anticipation Between the Ci and CyV Syllables**

<table>
<thead>
<tr>
<th>Speaker MN</th>
<th>F(2,12)</th>
<th>[n]</th>
<th>[c]</th>
<th>[k]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONE-WAY ANOVA</td>
<td>155.461, p&lt;0.0001</td>
<td>9.601, p=0.003</td>
<td>1.633, p=0.235</td>
<td></td>
</tr>
<tr>
<td>t-TEST (df=8)†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ci-Cya</td>
<td>21.057, p&lt;0.0001</td>
<td>3.080, p=0.007</td>
<td>3.268, p=0.005</td>
<td></td>
</tr>
<tr>
<td>Ci-Cyu</td>
<td>9.763, p&lt;0.0001</td>
<td>1.003, p=0.172</td>
<td>3.129, p=0.007</td>
<td></td>
</tr>
<tr>
<td>Cya-Cyu</td>
<td>6.419, p&lt;0.0001</td>
<td>3.414, p=0.004</td>
<td>0.301, p=0.385</td>
<td></td>
</tr>
<tr>
<td>Relation‡</td>
<td>Ci &lt; Cyu &lt; Cya</td>
<td>Ci = Cyu &lt; Cya</td>
<td>Ci = Cyu = Cya</td>
<td></td>
</tr>
</tbody>
</table>

†unpaired and one-tailed; ‡for the relation between the particular pairs the symbol = indicates no significance; the symbol < indicates significance at the p<0.01 level (one-tailed).
Such a general pattern, however, was not obtained in the results of SEQUENCE OVERLAP for the consonant [ç], as one of the CyV syllables, /asja/, is isolated from the others. The articulatory overlap becomes similar between /asi/ [açi] and /asju/ [açu]. Accordingly, there is no significant difference in SEQUENCE OVERLAP for all the contexts. We can argue that this tendency is due to the inherent articulatory property of [ç], the strong coupling effect between the tip/blade and dorsum component of the tongue. Since the alveolopalatal [ç] that we characterised as an alveolarised palatal, has a more retracted place of articulation than [p] (see Figures 6-6 and 6-7), it necessarily involves the raising of the tongue dorsum in the articulation. Also, the coarticulatory effects of the vowel /u/ lengthen the tongue-palate contact of the palatalising gesture (see Figures 6-12 and 6-17) and enhance coupling effects to cause lingual gestures for /si/ and /sju/ to become similar.

The consonant [ɾ] reveals an uncommon characteristic in SEQUENCE OVERLAP and PERCENT ANTICIPATION. Recall our earlier finding that, while the dorsum gesture is tightly constrained, timing of the tip/blade gesture is typically constant when compared to the other palatal(ised) consonants. Such manner requirement suppresses the overlap between the two gestures. Consequently the tip/blade and the dorsum gestures are activated in this order. Thus, the lack of interarticulator overlap is due to the production constraint specific to the consonant. This is also related to the parameter PERCENT ANTICIPATION, the result of which turns out to be non-significant between Ci and CyV syllables. It is worth noting that the percentage contact in the central region is not large either at the left-edge or at the MAX point: in the pooled syllable context, 24.45% and 35.84% (speaker MN); 24.16% and 34.45% (speaker TM) respectively. These values are roughly half as small as those, for instance, in the /aɾja/ tokens: 51.67% and 68.5% (MN); 44.14 % and 60.28% (TM). Most significant is the reduction in /arja/, which paradoxically results in the highest value of PERCENT ANTICIPATION: 24.18% at the left-edge and 30.02% at the MAX point (MN); 22.50% and 24.16% (TM) respectively. This suggests that the general distinction between Ci and CyV syllables depends on the extent of the vocalic gesture that is executed after the consonantal gesture by the tip/blade ceases (see Figures 6-16 and 6-17).

Similar patterning is observed in the results of the palatalised velar [kʰ]. The ANOVA results to examine the effect of the syllable type show no significant differences in SEQUENCE OVERLAP. The pair-wise comparisons indicate the distinction between Ci and CyV syllables for speaker TM but not for speaker MN. As for PERCENT ANTICIPATION, significant effects were found when it is measured at the temporal mid-point of the articulatory plateau in the front region, but not at the left-edge. We should notice that the contact degree in the central region at the left-edge is, surprisingly, smaller for speaker TM than for speaker MN (see Figure 6-22(a)): the mean values across the vowel contexts are 34.44% (s.d. 5.15) for speaker MN and 7.52% (s.d. 2.23) for speaker TM. These results essentially come from different
co-production strategies used by a speaker to make a constriction further forward with the tongue medio-dorsum raised. Typical contact patterns are given below.

![Typical EPG Patterns of the [a] Syllable](image)

(a) Speaker MA and (b) Speaker MB.

The symbols 'C' and 'R' indicate the first frame of complete closure and that of closure release.
Differences between the two speakers are most evident in the approach phase to the complete closure that is labelled as 'c' in the above Figure. It is obvious that, for MN, the forward extension of contact gradually increases and the complete closure comes after it. This ordering was observed for the \([k^l]\) in all the vowel contexts and the contact degree effectively becomes indistinguishable. For TM, in contrast, the forward contact extends after the attainment of complete closure. This results in few values of the percent anticipation at the left-edge.

6.5 Summary

6.5.1 Two Palatalising Gestures \([i] \) and \([j] \)

A high front vowel \(/i/\) and a palatal approximant \(/j/\) are distinctive gestures by themselves and also function as a palatalising gesture for the given consonantal articulation. Browman & Goldstein (1989) propose three parameters to specify articulatory movement in the vocal tract: constriction location (CL); constriction degree (CD); and stiffness. They hypothesise that \(/j/\) is characterised as having the same CL and CD as \(/i/\), but differs in stiffness: \(/i/\) is specified as a tongue body [narrow palatal] gesture, while \(/j/\) is specified as a tongue body [narrow palatal] gesture with an [increased] value of stiffness. The ‘stiffness’ is a control variable of a gesture that ‘specifies (roughly) the time required to get to target (Browman & Goldstein, 1989; 208)\(^7\)’. Therefore, the ‘increased’ value of stiffness in the \(/j/\) gesture phonetically results in faster movement towards the target than the \(/i/\) gesture. On the basis of the present results, we develop our discussion by evaluating the above articulatory specifications of the two vocoids.

We have shown that differences between the articulatory gesture for \(/i/\) and that for \(/j/\) are paralleled by differences in the degree of coarticulatory resistance and the duration of the hold phase (Figures 6-4-1 and 6-4-2 and Table 6-5). For the latter, the contextual vowels, particularly \(/u/\), affect the duration of the hold and release phase, but not the approach phase: the two phases become longer but the durational hierarchy \(/j/</i/\) is constantly retained. The main constriction location (i.e. rows 5-8), and the hold phase duration specific to each vocoid (i.e. \(/i/>/j/\)), are systematically maintained with the changes of vowel contexts. There is a broad distinction for the oral constriction gestures vocalic and consonantal (Browman & Goldstein, 1989, 1990). When compared to \(/i/\), a gesture for \(/j/\) is characterised as the consonantal gesture that has a shorter time constant than the vocalic gesture.

Contrasting \(/j/\) with \(/i/\) in the [increased] stiffness value by Browman & Goldstein seems to be workable. However, it must be noted that, when the three articulatory phases are considered, the shorter duration of the hold phase accurately distinguishes \(/j/\) from \(/i/\): the

---

\(^7\) Browman & Goldstein (1989; 229) suggest that there may be a relationship between the (abstract) stiffness used for a gestural specification and the actual biomechanical stiffness of the muscle tissues.
duration of the other two phases can be considered as relatively invariant across the two palatal gestures. This fact of 'relative invariance' (Fujimura, 1986) is obscured by the definition of the stiffness above. Given that articulatory 'target' corresponds to the midpoint of the articulatory plateau (or the hold phase), the specification of the [increased] stiffness predicts that 'the time required to get to target' in the actual implementation is shorter in /j/. However, this is because the articulatory plateau is the only parameter which is constantly shorter in the /j/ gesture regardless of the surrounding vowels; not because the transition to (and from) the maximum constriction increases its speed. Note in passing that the coarticulatory effects appear in the release phase only. This may be related to the fact that the CV syllable structure, rather than (C)VC, is a potential segmentation unit in Japanese.

According to Browman & Goldstein, the two vocoids are assumed to have the same constriction degree, namely [narrow], or approximant stricture, at the palatal region. This is generally supported by the results of the present experiment. Yet, we found that the actual implementation of the CD [narrow], which is depicted by the amount of tongue-palate contact, varies across the vowel contexts and individual speakers: while the pattern and degree of tongue contact were relatively stable in /i/, the forward extension of tongue-palate contact was significant in the /uju/ sequence. Nevertheless, the essential stricture type (or the lingual contact pattern) of the approximant can be attained for the two distinct palatal gestures.

Whereas there are inter-speaker differences in the configurational patterns of tongue-palate contact at the MAX point, they are largely similar to each other, when the two speakers are considered independently. This, however, is not entirely supported by statistical examinations: one speaker shows significant differences in the pairs /aia, aja/ and /oio, ojo/ respectively, while the other shows no significant differences between the two vocoids across the vowel contexts. This raises a question as to the relationship between the duration of articulatory movement and the magnitude of articulatory gesture. Our results imply that the shorter articulatory duration is not necessarily implemented as the smaller contact degree. It is difficult to assert from the present results whether time constants different between /i/ and /j/ correlate with different degrees of tongue-palate contact. Instead, the contrastive articulatory durations between the two vocoids appear to be related to their susceptibility to the coarticulatory effects from the surrounding phonetic segments.

6.5.2 Palatal(ised) Consonants and Palatalisation

At the beginning of this chapter, we introduced an exploratory definition of palatalisation: a specialised use of the raising gesture of the tongue body. It was hypothesised that there are two kinds of (secondary) dorsum gesture during the production of [n, c, r̃, k̃], the pattern that corresponds to the two basic syllable types, CI and CyV, in Japanese. The EPG evidence reveals that the gestures are precisely timed with respect to the syllable type: the contrastive
timing and magnitude of dorsum activities underlie the distinction between the consonantal articulations in the two syllables. This suggests that the secondary articulation is consciously controlled. Substantial divergences were also found for intergestural coordination and timing between the primary articulatory gesture and the secondary dorsal gesture during the palatal(ised) consonants.

We may summarise general spatial and temporal properties of [p, c, k] in each syllable, with reference to the present results: (i) similarities that are shared by the palatalised consonants across the two syllable types are the centre of articulatory gravity, which was computed as a function of the overall electrode activation and the peak COG values which were distributed tightly over rows 3 to 5 (see the diagram in (6)); (ii) the percentage contact in the central region shows no significant differences between /i, j/ and the palatalised consonants; they are all complex segments (see 6.2.2); (iii) the tongue-palate contact in two relevant regions lasts longer in Ci syllables (REGION and SEQUENCE DURATION); (iv) the tongue tip/blade (or post-dorsum) and the (pre/-)medio-dorsum are more co-produced in the CyV syllables (PEAK INTERVAL and SEQUENCE OVERLAP); (v) the activity of the tongue medio-dorsum is more anticipated in CyV syllables (PERCENT ANTICIPATION); and (vi) the pattern of significant differences indicates the distinction between the effects of an [i]-gesture for Ci syllables and those of a [j]-gesture for CyV syllables. These coarticulatory effects can be explained by the temporal mechanism involved in the dorsum raising gesture during the production of the palatal(ised) consonants.

The characteristic of the constriction place observed in the linguopalatal contact pattern for the consonant [n] is shown to be language-specific. We mentioned in chapter 3 that the place-neutral status of [n] (i.e. a dorso-palatal nasal) had been questioned (e.g. Catford, 1977; Recasens, 1990): there is a fairly general tendency for the place to shift forwards. As discussed in chapter 4, the complete occlusion of the Japanese [n] is formed at rows 1-4, while that of the Catalan [ɲ] formed at rows 2-5. We suggested that the frequent involvement of the tip contact was primarily due to the frontal constriction formation in Japanese. Such variability in constriction place between the languages supports the concept of ‘articulatory settings’ (Honikman, 1964), as well as ‘the principles of susceptibility and compatibility’ (Laver, 1980, 1994). Given these concepts, the forward constriction of the Japanese [n] could be understood as the effect of the tongue settings: a preference for dentalised and palatalised laminal articulation (Edwards, 1903). This is also supported by the results of the COG analysis. Furthermore, if this idea conforms with the coarticulatory strategies used by the speaker, then the issue of the articulatory controls for [n] becomes more complicated. Since the tongue blade is arguably anchored at the anterior region in Japanese, it is rather implausible to assume that in the production of the consonant, the participation of the tip and post-dorsum gestures are mechanical responses to the primary activity executed by the tongue
predorsum, the claim made by Recasens (1990). Although there is not much evidence to
decide the matter, it can safely be stated that the articulatory setting is a valid concept in
speech analysis, not only because it provides a parametric view for describing language-
specific articulatory characteristics, but also because it offers a substantial base for
quantifying language-specific coarticulatory characteristics.

Interesting findings are the behaviour of [ç] and [r\^]. The alveolopalatal [ç] in Ci and
Cyu syllables tends to be indistinguishable in the parameter SEQUENCE OVERLAP. This
implies that the distinction between these two syllables largely depends on the active
involvement of a labial gesture for /sju/, the effect of which makes fricative quality darker (or
less bright). This activity, though strangely ignored in common phonetic descriptions, can
easily be confirmed by observation. The finding has an important consequence for the issue of
vowel devoicing in Japanese. Recall that there is a sound change in progress, what we call
CyV variability (see Chapter 2, section 2.3.3): for instance, \[çu,kuadai\]—\[çi,kudai\]
‘homework’, \[çu,utsui\]—\[çi,utsui\] ‘expense’, \[(d)çu,kuiren\]—\[(d)zi,kuiren\] ‘expert’,
\[(d)çun'oo\]—\[(d)zin'oo\] ‘lecture’. In the next chapter we will expand the discussion
regarding the articulatory and coarticulatory base underlying this phenomenon, relating it to
palatalisation and vowel devoicing.

The consonant [r\^] indicates the above general properties in both Ci and CyV
syllables. However, one of the EPG parameters, SEQUENCE OVERLAP (above the 50% cut-off
point), was not obtained. It appears that the dorsum raising gesture to effect palatalisation is
incompatible with the single ballistic movement of the tongue tip/blade for [r]. The important
point is that during the production of a tap, either palatalised or non-palatalised, the tongue
dorsum is actively and consciously controlled even though the degree of tongue-palate
contact is smaller. Although the momentary gesture of the tongue tip/blade is commonly
considered as the distinctive property of [r], it is reasonable to suppose that the gesture may
not be realised without the conscious and tight control of the tongue dorsum. This idea will be
elaborated in the next chapter with the example of child phonological acquisition.

The above production constraint on [r] yields a consonant-specific co-production
strategy. There are three types of coordination between the two components of the tongue:
gestural blending for the alveolopalatals; fronting for the palatalised velar; and gestural
sequencing for the palatalised tap. Since the palatalisation assigns competing or conflicting
demands to the tongue body, those three types of intergestural coordination are considered as
the ways of resolving antagonistic gestures, the choice of which depends on the place and
manner of articulation. In addition, the characteristic of [r\^] suggests a refinement of a
hypothesis that the degree of the dorsum involvement in the consonantal articulation and that
of the coarticulatory effects from the vowels are correlated inversely and monotonically (for
Catalan, Recasens, 1984; Recasens et al. 1997): the more the tongue-dorsum contact increases,
the smaller the V-to-C coarticulation. This is because the dorsum gesture for the dorsal consonant strongly resists, or readily incorporates, that for the surrounding vowels (the following vowels in our experiments). However, the consonant [r] is an apparent counter-example for the monotonic-inverse hypothesis. The manner of articulation, as Recasens et al. (1997) acknowledge, must be taken into account of coarticulatory resistance.

6.6 Conclusion
The primary question we addressed in this chapter was how the (secondary) dorsal articulation is spatio-temporally coordinated with the primary articulation during the production of the four palatal(ised) consonants, [n], [ç], [r], and [k]. We formulated the working hypothesis that palatalisation could be understood as a specialised use of the raising gesture of the tongue dorsum. The results of the EPG experiments, as expected, indicated that: (i) the (secondary) dorsal gesture was classified into two groups, an [i]-like gesture and a [j]-like gesture; (ii) these two gestures differed significantly in the duration of the articulatory plateau, not the transition to and from it; (iii) the spatial configurations, similar between Ci and CyV syllables, were derived from the different temporal coordination between the two components of the tongue relevant to the given consonantal articulation, blending, fronting and sequencing; (iv) two kinds of palatalisation can be accounted for by assuming two timing mechanisms specific to the syllable types. We argued that the concept of articulatory settings raised an important issue, that of the speakers' control strategies used to effect palatalisation.

The generalisations derived from the analyses of the various spatio-temporal parameters lead us to a review of the speech production model described at the beginning of this chapter. The articulatory plan that is represented by the features does not specify any of the timing distinctions observed for Ci and CyV syllables. Fowler (1977) and Fowler et al. (1980) call this model 'translation theory (or extrinsic timing model)'. The characteristic features of translation theory, as its name implies, are translations from the abstract representation to the physical realisation. In the course of the derivation the representation 'increases' the pronounceability (e.g. Keating, 1990). If timing information is to be written in the articulatory plan, a further (timing) component for the translation process must be devised. The translation-theoretic view of speech is not the only one to explain the production of speech. As discussed in the previous chapters, there is a tight association between the linguistic and articulatory aspects of speech. What we need is a technique to describe the events in the vocal tract more directly by referring to systematic movements performed by a speaker. In the next chapter we shall explore this theory by presenting analyses of some phonetic and phonological phenomena, based on our findings of the EPG experiments.

---

8) See Fowler et al. (1980) and Fowler (1980) for a detailed criticism; See also Nolan (1982) for a criticism of Fowler's view (Action Theory).
Chapter 7

Coarticulation and Some Phonological Processes

7.1 Introduction

In this chapter we argue that it is possible to give an improved explanation for phonological phenomena by taking into account the systematic articulatory controls underlying them. This involves integrating a parametric approach to phonological description. We approach our theme through a detailed analysis of the three phonological phenomena in Japanese: /r/ variability in adult and child phonology (section 7.2); CyV variability and de-palatalisation (section 7.3); and vowel devoicing (section 7.4).

The proposal formulated in this section is based on the idea that coarticulation is realised as overlapping and blending of articulatory gestures. This idea originates from the pioneering work of Öhman (1965), in which he quantitatively analysed the VCV sequences produced by American English, Swedish and Russian. Recently, Saltzman & Munhall (1989) and Browman & Goldstein (1995, 1992, 1990a) have developed the idea within the framework of task dynamics. Task dynamics is a general model of skilled movements and describes them in terms of the tasks to be accomplished (Hawkins, 1990). Invariant gestural units are specified by a set of articulators and their dimensions: constriction location (CL), constriction degree (CD) and stiffness (duration of a gesture). These are associated with the variables in the vocal tract and attempt to shape its movements. The gestures for a given word, or utterance, are organised together into larger coordinative structures, which are represented in a gestural score. The theoretical basis of Articulatory Phonology proposed by Browman & Goldstein is that articulatory gestures are the units of phonological representation as well as the units of action in the vocal tract. Browman & Goldstein (1991, 1990a) use this model to describe various synchronic and diachronic phonetic processes such as assimilation and deletion, showing that these seemingly unrelated phenomena can be explained by a single mechanism, the modification of the magnitude of gestural overlap.

Based on the results of our EPG experiments, we discuss the patterning of articulatory gestures underlying the three different phenomena in Japanese: child phonology (i.e. /r/ variability); sound change in progress (i.e. CyV variability and de-palatalisation); and a phonological rule (i.e. vowel devoicing). Assuming the gestural model for describing changes in articulatory gestures and their timing, we show that the model reasonably represents the gradient and developmental nature of these phenomena. We propose some amendments to the dynamic parameters. This chapter can be considered as an attempt to bring together cognitive aspects and motor behavioural aspects of speech production.
7.2 /r/ Variability in Adult and Child Phonology
This section deals with the variability of /r/ ([r]) that is observed in child phonological acquisition, regional accents, and the careless pronunciation of adults. Based on the results of our EPG experiments in chapter 4, we consider the spatiotemporal characteristics of the Japanese [r], relating those lingual activities to factors of the articulatory control mechanisms. Given a parametric, or dynamic, perspective of speech, an explanation is proposed based on the rules in the vocal tract. We will show that the proposal can attain a unified account for the /r/ variability that is observed in the pronunciation of children and adults.

7.2.1 The Problem of Rule-based Generalisation
The data of children’s pronunciation is given in (1-5). The /r/ variability falls into five groups in terms of the type of exchanged segment. All the examples are taken from Utsunomiya (1998), in which phonological acquisition of three-year-old children was investigated.

(2) [ko,te]→[ko,de] ‘this’, [bo,oru]→[bo,odu] ‘ball’, [m’idori]→[m’idodi] ‘green’, [t’erebi]→[t’edebi] ‘T.V.’, [’toire]→[’toide] ‘toilet’, [ki,i,ro]→[ki,i,do] ‘yellow’
(3) [ko,te]→[ko,je] ‘this’, [o,renz]i→[o,jenzi] ‘orange’, [’meron]→[’mejon] ‘melon’, [so,te]→[so,je] ‘that’, [k’i,irei]→[k’i,ije]2 ‘beautiful’, [i,rete]→[i,jete] ‘let me join in’
(4) [’dareka]→[’dakeka]3 ‘who’, [m’idori]→[m’idoki] ‘green’, [o,ri,rus]→[o,ri,dku] ‘fall’, [t’uru,ippu]→[t’uru,ippu] ‘tulip’, [ja,ri,takattara]→[ja,ji,takattara] ‘if you like’

The phonetic regularities of the above segmental processes are summarised as rewrite rules in (6a-e) below.

1) Utsunomiya’s informants are two girls: both were aged three years and four months old and lived in London at the time of data collection. The data was collected during one month (one session per week) and the utterances were tape-recorded. The phonetic transcription was carefully made, using the IPA, on the basis of actual pronunciations and tape recordings. Fortunately I had a chance to listen to the tape recordings and studied children’s production of the utterances that contain various realisations of [r]. Thus, the phonetic transcription of the examples in (1-5) were first normalised by Utsunomiya (1998) and then by me. I gratefully acknowledge helpful discussions with Tomomi Utsunomiya.

The total number of words collected by Utsunomiya (1998) was 672, 250 words of which contain the consonant [r]. The incorrect production of [r] is found in 115 out of those 250 words. The examples in (1-5) above are randomly chosen from those incorrect productions.

2) In standard Japanese the diphthong [ei] and the long vowel [ee] are free variations as in [k’irei→k’iree]. From the tape recordings a strong preference for [ei] was observed for the two informants.

3) This symbol is used to imply a very brief [j]-like transition. The phonetic forms in (4) differ from those in (3) in the duration characteristic of the glide.
The replacement by a voiced alveolar stop [d] in (6a,b) occurs most frequently and extensively. In contrast, the frequency of the replacements by [j] or [l] in (6c,d) and that of deletion in (6e) depend on idiosyncratic preference and the developmental stages of the [r] articulation. Whereas the alternation between [r] and [d] was persistently observed throughout the period of the data collection, the other patterns gradually decreased (Utsunomiya, 2000, pc). These observations agree with the results reported in previous studies (e.g. Nakazima, 1972), on the point that the /r/ variability, [r]→[d] in particular, is commonly observed in the pronunciation of Japanese children.

In an SPE framework the rewrite rules in (6a-e) are considered as ‘realisation rules’ (Smith, 1973, 13) that ‘take the adult surface form as input and give the child’s form (prior to the application of phonetic rules) as output’. These mapping rules are unlearned in the course of development. Similarly Stampe (1973) calls them natural rules that are suppressed in the course of phonological acquisition. Thus, the presence of realisation rules guarantees the incorrect outputs of [r] described in (1-5) above. This approach predicts that various patterns of substitution do not occur after children produce the correct [r] regularly.

However, there are adult pronunciations that contradict the prediction above. The phonetic forms in (7) are found in the regional accent of the Fukuoka, Yamaguchi, Wakayama, Mie, and Niigata area (examples are from Fujiwara (1997): phonetic transcriptions are mine). And the forms in (8) are easily observed in everyday conversations between adults (examples are based on my informal observation).


The alternation occurs between [d] and [r]. This is essentially the same as the pattern found in child pronunciation (1, 2). Notice here that only [d] and [r] are involved in the process: the

4) It must be noted that there are considerable idiosyncratic differences. One informant (A) prefers the replacement types in (6a,b), the frequency of which is 70%. The other informant (B) shows this tendency less clearly and uses the other types of substitution in almost equal frequency. For instance, the word midori ‘green’ is pronounced as [ˈmɪdɔɾi] by informant A, but [ˈmɪdɔdɪɾi] by informant B; [ko,re]→[ko,de] and [so,re]→[so,je] (A) but [ko,re]→[ko,de], [ko,e], or [ku,je] and [so,re]→[so,de] or [so,je] (B).
child's patterns in (6c-e) are not found in adult speech. These examples lead us to a review of the nature of the realisation rules that are assumed in the process of phonological acquisition.

There are two problems in the rule-based account of child phonology. First, it has been questioned whether the realisation rules are psychologically real. The ‘realisation’ rules are different from the ‘phonological’ rules. On the one hand, the phonological rules are assumed to reflect the speaker’s knowledge of the sound structures of a given language, and to represent the language-specific characteristics of a given language. The ‘realisation’ rules, on the other hand, lack these two properties. Because the rules are ultimately unlearned in the developmental process, they simply represent the patterns of incorrect pronunciations: the rules can be considered as a kind of phonetic description. Second, it is not at all clear what is acquired after throwing away the realisation rules that are tentatively formulated by a child. It does not make sense to assume that phonological contrasts of a given language are established. This is because there is no direct evidence to show that a child does not build up the system of phonological contrasts during the period when the realisation rules are considered to be active. In addition, if we assume that the object of phonological acquisition is the system of phonological contrasts, it is likely that the (tentative) formulation of realisation rules is nothing more than an obstacle. The rules that generate the incorrect phonetic forms actually neutralise phonological contrasts in the standard pronunciation. Therefore, the realisation rules are justified only if the system of phonological contrasts already exists.

It is important to distinguish the (psycho)linguistic aspects of phonological development on the one hand, from the physical, or motor control, aspects on the other. The former deals with the linguistic contrasts in child phonology, an issue related to the communication level. The latter, in contrast, concerns the development of the motor representation and control, a question related to the signal generation level (McNeilage & Davis, 1990). It appears that the function of the realisation rules is more appropriately captured when they are considered as the segmental descriptions of the developmental aspects of motor control.

7.2.2 A Gesture-based Generalisation of /r/ Variability in Adult and Child Phonology

We shall attempt to formulate the developmental process of the motor representation by focusing on the systematic articulatory control underlying /r/ variability in child language. Based on the findings of our EPG experiments in chapter 4, the articulatory gestures of [d] and [r] are parameterised by three points: constriction degree (CD), constriction location (CL), and stiffness.

(9a) [d]: tongue tip [CD=close; CL = dentalveolar]

(9b) [r]: tongue tip [CD=close; CL = alveolar; increased stiffness]; tongue body [suppressed]
Both [d] and [r] are stop consonants, the complete occlusion of which is formed by pressing the tip/blade component against the (dent)alveolar region. They distinctively differ from each other in the aspectual process, non-tapped or tapped. Browman & Goldstein (1992, 170) propose the increased value of [stiffness] for [r]. Underlying this proposal is the assumption that a tap/flap is a 'reduced' form of the corresponding stop: a closing gesture of the tongue tip, articulatory duration, and displacement are reduced. Because the [increased stiffness] is specified relative to [t] and [d], this is clearly based on an analogy of the 'flapping' rule in American English: /wɔ:tər/ → [wɔːtə] or [wɔːtə]; /raɪdər/ → [raɪdə]. Our EPG data generally supports Browman & Goldstein's proposal: both closure and gestural durations are shorter in [r] than in [t, d]. However, it is questionable, in the case of the Japanese [r], whether the articulatory gesture for taps/flaps contrasts with that for stops only in the shorter duration of movement. As discussed in chapter 4, the activity of the tongue dorsum is highly constrained in tap/flap articulation. It is reasonable to suppose that the suppression of the raising gesture of the dorsum is prerequisite for the successful flicking, or transient, gesture of the tongue tip/blade. Therefore, we propose the feature [suppressed] for the tongue body gesture of [r] and this specification co-occurs with the feature [increased stiffness] for the tongue tip/blade. Thus, a tapped stop is distinguished from a non-tapped stop not only by the rapid movement of the tip/blade but also by the constrained movement of the tongue body.

The diagrams in (10a–e) demonstrate the hypothesised gestural scores for various realisations of the consonant /r/ in child's pronunciation. The basic score for a stop [d] is in (10a) and that for a tap/flap [r] is in (10b): the feature [+stiffness] is used to mean the increased value of the feature [stiffness]. The tongue body gestures for contextual vowels (V1 and V2) are specified as [a raised] corresponding to the height of the given vowels. The glottal gesture involves the feature [opening-closing] to indicate the abduction and adduction of the laryngeal muscles for the continuous voicing. It is assumed that these articulatory gestures together form a larger coordinative structure and associate with the prosodic category, syllable, at the higher level. An example is given for all five phonetic forms observed in the /r/ variability for each of the gestural scores.

The gestural scores in (10a,b) illustrate the most frequent pattern of the alternation between [r] and [d]. This pattern is regarded as the complete substitution of the articulatory parameters of the tip/blade component. This is closely related to the control of the tongue body. If the tongue body is unconstrained and coarticulates freely with the surrounding vowel gestures, then the freedom of the tongue tip/blade is reduced, and hence, the flicking (or transient) movement cannot be executed successfully. Given the mutual dependency between the [+stiffness] feature of the tongue tip/blade and the [suppressed] feature of the tongue dorsum, the substitution pattern is primarily due to the unqualified, or imprecise, control of
the tongue dorsum. Support for this idea comes from the tendency that the /r/ variability occurs mainly in the non-low vowel contexts.

(10a)

(10b)

(10c)

(10d)

(10e)
It is therefore reasonable to assume that the /r/ variability in child’s speech depends on the extent to which the co-production between a tap gesture and a vocalic gesture is acquired. The ideal development takes the following steps: (i) the dorsum gesture dominates the tip/blade gesture; (ii) the dorsum activity is constrained and the tip/blade gesture takes part in the production; and (iii) the parameters [+stiffness] and [suppressed] are regulated. This development may become clear if we analyse the cases where [r] is replaced by [j] and is deleted to generate a vowel sequence.

Compare the gestural score in (10e) with that in (10b): they show the substitution pattern [r]→[0] (i.e. deletion). This pattern was not observed at the end of the data collection period (Utsunomiya, 1998). The realisation of *[k’iei] for *[k’irei] is accounted for by assuming the absence of the specification for both the tongue tip/blade and body gestures. The unspecified gesture is filled with the smooth transition from V1 to V2. In contrast, the manifestation of the forms with the palatal approximant in (10c,d) involves the dorsum gesture that is not specified as [suppressed]. For *[k’i’ei] in (10c), the tongue body is raised for the V1 /i/ and the open approximation is released rather slowly. This may generate a percept of an epenthetic [i] before attaining the V2 /e/. If this sliding of the V1 gesture becomes regulated and the articulatory parameters are ‘re-organised’ (Browman & Goldstein, 1991), then the palatal approximant emerges as seen in the form *[k’i’i’ei] in (10d).

The /r/ variability in the regional accents and adult’s careless pronunciation can also be explained in the same way as in the gestural scores in (10a,b): the parameters for the dorsum gesture are misaligned to generate a percept of [d] or [r]. It must be noted that there is a difference between the /r/ variability in child speech and that in adult speech. In the regional accent the pronunciation with the variation is socially acceptable and can be considered as lexicalised at the allophonic level. In contrast, the /r/ variability in children’s pronunciation demonstrates that the distinctive parameters and their co-production with the vocalic gesture

---

5) There is reason to assume that the tongue dorsum has a priority in the control of articulatory gestures at the earlier stages of acquisition of the motor representation. It has been reported that the child avoids the alveolar, velar, and glottal consonants and they are replaced by the palatals or palatalised consonants: for instance, [i,nu]→[i,nu] ‘dog’, [‘seror’i]→[‘cer’i’i] ‘celery’, [ne,zum’i]→[ne,zum’i] ‘mouse’, [heb’i]→[’ci’i’i] ‘snake’, [k’i,iro]→[t’i,iro] ‘yellow’ [tsu,me’k’i’i]→[tsu,me’k’i’i] ‘nail trimmer’, and [‘dzoosan]→[‘dzoosan] ‘elephant’ (examples are from Utsunomiya (1998)).

According to Jakobson (1968; 78), ‘palatalisation’ in these examples is observed particularly in children who acquire Russian, Polish, and Japanese. In other words, if the target language has the opposition between palatalised and non-palatalised, these consonants tend to be acquired in the earlier period. I cannot readily agree with this claim. This is because such palatalisation phenomena seem to be a universal tendency in child pronunciation. In English there are examples essentially similar to those in Japanese above: for instance, [dres]→[da], [noiz]→[nɔi], [kis]→[ki], and [pliz]→[bi] (Moskowitz, 1970). These examples suggest that the priority lies in the tongue dorsum that generates the palatal consonants, rather than the priority established by the phonological system of a language.
are not regularised. Thus, the articulatory nature of the /r/ variations in adult and child language is the same but its level of lexicalisation differs.

7.2.3 Summary
We focused on the acquisition of the Japanese [ɾ] in child phonology and discussed the (psycho)linguistic and physical problems in modelling the process of phonological acquisition. In previous approaches, based on phonemes or features as atoms of phonological acquisition, it is impossible to predict the variability in child speech. This is because those approaches commonly assume the input form (or feature) must be identical with the output form (or feature). To overcome such a conflict, the realisation rules that are able to change features are proposed. However, this formulation has been criticised because it is psychologically unreal.

We attempted to apply the concept of articulatory gestures (Browman & Goldstein, 1989) to the description and analysis of children’s pronunciation. We showed that the variability in child’s phonetic forms could be described reasonably as the progressive modification of the control parameters of the given articulatory gestures. It was proposed that the articulatory gestures for the Japanese [ɾ] contrasted with those for [d] in terms of the parameters [tongue tip, +stiffness] and [tongue body, suppressed]: these two parameters are assumed to co-occur in the articulation of [ɾ]. It was then shown that the proposal can attain a unified account for the /ɾ/ variability in child language acquisition, regional accents, and the careless pronunciation of adults.

7.3 CyV Variability and De-palatalisation
CyV variability in standard Japanese is a synchronic phenomenon, in which the /Cju/ syllables [ɕʰu] and [(d)zʰu] are phonetically realised as the /Ci/ syllables [ɕi] and [(d)zi] respectively. We outlined this phenomenon in chapter 2 (section 2.3.3). The examples are repeated here, with some additions, as (11) and (12) for convenience (examples from Nakajô (1980), Sakurai (1966 ), and my informal observations: phonetic transcriptions are mine). The standard pronunciations are written on the left-hand side.

(11) /sjukudai/ [ɕʰu,kudai]-[ɕi,kudai] ‘homework’
/sjuzjutu/ [ɕʰuzʰu,tsu]-[ɕizʰu,tsu]-[ɕiz̪u-tsu] ‘medical operation’
/sjuru/ [ɕʰurui]-[ɕi,ru] ‘kind (n)’
/sisjutu/ [ɕi,ɕʰu,tsu]-[ɕi,ɕi,tsu] ‘expenditure’
/sjhu/ [ɕʰufu]-[ɕifu] ‘house wife’
/seesjuku/ [ɕe,ɕʰu,ku]-[ɕe,ɕifu] ‘quietness’
/gaisjutu/ [ɡa,ɕʰu,tsu]-[ɡa,ɕi,tsu] ‘going out’
/gesjuku/ [ɡe,ɕʰu,ku]-[ɡe,ɕi,ku] ‘lodgings’
/sjutubotu/ [ɕʰu,tsu,botu]-[ɕi,tsi,botu] ‘infestation’
The CyV variability makes some words homophonous: jugyoo ['dz^u,siu,oo]—'life span' /zjumjoo/ ['dz^u,miu,oo]—'class, lecture' sounds the same as jigyoo ['dz^i,siu,oo]—'business'; [dz^w,siu,oo]—'wining a prize' and [dz^i,siu,oo]—'self-professed'; [dzi,siu,oo]—'plan' and [dzi,siu,oo]—'taste'; [dzi,tsiu,oo]—'business trip' and [(e,e)zi,tsiu,oo]—'malnutrition'.

Further examples are found in the regional accents of the south Tohoku and Kanto areas6). In the case of the regional accents, the consonant [dzi,tsiu,oo], as well as [dzi,tsiu,oo] and [(dzi,tsiu,oo)], is susceptible to the CyV variability. Its output forms fall into two groups: the ones in (13a) are essentially the same alternation pattern as in standard Japanese above, and the others are 'de-palatalisation' in (13b) that are particularly observed in the pronunciation of the south Tohoku area. The phonetic forms in standard Japanese are on the left-hand side (examples are from litoyo (1975): phonetic transcriptions are mine).

(13a) Kanto area
/sujuizin/ ['dzi,siu,tsiu,oo]—[dzi,tsiu,oo]—'husband, owner'
/tjuugakkoo/ [dzi,tsiu,oo,tsiu,oo,tsiu,oo,tsiu,oo,tsiu,oo]—'junior high school'
/sanjuizin/ [dzi,tsiu,oo,tsiu,oo,tsiu,oo,tsiu,oo]—'mathematics (archaic)'

(13b) south Tohoku area
/sujuizin/ ['dzi,siu,tsiu,oo]—[dzi,tsiu,oo,tsiu,oo]—'husband, owner'
/sajasin/ ['dzi,tsiu,oo,tsiu,oo]—'photograph'
/tjuugakkoo/ [dzi,tsiu,oo,tsiu,oo,tsiu,oo,tsiu,oo,tsiu,oo]—'junior high school'
/zjuubako/ [dzi,tsiu,oo,tsiu,oo,tsiu,oo,tsiu,oo,tsiu,oo]—'lacquered boxes'
/zjuunzio/ ['dzi,tsiu,oo,tsiu,oo,tsiu,oo,tsiu,oo,tsiu,oo,tsiu,oo,tsiu,oo,tsiu,oo]—'order'

litoyo (1975, 204) mentions that 'the high vowels [i] and [u] are centralised as [i] and [u] in many regions: the distinctions between Cyu and Ci syllables tend to be lost, especially when the consonants are fricatives or affricates' (my translation). Although there seems to be a link between the CyV variability and de-palatalisation, these phenomena have not been examined seriously in previous studies.

---

6) The south Tohoku area includes the four neighbourhood prefectures: Iwate, Miyagi, Yamagata, and Fukushima. The Kanto area includes the eight prefectures: Ibaraki, Chiba, Tochigi, Saitama, Gunma, Tokyo, Kanagawa, and Yamanashi. These areas are divided by geographical boundaries rather than isoglosses of the given phonetic characteristic. The details of the subdivisions are irrelevant for our discussion. See litoyo (1975: 201) for further discussion.
In the following sections we argue that a parametric perspective can attain a better account of the phenomena than posture-based and feature-based approaches. We first consider the acoustic and articulatory characteristics of Cyu and Ci syllables. Our discussion is limited to the case where the syllable [ç^u] alternates with the syllable [çi], while a similar mechanism can be applied to the cases of the voiced alveolopalatal fricative (and affricate). We discuss the conditioning factors for the CyV variability in standard Japanese and the Kanto area. Then, we move on to the de-palatalisation in the pronunciation of the south Tohoku area.

7.3.1 Spectral and Tongue Contact Patterns of the Alveolopalatal [ç]

Figures 7-1 to 7-4 show the spectrograms and the EPG full-printouts for the sequences [aci] and [ac^u] produced by the two speakers. We will examine these particular tokens and consider the acoustic and articulatory characteristics of the alveolopalatal in different phonetic contexts. This comparison may help us to infer the occurrence of the CyV variability.

The alveolopalatal [ç] in the sequence [aci] exhibits a long spectrum length of turbulence noise, ranging approximately from 6500Hz to 3000Hz. It can be seen from Figure 7-1 (MN) and Figure 7-3 (TM) that the lowest frequency noise stays at about the 4000Hz level throughout the fricative. This turns into the regular vertical striation of the formants when the voicing starts for the following vowel [i]. The EPG printout reveals that the narrow constriction in the front region is formed for [ç]. This constriction is released as the contact in the posterior regions increases. The increase of the posterior contact, however, tends to be less clear. This is because the consonant already involves the dorsum activity. Thus, the widening of the constriction in the front region is more obvious for the preparatory activity of the vowel [i].

In contrast, the alveolopalatal [ç] in the sequence [ac^u] displays a distinctively lower peak of noise frequency. Whereas the spectrum length is long, similar to the [aci] sequence (the range approximately 3000-6500Hz), the bottom line of the fricative spectrum drops progressively towards the following vowel. It can be observed in Figures 7-2 (MN) and 7-4 (TM) that the downturn is realised by the active lowering of the frequency to around 3000Hz, as well as the reduction of the higher frequency above 4000Hz. This continuous modification is more obvious in the spectrogram of MN than in that of TM. Because the bottom line of the noise resonance is already low in the production of speaker TM, the downturn does occur continuously but not so drastically. The EPG contact pattern for [ac^u] is so similar to that for [aci] that the two patterns are hardly distinguishable by visual inspection alone.
The sampling rate is 16KHz. Frequency and amplitude were measured by a single spectral slice at the temporal midpoint of the first half (a) and latter half (b) of the friction noise: at the point (a), the highest peak is 6437Hz (42.4dB) and the lowest peak is 3562Hz (37dB); at the point (b), the highest peak is 5312Hz (48dB) and the lowest peak is 4187Hz (47dB).
The sampling rate is 16KHz. Frequency and amplitude were measured by a single spectral slice at the temporal midpoint of the first half (a) and latter half (b) of the friction noise: at the point (a), the highest peak is 5312Hz (52dB) and the lowest peak is 3312Hz (47dB); at the point (b), the highest peak is 3562Hz (53dB) and the lowest peak is 2612Hz (46dB).
The sampling rate is 16KHz. Frequency and amplitude were measured by a single spectral slice at the temporal midpoint of the first half (a) and latter half (b) of the friction noise: at the point (a), the highest peak is 5812Hz (44dB) and the lowest peak is 4612Hz (34dB); at the point (b), the highest peak is 5312Hz (36dB) and the lowest peak is 4656Hz (34dB).
The sampling rate is 16kHz. Frequency and amplitude were measured by a single spectral slice at the temporal midpoint of the first half (a) and latter half (b) of the friction noise: at the point (a), the highest peak is 4562Hz (50dB); at the point (b), the highest peak is 4125Hz (51dB) and the lowest peak is 3625Hz (47dB).
Observation of the spectrograms and EPG full-printouts reveals that the crucial difference between [çi] and [ç'^u] is the phonetic quality of the friction: the former quality is sharp or acute and the latter is (slightly) dark or grave. This difference in the phonetic quality can easily be confirmed by observation of the actual pronunciation. Also, the active involvement of the lips is noticeable not only in [ç'^u] but also in all the Cy/u/ syllables. The spectrograms demonstrate that, while the alveolopalatal exhibits a long spectrum length, the noise frequency declines progressively towards the second vowel in the production of [ç^u]. This agrees with the results reported by previous studies (Beckman & Shoji, 1984; Yeni-Komshian & Soli, 1981; Soli, 1981 for [s, z, ʃ, ʒ] in English): the information of the second vowel is involved in the latter part of the fricative formant. In contrast, the continuous downtrend was not observed in the production of [çi]. We thus have good grounds for assuming that the loss of the frequency downtrend, or less bright phonetic quality, in the [ç^u] production gives rise to the CyV variability. It can be asked why the syllable [ç^u] in particular is susceptible to the phenomenon.

7.3.2 Distinctive Features and Enhancement

Stevens et al. (1986) and Stevens & Keyser (1989) assume that 'a minimal distinction between words can be carried out by not just one feature but by some combination of features. This redundancy could provide the listener with additional acoustic cues that could be used to reduce the possible confusion between words, particularly in situations in which there is noise or in which the speech is not clear for some other reason (Stevens et al, 1986; 427)'. The authors observed that the basic phonological contrast is often 'enhanced' by redundant features. According to this view, it is possible to formulate a hypothesis that, if the distinctive features become weaker for a certain reason and are NOT enhanced by relevant redundant features, then the basic contrast may readily be lost. We shall consider the interactions between the distinctive features of the consonant [ç] and those of the vowels [i, ui]. It will be shown that the CyV variability is a case where the enhancement by the consonant conflicts with the distinctive feature of the vowel and this conflict tends not to be supported by another enhancement.

The feature [+coronal] enhances the feature [-back] for high vowels. As Stevens et al. (1986) demonstrate, the backness of the high vowels is acoustically distinguished by the spacing between the formant frequencies: while F1 is generally characterised as a low-formant frequency for the vowels [i] and [ui], F2 is closer to F3 in [i] but is closer to F1 in [ui]. The frequency of F2 and F3, Stevens et al. (1986) mention, will further increase by raising the tongue blade, the movement which is the articulatory correlate of the feature [+coronal]: this raising gesture is interpreted as an enhancement of the feature [-back]. As discussed in chapters 5 and 6, the consonant [ç] is a coronal complex segment, the narrow
constriction of which is formed by the raising gesture of the tongue blade and the medio-dorsum against the hard palate. We therefore assume that this consonantal configuration enhances the feature [−back] for the given high vowel within the same syllable.

The above assumption can be illustrated in the diagrams in (14) below. The articulatory features and their acoustic consequences (i.e. auditory features7) for the [qi] syllable are shown in the diagram (14a) and those for the [çui] syllable in (14b) ([ant]= [anterior] and [distr]=distributed).

(14a)

(14b)

7) Flemming (1995) proposes a model of phonological representation in which articulatory and auditory features are paralleled. He systematically describes various auditory features such as [+highest F2] and [−low F2, F3]. We agree with the idea that the auditory features have an important role in phonological processes. Yet, our representations above are based on the observations of the spectrogram, as well as common acoustic descriptions of the consonant and vowels in question.

Using the features proposed by Jakobson et al. (1952), the segments in question are specified as follows: the consonant [ç] is [+sharp] and [−grave]; the vowel [i] is [+diffuse] and [−grave]; the vowel [u] is [+diffuse] and [+grave]. The value of the feature [grave] makes a contrast between the two high vowels. Also, the vowel [u] opposes the feature [−grave] against the consonant [ç].

314
The auditory features of the consonant [ç] describe its spectral characteristics and are tentatively specified here by three points: [High Intensity], [Long Spectrum Length], and [Strong Mid-frequency] (see Figures 7-1 to 7-4 later in this section; also see Figures 5-6 and 5-7 in chapter 5; Halle & Stevens, 1997; Stevens & Keyser, 1989; Strevens, 1960). The labial feature [±round] is not distinctive either for the consonant [ç] or for the vowels [i, u] in Japanese: the feature is filled in by the lower level phonetic rules.

The [coronal] feature of the consonant generates a situation where CyV variability is likely to occur. Given that the consonantal gesture and the vocalic gesture are coordinated with each other in the relevant timing, the coronality of the alveolopalatal [ç] has different effects upon the following vowels [i] and [u]. As shown in (14a), enhancement by the feature [coronal] would work well in the syllable [çi]. This is because the vowel [i] already contains the feature [−back]. Further strengthening of this feature is accomplished. In contrast, enhancement by the feature [coronal] causes a conflict in the coordination between [ç] and [u]: the coronality of [ç] strengthens the feature [−back] for the high vowel [u], in which the feature [+back] is distinctive. In other words, the feature [coronal] for [ç] does not ‘enhance’ but does ‘degrade’ the backness feature for the vowel [u]. Acoustically, the frequency of F2 in [u] may increase towards that of F3, rather than F1. This effect agrees with the place of articulation for the vowel [u] in Japanese: it is not actually back but central and is articulated by raising the post (and medio) dorsum towards the palatal region. Thus, both the enhancement by the feature [coronal] of [ç] and the advanced (or centralised) constriction place of [u] provide a potential condition for the manifestation of the CyV variability.

Furthermore, the above situation may cause neutralisation between the syllables [çi] and [çui] when the vowels are devoiced. This expectation, however, is disproved. Beckman & Shoji (1984) conducted production and perception experiments of devoiced [çi] and [çui] syllables in the minimal pairs such as [çık,koo] ‘invalidation’ and [çuk,koo] ‘leave a port’, [çıt,too] ‘performing a medical operation’ and [çüt,too] ‘appearing in court’, and [ˈkaçi] ‘song’ and [ˈkacui] ‘sing’. Beckman & Shoji found that: the ‘deleted’ vowel coloured the spectrum of [ç] in those words but the mean influence was small; the [çui] syllable indicated a significantly lower mean frequency of turbulent noise; these spectral effects (the velar or labiovelar timbre) tend to occur later during the production of the fricative; and the listeners used the different colourings of the consonant for identifying the deleted vowel. Assuming that the devoiced vowels are already ‘deleted’ at a higher level\(^8\), Beckman & Shoji concluded that: the two timbres are derived from a low-level anticipatory coarticulation; and that the

\(^8\) This assumption implies that vowel devoicing is a categorical phenomenon, rather than gradient. However, this in fact conflicts with their data. Beckman & Shoji’s data reveals that the vowel in the potentially devoiced environment does not always loose the voicing feature. This characteristic has been reported in a number of previous studies (e.g. Kondo, 1997).
information of the 'deleted' vowel is recovered primarily from the different co-production patterns of the lingual gesture\(^9\), since the laryngeal gesture for the syllable remains voiceless. Thus, the different co-production patterns prevent the two devoiced syllables from neutralisation. This implies that: if the lingual gestures maintain the coordinative patterns specific to the syllables [çi] and [çui], the CyV variability may be avoided, even though the feature [+back] for [u] is 'degraded' by the enhancement of the feature [coronal].

However, the results of our EPG experiment do not support the claim made by Beckman & Shoji (1984). We showed, in chapter 6, that the gestural coordination between the tongue tip/blade and dorsum was precisely timed with respect to a syllable type, namely Ci and CyV: the two components were more blended in CyV syllables than in Ci syllables. Only the alveolopalatal [ç] was found to be indistinguishable in the degree of co-production in the front and the central region (i.e. the parameter SEQUENCE OVERLAP). In addition, the duration of the dorsum contact for a Cy/u syllable was found to be fairly similar to that for a Ci syllable, but was substantially longer than that for a Cy/a syllable. We suggested that this tendency was primarily due to the manner requirement of [ç]: the fricative requires the precise shape of the tongue, the configuration that has to be maintained for a relevant time interval (see chapter 5). The consequence of these conditions for intergestural coordination is that the timing of the dorsum gesture, with respect to the tip/blade gesture, tends to be identical for both [çi] and [ç^u] syllables. Therefore, it is not plausible to assume, as in Beckman & Shoji (1984), that the different patterns of lingual gesture are primary cues for the listeners to identify the two devoiced syllables in question.

The apparent solution for the above 'disadvantages' of the [ç^u] syllable is to 'enhance' the feature [+back] for the high vowel [u]. As shown in (14b), the acoustic correlate of that feature is 'the closer proximity of F2 to F1 than to F3' (Stevens et al. 1986). This acoustic property can be enhanced by the feature [+round] (Stevens et al. 1986; Stevens & Keyser, 1989). The problem now becomes not only the enhancement mechanism within the syllable, but also the representation of the vowel feature.

Although the high vowels [i, u] in Japanese are commonly considered as articulated with the unrounded lips, and both are specified by the same feature [−round] as in (14a,b) above, the feature [−round] is not an appropriate descriptor. The two vowels differ essentially in the horizontal aspect of the lip gesture. The lips are vertically contracted, or compressed, in the high vowel [u] (see chapter 2, section 2.5.1). We have represented this lip activity by the phonetic transcription [ç^u], instead of [çu]. Ladefoged & Maddieson (1996) propose the

---

\(^9\) This conclusion relies heavily on the imprecise phonetic observation that: both [i] and [u] are produced with the 'unrounded' lips that are specified by the feature [−round]. This preconception may lead Beckman & Shoji (1984) to the idea that the two timbres of [ç] are maintained by different coordinative patterns of the lingual gestures.
features [compressed] and [protruded] for identifying the various degrees of lip rounding. This helps us to specify the vowel [ui] with the feature [compressed] and the vowel [i] without it: the two high vowels contrast with each other in the degree of the (vertical) compression of the lips, namely [±compressed]\(^{10}\), as well as the feature [±back]\(^{11}\).

It must be noted that the relationship between labial gestures and lingual gestures in sibilant productions is complex. Brown (1981) observes that the consonants in English fall into three categories in terms of the lip activities: (i) /ʃ, ʒ, tj, dʒ, r/ with lip protrusion and eversion; (ii) /f, v, s, z/ with slight protrusion and eversion of the lower lip; and (iii) /w/ with compression of the corners. Brown (1981; 68) claims that ‘most rounding in natural speech derives from ‘rounded’ consonants rather than from ‘rounded’ vowels’ on the grounds that ‘the consonants will retain their rounding even in the environment of non-rounded vowels’. It is desirable to distinguish the coarticulatory effects of the adjacent vowels on the one hand, from the inherent articulatory requirements (or pronunciation habits) for sibilants on the other. However, this distinction is not always clear-cut. Faber (1989a,b, 1990, 1991b) examined the contribution of lip gestures to sibilant articulations in Catalan, English, German and Italian, using a SELSPOT opto-electronic tracking system, spectrograph and EPG. It was found that [ʃ] involved more lip protrusion than [s] in the /a/ and /i/ symmetrical contexts, but not in the /u/ symmetrical context, where the difference tends to be neutralised. It is possible to assume that there is an articulatory target specific to the sibilant gesture. This target position, however, varies considerably with vowel contexts, speakers and languages. Faber’s data reveals the complex interactions between the three variables for lip gestures: the most consistent trend may be that lip protrusion is less involved in the /a/ context. Although we have no direct evidence to show the inherent characteristics of lip gestures for Japanese [ç], the downtrend of the fricative formant is arguably considered as the contextual effects of the following vowel [uu].

7.3.3 A Gesture-based Generalisation of CyV Variability and De-palatalisation

Given the [+compressed] lips for [ui], the time varying features of the [ç] articulations are represented in the diagrams in (15) and (16). They compare the hypothesised gestural scores

---

\(^{10}\) Using two features [protruded] and [compressed], Faber (1991a) attempts to describe the contextual, idiosyncratic and inherent properties of lip gestures in English consonants.

\(^{11}\) Schane (1973) discusses the relationship between [back] and [round]. Schane attempts to resolve the problem by assuming a hierarchy for those two features: all front unrounded vowels would be marked for frontness (i.e. [−back]) but unmarked for rounding, whereas all back rounded vowels would be marked for rounding (i.e. [+round]) but unmarked for backness (Shane, 1973; 180). This proposal does not help our situation: the back ‘unrounded’ vowel such as [uu] contrasts with the front ‘unrounded’ vowel [i] in the feature [back] only. The point is that the features for labiality and backness consist of subparts of a phonetic/phonological contrast (see De Jong (1995) for the discussion on the redundancy of the labiality in the production of the velars.)
for the syllable [çi] and [çʰu], followed by a voiceless consonant. The coronal complex segment [ç] contains the lingual and laryngeal gestures: the tongue tip/blade component

(15a)

TONGUE TIP

fricative

alveolar

TONGUE BODY

narrow

palatal

Lips

GLOTTAL

opening

closing

opening

(15b)

TONGUE TIP

fricative

alveolar

TONGUE BODY

narrow

palatal

Lips

GLOTTAL

opening

opening

(16a)

TONGUE TIP

fricative

alveolar

TONGUE BODY

narrow

fronted velar

Lips

GLOTTAL

opening

closing

opening

(16b)

TONGUE TIP

fricative

alveolar

TONGUE BODY

narrow

palatal-fronted velar

Lips

GLOTTAL

opening

opening

318
[fricative alveolar], the tongue body component [narrow palatal], and the glottal component [opening]. The high vowel [i] in (15) consists of the tongue body [narrow palatal], the glottal [opening-closing], and the lips [-compressed]. In contrast, the vowel [ui] in (16) is composed of the tongue body [narrow fronted-velar] and the lips [+compressed]. The lingual gestures for the following consonant (C2) are irrelevant and are left unspecified. The diagrams in (15b) and (16b) illustrate the substantial changes of the above parameters in the devoicing environment: the curves superimposed on the tongue tip/blade and body gestures indicate the ideal articulatory trajectories over time. The glottal parameter of the vowel incorporates that of the adjacent consonants to become voiceless, namely the glottal [opening].

Differences between the syllables [çiC] and [ç'^uC] can be sought in the patterns of the two independent gestures. An unappealing possibility is to suppose that the two components of the tongue are coordinated with respect to the syllable type, Ci and CyV (Beckman & Shoji, 1984). The articulatory trajectories in (15b) and (16b) reflect the results of our EPG experiments that the tongue tip/blade and the dorsum are more blended in CyV syllables than in Ci syllables. However, as mentioned above, the degree of overlap between the two components of the tongue tends to be indistinct. In addition, the duration of the dorsum contact tends to become fairly similar between the two syllables, when the consonant is the alveolopalatal [ç]. Thus, the pattern of lingual gestures is less likely to function as a primary cue for devoiced [çi] and [ç'^u] syllables. An alternative account is to assume that different lip gestures for the two high vowels manifest a phonological contrast between the syllables. As illustrated in (16b), the [+compressed] lips extend to the consonantal articulation. This extension may or may not start at the onset of [çi], because the downtrend of the fricative formant occurs primarily later during the narrow constriction. Given these temporal movements, the [+compressed] lips can characterise the [ç'^u] syllable entirely or partially by giving the dark (or less bright) phonetic quality. This explanation accords well with the facts that we observed in the spectral properties and the EPG contact characteristics of [çi].

It follows from what has been discussed above, that the CyV variability arises from the misalignment of the timing of the lip [+compressed] gesture for the syllable [ç'^u]. Speakers differ in their realisation of the feature [+compressed]. We have seen that, while the spectrogram of MN reveals the gradual downtrend of the fricative formant, the spectrogram of TM shows the less dramatic downtrend but the relatively low formant peak throughout the fricative noise. This difference implies that the vertical compression of the lips is activated earlier in TM than in MN. If the lips are not sufficiently compressed and therefore are unable to realise the darker timbre, then the phonetic quality of the [ç'^u] syllable becomes very close to that of the [çi] syllable. In order to avoid potential neutralisation, or CyV variability, it is likely that the gesture [+compressed] is 'enhanced' consciously by the gesture [+protruded]. Therefore, it is reasonable to suppose that the two specifications for the lip gesture,
[+compressed] and [+protruded], constitute a continuum; and that both lip gestures are phonological in the case of the distinction between the syllables [çi] and [ç^u] in Japanese. The explanation can be applicable not only to the cases with the voiceless fricative [ç] but also to those with the voiced fricative [z] (and affricate [dz]).

We now turn to depalatalisation in the regional accent of the south Tohoku area: /sjuzin/ ['ç''u(d)ziN]→[süü(d)ţiün] ‘husband, owner’, /tuugakko/ [tç''u,u'gakko→tç''u,u'ňakko]→[tsu'u,gakko→tsu'u,ňakko] ‘junior high school’, and /zjuubako/ [dz''u,ubako]→[dzüü.cli] ‘lacquered boxes’. The gestural scores below exemplify the portion of the change in ['ç''u(d)ziN]→[stü(d)ţiün].

(17a)

<table>
<thead>
<tr>
<th>TONGUE TIP</th>
<th>TONGUE BODY</th>
<th>Lips</th>
<th>GLOTTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>fricative</td>
<td>narrow palatal</td>
<td>+compressed</td>
<td>opening</td>
</tr>
</tbody>
</table>
| alveolar   | narrow fronted velar | | opening-
closing |

(17b)

<table>
<thead>
<tr>
<th>TONGUE TIP</th>
<th>TONGUE BODY</th>
<th>Lips</th>
<th>GLOTTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>fricative</td>
<td>half-close retracted palatal</td>
<td>+compressed</td>
<td>opening</td>
</tr>
</tbody>
</table>
| alveolar   | | | opening-
closing |

De-palatalisation in the south Tohoku accent is captured by the accent-specific articulatory control underlying the syllable [ç''u]: ‘[i]n a ‘thick’ accent the [accent-specific] setting might be given priority’ (Knowles, 1978; 90). In standard Japanese (17a) the tongue body gesture for [ç] is [narrow palatal] and is coordinated with the gesture [narrow fronted velar] for the following [tu]. In contrast, in the accent of south Tohoku (17b) the [narrow palatal] gesture for [ç] overlaps with the tongue body gesture for the centralised vowel [tı] that is specified by [half-close retracted palatal]. It is plausible to assume that the vowel has a priority in the timing of the tongue body gesture, on the ground that the centralised vowels [i, tı] are the characteristic of the south Tohoku accent (litoyo, 1975). Given this, during the production of the alveolopalatal the tongue body anticipates the following gesture for the centralised vowel [tı]. This anticipation effectively reduces the degree of the raising gesture.
[narrow palatal] that is required for [ç]. In consequence, the tongue shape for the narrow constriction of [ç] becomes fairly similar to that of [s]. There is no change in the specification of the tip gesture and the lip gesture may or may not be retained in the regional accent in question. Thus, de-palatalisation in the south Tohoku accent is characterised as the changes in articulatory parameters and overlap that are motivated by the accent-specific articulation of the high vowels. This modification is believed to be lexicalised at the allophonic level.

7.3.4 Summary
The CyV variability and de-palatalisation have been discussed in terms of the systematic articulatory control underlying them. Based on the observation of the spectrograms and EPG full-printout, we demonstrated that the consonant [ç] before [ui] differed in the phonetic quality of the friction from that before [i]: [ç^u] was acoustically characterised as having the lower peak, or progressive downtrend, of the fricative formant. We have assumed that CyV variability is the phonetic process that involves the loss of that characteristic timbre of darkness. The susceptibility of the syllable [ç^u] to the process is predictable from the concept of ‘enhancement’ (Stevens et al., 1986) and the articulation of [ui] in Japanese. We proposed that the vowel [ui] contrasts with the vowel [i] in terms of the lip gesture [+compressed], rather than being similar in the feature [−round]. We then suggested that the CyV variability was actuated primarily by the misalignment of the timing of the lip [+compressed] gesture for the syllable [ç^u]. In contrast, de-palatalisation in the regional accent of the south Tohoku area was characterised as the accent-specific modification of the articulatory parameter and its coordination pattern. We showed that the tongue body gesture for the centralised vowel [iui] suppressed and took over that for [ç]: de-palatalisation could be understood as undershoot of the raising gesture of the tongue body.

7.4 Vowel Devoicing
A common allophonic description of vowel devoicing in Japanese is that the unaccented high vowels [i, u] are devoiced when preceded by a voiceless consonant and followed by a voiceless consonant or a pause (e.g. Kondo, 1997; Vance, 1987). This phenomenon is not restricted to standard Japanese. The structural description takes the form of the rewrite rule in (18a) and the examples are given in (18b-d).

(18a) \[
\begin{array}{ccc}
\text{V} & \text{V} & \text{C} \\
+\text{high} & -\text{accented} & \\
& -\text{voice} & -\text{voice} \\
\end{array}
\] \rightarrow \left\{
\begin{array}{c}
\text{C} \\
-\text{voice} \\
\text{pause}
\end{array}\right\}
\]

(18c) [ki,çi] ‘riverbank’, [tçi,see] ‘reign’; compare [kici] ‘knight’, [tçi,see] ‘intellect’
(18d) [tçi,gi’mas̄u(t)] ‘That’s wrong’; compare [tçi,gi’mas̄u] ‘Is that wrong?’

321
The rule in (18a) is just a gross generalisation that misses the following points: (i) unaccented non-high vowels such as in [ka, karui] 'hang' and [ko, ko] may sometimes be devoiced (Sakuma, 1929; Kawakami, 1977); (ii) accented high vowels are often devoiced for example in ['cik^o-'cik^o] 'death' and [su,i'tsukyuu-su,i'tsukyuu] (NHK, 1998); (iii) unaccented high vowels can be devoiced when preceded by a pause and followed by a voiceless consonant such as in [i,k^i-j,k^i] 'smart' and [ui,çiro-ju,çiro] 'back' (Imada, 1981); (iv) there may be a hierarchy of devoicing rate by the preceding voiceless consonants (fricatives>affricates>stops) and by the following consonants (affricates>stops>fricatives) (Han, 1962); (v) devoicing applies gradiently, rather than categorically, showing the various grades, or spectral characteristics, such as voiced (clear voice bar with glottal pulses), partially devoiced (one or two weak glottal pulses), completely devoiced (no glottal pulses visible on spectrogram) (Beckman, 1996; Beckman & Shoji, 1984; Kondo, 1997); (vi) devoicing may scarcely occur when the following consonant is [h] such as in [ei,harai] ‘payment’ and [e=,u,han] ‘ringleader’ (Tsuchida, 1997); and (vii) the consecutive devoicing does occur as in [ku,tsuc^ita-ku,tsuc^ita] ‘socks’ but the pattern may depend on idiosyncratic and stylistic factors such as speaking rate (Han, 1962; Kondo, 1997; Sakuma, 1929)\(^{12}\).

\(^{12}\) Recently Coleman (1998, 258ff.) has proposed a completely different conceptualisation. Instead of viewing devoicing as the loss of the feature [+voice], Coleman claims that the vowels in Japanese are ‘exceptionally’ voiced. This idea comes from the orthographic convention in Japanese. By kana characters the phonemic sequences /ku/ and /gu/, for instance, are written as ‘<’ and ‘C’ respectively. The distinction is made by adding or omitting a dot-like diacritic ‘.’ at the upper-right corner: this diacritic is called dakuten. Coleman interprets this as marking the laxness of the onset consonant (i.e. /g/), as well as the information that the nucleus is voiced without fail. This leads Coleman to the idea that phonological voicing is operative at the mora level, rather than at the segmental level. Therefore, the consonant and vowel elements of unmarked mora agree with each other in voicing, whereas those of marked mora do not: the combination of voiceless onset and voiced nucleus is exceptional. Coleman states that ‘the grammar should define not when vowels are devoiced, but rather when voiced vowels occur in voiceless moras (the marked case) (261)’. The specification of voiced, or voiceless, nuclei depends on the presence of an accent nucleus and on tempo: the hierarchy of the voicing feature is in the order ‘accented (voiced vocalic nuclei) > unaccented (whispered vocalic nuclei) > unaccented/accelerated speech (fricative nuclei)’. Phonetically the temporal overlap of ‘resonance’ is assumed for the realisation of ‘fricative nuclei’.

Although it is beyond the scope of this discussion to examine this radical conceptualisation in detail, I would like to mention one important implication of the above idea. The following examples of the compound formation and the loanword pronunciation involve an accent shift that is induced by vowel devoicing.

(i) [k,i,m,i]+[tat^ci]→[k,i,m,ïtat^ci] ‘you’; compare [u,ta,ta^ci]+[‘tat^ci]→[u,ta^ci,ïtat^ci= u,ta’ta^ci,ïtat^ci] ‘we’
(ii) [ei,zu,ko]+[‘ken]→[eiuzu,ko’ken] ‘Shizuoka Prefecture’; compare [na,ñasak,i]+[‘ken] →[‘na,ñasa,k,i,ken=ña,ñasa,k,i,ken] ‘Nagasaki Pref.’

There are general rules of accent assignment for each case: for (i), the morpheme /tat^ci/ ([‘tat^ci]) assigns the accent nucleus to the mora immediately preceding it; the same for the morpheme /ken/
A considerable number of studies have been made with an aim of generalising the laryngeal mechanisms of vowel devoicing. However, to my knowledge, no serious attention has been given to lingual gestures during the production of devoiced vowels. The rule in (18a) explicitly specifies the state of the laryngeal gestures for the target vowel and the contextual consonants, while no statement is given regarding the state of the supralaryngeal gestures that are certainly involved in the production of the devoiced syllable. Given the legitimate relationship between oral and glottal events, the phenomenon can be understood as an extreme case of vowel shortening: the articulatory gestures for the high vowel are coproduced with those for the neighbouring consonants. To put it the other way round, the supralaryngeal gestures for the preceding and following consonants are activated within the minimum temporal interval. This view is based on that developed by Fowler (1981, 1983), in which she suggests that vowel shortening in English can be explained partly by the extent to which a stressed vowel coarticulates with neighbouring unstressed syllables. An adequate description of the mechanisms of vowel devoicing should include both laryngeal and supralaryngeal aspects. It is not known how the oral-laryngeal coarticulation is systematised in the production of the devoiced vowel in Japanese. We attempt to develop the idea that vowel devoicing in Japanese is a part of the vowel shortening process, with special reference to lingual gestures.

We begin by reviewing the characteristics of the laryngeal gesture during the production of devoiced vowels in section 7.4.1. We discuss implications of the recent proposal that devoicing is derived essentially by a coarticulation, or gestural overlap, of the laryngeal gestures between the (preceding) consonant and vowel (Beckman, 1996; Jannedy, 1995; Jun et al., 1998a,b). In section 7.4.2 we report the results of preliminary experiments using the combined technique of EPG and laryngograph (section 7.4.2). In section 7.4.3 we argue that the coordinative pattern of lingual gestures plays an important role in some phonetic processes and synchronic sound changes.

7.4.1 The Laryngeal Gesture in Devoiced Syllables

There are two possible lines of experimental research: one into the laryngeal activities of the devoiced high vowels themselves, and the other into those of the sequence containing a devoiced vowel. Reviewing both types of studies we discuss an application of the concept ‘gestural overlap’ (Browman & Goldstein, 1990a) to vowel devoicing.

The laryngeal adjustment for vowel devoicing is an active widening of the glottis.

(\textquoteright{k\textsc{e}N}) in (ii); and as shown in (iii) loanwords generally have an accent, either LH or HL, on the antepenultimate mora, if that mora is not moraic. However, in all cases, the accent location moves leftwards by one mora if the predicted position is the potential site for devoicing. The occurrence of this shift is reasonable if we assume that a devoiced vowel, or mora, is unmarked, that is, resistant to the accent assignment.
Sawashima (1971), in the fibercopic study of the laryngeal activity, found that the degree of the glottal opening for a devoiced vowel was fairly similar to that for a voiceless consonant. Hirose (1971) obtained electromyographic (EMG) recordings of the vocalis muscle (the internal thyroarytenoid muscle) that is considered to be responsible for tensing the vocal cords and of the lateralis muscle that rotates the arytenoids cartilage for narrowing the glottis (Hardcastle, 1976, 77ff.). Hirose's EMG data reveals that there is no peak in muscle activity for the devoiced syllables but there is a clear burst for the voiced syllables. Because both vocalis and lateralis muscles cause the adduction effect the devoicing gesture is considered as the abduction of the glottis. Thus, a speaker consciously controls devoicing of the high vowels; it is not an automatic accommodative process. Note in passing that whispered vowels differ from devoiced vowels: the former involves more constricted glottis and the latter does occur in whispered speech, although the basic mechanisms for devocalisation are considered to be the same (Wietzman et al., 1976).

Among the voiceless consonants the laryngeal gesture for the consonant [h] is unusual and does not always provide a potential condition for vowel devoicing. Using the EMG technique Yoshioka (1981) obtained the movement pattern of the abducting muscle the posterior cricoarytenoid (PCA) and that of the adducting muscle the interarytenoid (INT). During the production of the words such as [çi,see] 'posture' and ['çisee] 'life and death', Yoshioka found that the two muscles were complementary in the production of voiced and voiceless high vowels; for devoiced high vowels the PCA is activated but the INT is suppressed; the activities are reversed for voiced high vowels; the accent nucleus is reflected in the strong activity of the INT. In contrast, these results were not obtained for the sequences containing [h]. The consonant was often voiced in words such as [çi,hee~çi,fiee] ‘private soldier’. Furthermore, Yoshioka’s data reveals that there is no significant divergence in the activities of the two muscles during the production of ‘voiced’ [fi]: the PCA is clearly activated to open the glottis, while the INT is not positively activated to close the glottis. That is, a sort of vocal folds vibration for intervocalic [fi] is caused by the contradictory activities of the laryngeal muscles: the wide-open glottis similar to other voiceless fricatives produces the voicing effects\(^{13}\). The extent to which this exceptional voicing of [fi] is consciously

\(^{13}\) Akamatsu (1997; 97) mentions that ‘[i]n the articulation of [fi], the vocal folds are loosely held against each other so that the airstream passing between them makes them flap against each other and thus causes vocal vibration...The reason why [fi] occurs in such [sic] case [e.g. Takahashi [ta,kaʃaçi]] is that it is more economical for [sic] vocal vibration necessary for the articulation of the two surrounding vowels to be maintained throughout the articulation of the intervening glottal fricative’(italics are mine). It is reasonable to explain the occurrence of [fi] as laryngeal coarticulation between the glottal fricative and the surrounding vowels. However, we cannot readily the idea that continuous voicing is more economical, particularly if it implies automatic biomechanical response. This view, in itself, actually contradicts Akamatsu’s description. The laryngeal gesture that ‘the vocal folds are loosely held against each other’ can be considered as an actively controlled activity. It would be misleading to appeal to the concept of economy of speech without discussion.
controlled is a matter of controversy. It is, however, clear that the consonant is not always counted as one of the conditional factors for vowel devoicing.

Let us now extend the discussion into the laryngeal activity during the production of the sequence containing voiceless consonants and a devoiced vowel. Systematic studies on devoicing patterns in voiceless clusters have been carried out for Swedish (Löfqvist & Yoshioka, 1980), American English (Yoshioka et al., 1981), Dutch (Yoshioka et al., 1982a), and Japanese (Yoshioka et al., 1982b). There are two advantages in these studies: (i) the experiments use the combined technique of photoelectric glottography, fiberoptic filming and laryngeal electromyography; and (ii) the basic combinations of a fricative /s/ and a stop /k/ (or /p/) are used for the test utterances: the individual consonants and their combinations are located in word-initial or word-final position, thus /#s-/, /-s#/, /#sk- and /-ks#/. For Japanese the CVC and CVCC sequences are made from the consonants [s], [Ç], [k] and a high vowel. Under these conditions a series of experiments were conducted to explore the temporal aspects of laryngeal abduction and adduction. In addition, the comparable methodology and speech items enable us to compare the languages with each other. We contrast the data reported for Japanese (Yoshioka et al., 1982b) with that for American English (Yoshioka et al., 1981).

The opening gesture of the glottis is characterised by the number of peaks that vary systematically with the structure of the voiceless consonant sequence. The data reported by Yoshioka et al. (1982b) reveals that, for Japanese, the glottal opening gesture falls into two patterns: a monomodal pattern in which a single peak of the opening gesture is observed; and a plateau-like (bimodal) pattern in which a gradual peak slope is observed. A monomodal pattern was found for the $C_1VC_2$ sequences where $C_2$ was not a part of a geminated consonant: [kj,kee] 'malformation', [k'j,see] 'regulation', [çj,kee] 'verse form', and [çj,see] 'posture'. In contrast, a plateau-like pattern was found for the words such as [kjk,kee] 'felicitous (event)', [kjs,see], [çjk,kee] 'rude' and [çjs,see] 'misgovernment'. In both cases the peak glottal opening was attained during the devoiced vowel, rather than the preceding or following voiceless consonant. The peak velocity and size were found to be faster and greater when the fricative precedes the devoiced vowel.

For American English Yoshioka et al. (1981) reported that only one peak of the glottal opening was found for the 'autosyllabic /sk/ sequence [skeil] and [mæsk]'\(^{14}\); but that\(^{325}\)

---

\(^{14}\)These results simply demonstrate the fact that the Japanese [çjk] and English [sk] are both produced by a single opening gesture for the glottis. The sequences [su:k] and [ku:s], for instance, may sound fairly similar to [sk] and [ks], yielding an acceptable impression as a consonant cluster: for instance, compare the loanwords [su:ç,keeto] 'skate' and [’bokkutsu:] 'box' with the corresponding original English words [skeil] and [boks]. Even though their phonetic quality is similar to each other, the timing control may not be compatible: the peak velocity of the laryngeal gesture and other parameters of supralaryngeal gestures may be controlled language-specifically.
two distinct peaks were observed for the heterosyllabic /-s#k-/ and /-k#s-/ sequences such as in My ace caves [eis keivz] and I make sale [meik seil]. Yet, the boundary does not always require a separate opening gesture. For a long consonant /-s#s-/ and /-k#k-/ such as in My ace sales [eis seilz] and I make cave [meik kerv], the EMG and glottographic curves showed the plateau-like pattern with one single peak, similar to that for a geminate consonant in Japanese. However, the sequences /-sk#k-/, /-sk#s-/ and /-ks#k-/ always exhibited the two-peak pattern, each peak of which corresponded to the word-final cluster and the word-initial consonant: I mask cave [mæsk keiv], I mask sale [mask seil] and He makes cave [meiks ker]. Thus, whether the given consonant involves an independent opening gesture or not, depends on the phonetic feature that is determined at the prosodic level. Yoshioka et al. (1981) conclude that the aerodynamic requirements for the voiceless obstruents, aspiration and friction noise, necessitate a single opening gesture of the glottis.

Munhall & Löfqvist (1992), using transillumination and fiberoptic video recording, examined the combination of the successive opening and closing movements of the larynx during the utterance Kiss Ted produced at various different speaking rates. It was found that a single opening movement occurred at fast speaking rate; two separate movements took place at slow speaking rate; and at intermediate speaking rate, a single, but not plateau-like, movement was involved. When [s] is produced by a separate movement (i.e. two-peak pattern), both the friction onset and peak opening tend to occur almost simultaneously. In contrast, when the sequence [-s#t-] forms a single movement, the peak opening tends to occur at the transition between the friction offset and the complete closure. These results indicate that the underlying two gestures of the glottal opening coalesce to become a single composite gesture. Munhall & Löfqvist simulated this blending by combining the two laryngeal movements with each other at various degrees of overlap. It was found that: (i) ‘the two underlying gestures’ time courses were unaltered by the overlap; (ii) the second gesture’s onset is unconstrained; that is, it can begin at any phase of the first gesture; and (iii) the observed movement is the vector sum of the two underlying patterns for each degree of overlap’ (Munhall & Löfqvist, 1992; 117). Munhall & Löfqvist suggest that a single smooth glottal movement could be considered as the manifestation of the two underlyingly separate gestures of the larynx. Saltzman & Munhall (1989) take the same view for a plateau-like pattern observed for the within-word clusters.

All the evidence above suggests that it is possible to apply the concept of ‘gestural

15) We may doubt this point in the case of Japanese. Although Yoshioka et al. (1982b) examine the Japanese voiceless cluster only within a word (e.g. [cij,keel]), it is plausible to assume that the separate opening gestures are necessary for the sequence [-cij#k-]: for instance, [cij'kaeij k'i,m'i=iqa] ‘But you...’ and [nasu=kat,tajo] ‘(I) bought aubergines’. There are two reasons. One is that [k] may be weakly aspirated. The other is that the word (or phrase) boundary differs in strength (cohesion) from the syllable boundary.
overlap’ (Browman & Goldstein, 1989, 1990a) to vowel devoicing. Jun & Beckman (1993, cited in Kondo, 1997) first proposed such a view for Japanese and Korean. In this approach the phenomenon is considered as ‘undershoot’—that is, as a failure to achieve an interval of audible voicing due to overlap between the glottal adduction target in the vowel and the glottal abduction target for a neighbouring voiceless consonant (Jun et al., 1998; 1). Jun et al. refer to two facts as evidence to support their analysis: (i) the degree of devoicing constitutes a continuum ranging from fully voiced to completely devoiced; and (ii) the high vowels are most susceptible to devoicing. This is because they are phonetically so short that ‘a neighbouring consonant’s devoicing gesture covers a greater portion of the vowel’s gestural activation interval’ (Jun et al.; op.cit., 2). Thus, vowel devoicing is not a phonological process that is modelled by the rewrite rule in (18a), but ‘another example of the kind of gradient phonetic effect resulting from more or less subtle adjustments to the timing of otherwise invariant gestural specifications’ (Jun et al., ibid.). The gestural overlap analysis elegantly captures the gradient nature of the laryngeal gestures during vowel devoicing in Japanese, as well as other languages such as Korean (Jun & Beckman, 1993; Jun et al. 1998a,b), Montréal French (Beckman, 1996), and Turkish (Jannedy, 1995).

However, the analysis fails to account for various important aspects of vowel devoicing in Japanese. First, the strong hypothesis that devoicing is the results of subtle biomechanical (mis)alignments is simply wrong. As explicitly specified in the rule (18a) above, devoiced vowels are conditional allophones. This does not mean that the phenomenon is totally phonological (i.e. a categorical phenomenon). Kondo (1997) showed that devoicing occurs both in fast and in slow speech, although devoicing rates increase in fast speaking tempo. As Kondo mentions, the fact that the frequency of occurrence varies with speaking rate suggests that vowel devoicing is not constrained solely by phonological factors. Thus, various degrees of devoicing indicate that the phenomenon is gradient in nature, but do not necessarily imply that the process is ‘phonetic’ in the sense that it is totally induced by biomechanical constraints. Second, the gestural analysis ignores devoicing that occurs in word-initial and word-final position: for instance, [j,ka] ‘squid’ and [o,hajoo go,zaimasu] ‘good morning’. It might be possible to assume that the glottal stop precedes and follows the high vowel in the initial position of utterance or (intonational) phrase. Yet, we have no evidence of how the larynx moves from the resting, or constricted state, through the devoiced vowel to the consonantal constriction, or vice versa. Third, the consecutive devoicing environment is not accounted for adequately (Kondo, 1997). Finally, there is no satisfactory discussion at all for supralaryngeal gestures in general, or lingual gestures in particular. It might be the case that lingual activities are taken for granted as paralleled completely with laryngeal activities. However, there is no empirical evidence to support this expectation. Nevertheless, the concept of gestural overlap is attractive and provides strong support for
capturing gradient characteristics of the laryngeal gesture. In the next section we are principally concerned with the fourth point, above, that of lingual gestures during the production of devoiced syllables.

7.4.2 Lingual Gestures in Devoiced Syllables

7.4.2.1 Tongue-palate Contact Patterns

The experiment was carried out to examine how oral and glottal gestures are coordinated during the production of devoiced syllables. The data was obtained from one speaker (MN), using the combined technique of EPG and electrolaryngography (Abberton & Fourcin, 1997), with an acoustic recording made simultaneously. The speech items consisted of $C_1V_1C_2V_2$ disyllabic (two-mora) words. In order to see the effects of accent nucleus and consonant voicing, various combinations of fricatives, stops and accent nucleus were used. In this section we limit our observation to the results obtained from the following test words.

<table>
<thead>
<tr>
<th>DECOICEABLE</th>
<th>C2=VOICED CONSONANTS</th>
<th>ACCENTED</th>
</tr>
</thead>
<tbody>
<tr>
<td>(19) a. [su,tj,i] 'plough'</td>
<td>b. [su,tj,i-sui,ije,i] 'cedar'</td>
<td>c. *[su,tj,i]</td>
</tr>
<tr>
<td>(20) a. *[k3tj,sui]</td>
<td>b. [k3tj,zui] 'injury'</td>
<td>c. *[k3tj,sui] 'kiss'</td>
</tr>
</tbody>
</table>

(*=nonsense word)

The constriction places of the contextual consonants are maximally distant within the range that sufficient contact is obtained by EPG: when the consonant [k] is either in the $C_1$ or in the $C_2$ position, the following vowel must be [i], because otherwise complete occlusion is not obtained. For the words in (19a,c) and (20a,c), the high vowels [i] or [u] are placed between voiceless consonants [k, s] with or without an accent nucleus on the first mora. For the words in (19b) and (20b), the second consonants are [g (g)] or [z] without an accent nucleus on the first mora: the voiced velar stop in the intervocalic position (19b) was slightly nasalised as we mentioned in chapter 2. These test words were embedded in a carrier sentence '[moo bakar'ida] (There is/are only ____ now)', repeated six times at normal speed.

In an attempt to characterise the patterns of contact for alveolar fricatives [s, z] and velar stops [k, g(ŋ)], two measures were used: (i) the amount of contact in rows 1-3 for alveolar fricatives, the MAX frame was determined as the EPG frame showing the maximum number of contacted electrodes within those rows; and (ii) the complete central occlusion at row 8 for velar stops. Temporal overlap of the $C_1$ and $C_2$ gestures was analysed for the temporal distance between the following two points. For the fricative-stop sequences in (19), the index of overlap was the temporal interval between the last frame of the maximum contact for the fricative and the first frame of the complete central occlusion for the velar. For the stop-fricative sequences in (20), the index of overlap was the temporal interval between the first frame of the closure release for the velar and the first frame corresponding to the beginning of the friction: the latter frame was specified by acoustic recordings made.
simultaneously. Since the analysis was limited to one speaker the results should be interpreted as preliminary.

We begin by the observation of the fricative-stop sequences. Figures 7-5, 7-6 and 7-7 present the spectrogram, laryngogram (LX contour) and EPG pattern of [sɯᵣkʰᵣi], [sɯᵣ̃nᵣi] and [ˈsɯᵣkʰᵣi]. Tongue contact is represented only for those target words, the section specified by an arrow beneath the LX contour. On the EPG printout the frames showing maximum contact for [s] and complete occlusion for [k] or [g (ŋ)] are indicated by brackets: the two arrows, up and down, indicate the first closure frame for [k] and the last MAX frame for [s] respectively.

It is evident that the temporal interval between the two gestures is shorter in the sequence [sɯᵣkʰᵣi] than in [sɯᵣ̃nᵣi] and [ˈsɯᵣkʰᵣi]: 26.81msec for [sɯᵣkʰᵣi]; 61.29msec for [sɯᵣ̃nᵣi]; and 103.43msec for [ˈsɯᵣkʰᵣi]. For the sequence [sɯᵣkʰᵣi] in Figure 7-5, the complete occlusion for [k] at row 8 is formed within the minimum temporal loss, after the narrow constriction for [s] in rows 1-3 begins to decrease. The variability of this sequential gesture will be discussed later. The spectrogram exhibits only the fricative formant for the sequence [sɯᵣ] and its LX contour shows no activation. These characteristics are considered as indicative of devoiced vowel. Because there is no obvious trace of the high vowel [uᵣ], it is often claimed that the vowel is deleted at the higher level of the speech production process (e.g. Beckman & Shoji, 1984). However, it is clearly shown on the EPG pattern that during the latter half of the maximum narrowing for [s], the tongue post-dorsum reflects the articulatory gesture for [uᵣ] at row 8: this contact cannot be considered as a preparatory activity for the following [k] (see chapter 5). It is supposed that the fricative resonance may reflect the characteristic constriction. These results explicitly reveal that, while the laryngeal gesture degrades the vowel, the lingual gesture generates its typical shape.

For the sequences [sɯᵣ̃nᵣi] and [ˈsɯᵣkʰᵣi], both the effects of consonant voicing and those of accent nucleus clearly manifest the longer time interval between the fricative and stop gestures. It can be seen in Figures 7-6 and 7-7 that the release, or widening, of the fricative constriction at rows 1-3 is followed by the vowel gesture at row 8. There is no overlap between the narrowing gesture for [s] and the complete closure for [k]. It is worth mentioning that in the sequence [sɯᵣ̃nᵣi], the coarticulatory effects of the vowel [uᵣ] were found during the latter half of the maximum narrowing for [s]. This is similar to the sequence [sɯᵣkʰᵣi]. In contrast, in the production of [ˈsɯᵣkʰᵣi], such coarticulatory effects were not observed during the maximum narrowing: the contact at row 8 increased much later in the release phase of [s]. If we assume that the laryngeal and lingual gestures are cohesively activated in the ‘voiced’ vowel, it is possible that the realisation of the accent nucleus requires a relatively independent gesture for the vowel [uᵣ]: the fricative gesture has only a minimal overlap with the following vowel gesture. In contrast, such a cohesive activation is not true for the ‘devoiced’ vowel.
On the EPG pattern the upper brackets indicate the MAX frames for [s]; the lower brackets indicate the frames for the complete closure of [k]; the down-arrow points to the last MAX frame; and the up-arrow points to the first frame of the complete closure.
The sampling rate is 16KHz. On the EPG pattern the upper brackets indicate the MAX frames for [s]; the lower brackets indicate the frames for the complete closure of [ŋ]; the down-arrow points to the last MAX frame; and the up-arrow points to the first frame of the complete closure.
The sampling rate is 16KHz. On the EPG pattern the upper brackets indicate the MAX frames for [s]; the lower brackets indicate the frames for the complete closure of [k]; the down-arrow points to the last MAX frame; and the up-arrow points to the first frame of the complete closure.
Figure 7-8 above presents three productions of the sequence [sušk]. It is evident that the degree of gestural overlap between the fricative and the stop varies considerably. In (a) the release of the maximum narrow constriction for [s] is followed by the complete closure for [k]. This time interval is shortened in (b). In (c), however, the two gestures completely overlap each other: the releasing activity for [s] occurs after the complete closure for [k]. This suggests that there is much freedom for the timing of the following stop consonant, if the earlier activation does not interfere with the preceding fricative consonant.
Turning now to the stop-fricative sequences, the spectrogram with LX contours and EPG patterns are presented for \([k^i,sm]\) in Figure 7-9; partially devoiced \([k^ii,sm]\) in Figure 7-10; \([k^iizu]\) in Figure 7-11; and \('[k^isui]\) in Figure 7-12. The analytic points are indicated on the EPG pattern: the lower bracket covers the frames of the complete closure for \([k]\); the rising arrow points to the last frame of the closure; the falling arrow points to the first frame corresponding to the beginning of the friction; and the upper bracket covers the frame corresponding to the friction midpoint. Acoustic recordings specify the latter two frames.

The stop-fricative sequences generally show results similar to those of the fricative-stop sequences: the time interval between the two gestures was found to be shorter in the devoiced sequence \([k^i,stu]\) (91.94msec) than in the voiced sequences \([k^izuu]\) (118.76msec) and \('[k^isuu]\) (153.24msec). It must be noted that the overall interval is longer in the stop-fricative sequences than in the fricative-stop sequences. This is due to the (weak) aspiration of \([k]\) in the initial position: after the articulatory release of the complete closure there is a moment at which there are no changes of the contact pattern (1 to 3 frames): a similar ‘freezing’ was also observed for affricates (see chapter 5). This freezing of the gesture is presumed to prevent the two consonantal gestures from overlapping freely. The velar stop is palatalised before \([i]\): the tongue medio-dorsum is involved in the constriction formation and the complete closure is effectively fronted. This positioning of the tongue gives rise to the friction after the release burst and manifests the typical formant transition to the following vowel. In contrast, the partially devoiced sequence \([k^ii,stu]\) reveals intermediate characteristics: the time interval is longer than the completely devoiced \([k^i,stu]\) but is shorter than the voiced sequences \([k^izuu]\) and \('[k^isuu]\). This suggests that during the vowel production, the larynx gesture is directly related to the timing of the lingual gesture: the presence of voicing requires a certain amount of time for the tongue activity.

Overall, both the fricative-stop and the stop-fricative sequences exhibit a shorter time interval between the two consonantal gestures when the high vowel is devoiced. It was found that, when vowel devoicing occurs, the timing of the second consonant was less constrained in the fricative-stop sequences than in the stop-fricative sequences: this is presumably due to the presence of (weak) aspiration. It is supposed that this characteristic might be different from language to language. Although the prosodic system is different, similar variation is found in the data reported by Byrd (1994) for American English /#sk-/ in the word-initial position. Also, Gibbon et al. (1993) show that temporal overlap of the sequence /kl/ varies considerably with speakers and across languages. It is therefore reasonable to assume that, even if the vowel undergoes devoicing, the timing of coordinative activities for \([sV\k]\) and \([kV\ys]\) differs from the true consonant clusters \([sk]\) and \([ks]\). This is suggested by the fact that the realisation of a devoiced vowel is largely hidden on the spectrogram but is explicitly reflected on the EPG pattern.
FIGURE 7-9: SPECTROGRAM, LARYNGOGRAM AND EPG PATTERN OF [k][s][u]

The sampling rate is 16KHz. On the EPG pattern the lower brackets indicate the complete closure for [k]; the upper bracket indicates the midpoint of [s]; the up-arrow points to the last frame of the closure; and the down-arrow points to the frame corresponding to the beginning of the friction.
Figure 7-10: Spectrogram, Laryngogram and EPG Pattern of the Partially Devoiced [k̃], [s̃], [ũ]: the sampling rate is 16KHz. On the EPG pattern the lower brackets indicate the complete closure for [k]; the upper bracket indicates the midpoint of [s]; the up-arrow points to the last frame of the closure; and the down-arrow points to the frame corresponding to the beginning of the friction.
The sampling rate is 16KHz. On the EPG pattern the lower brackets indicate the complete closure for [k]; the upper bracket indicates the midpoint of [s]; the up-arrow points to the last frame of the closure; and the down-arrow points to the frame corresponding to the beginning of the friction.
Figure 7-12: Spectrogram, Laryngogram and EPG Pattern of [k]isuu]
7.4.2.2 $\text{[ç]}$-$\text{[ç]}$ Variability and Moraic Consonant Formation in Connected Speech

We have discussed the characteristics of laryngeal gestures and have observed those of lingual gestures during the production of the devoiced syllables. Based on the observations above we consider two phenomena involving vowel devoicing: $[\text{ç}]-[\text{ç}]$ variability; and moraic consonant formation in connected speech. We will show that a parametric approach can account for the instability of the phonetic forms in a unified way.

It has been mentioned that the pronunciation of a palatal fricative $[\text{ç}]$ is frequently replaced by an alveolopalatal fricative $[\text{ç}]$: the reverse process is very rare (e.g. Akamatsu, 1997; Kawakami, 1977; Sakuma, 1929; Vance, 1987). Edwards (1903: 37) describes this confusion as follows:

> Les sons (s), (s'), (s''), (s) se confondent assez souvent en japonais; de plus, (s) devant (i) se confond avec (ç) en devenant (ç'), son qu'on entend quelquefois en allemand, au lieu de la forme régulière çemi: (Chemie).

Voici quelques exemples qui montrent la difficulté de déterminer les limites de ce groupe de sons en japonais (la première forme est la plus usuelle, les autres sont des variants). çi (feu), çi et quelquefois presque jì; çto (personne), ç'to, ç'to samisej (instrument de musique), çamisej; ç'ite (étendre), ç'ite; s(u,)kôjì (peu) jkôjì; osorojsa (crainte) osoros'sa; jimbajî (pont voisin de la gare principale de Tokyo) j'imbajî et rarement sjimbajî; ç(i)tōtsù (un) j'tōtsù; 3u:ji,tjî (dix-sept), 3u:çi,ti.

litoyo (1975: 208) states that this phenomenon is readily observable all over the Kanto area but is restricted to the pronunciation of particular words (my translation). However, litoyo fails to consider the conditions for eliminating those lexical items. When we observe the examples listed by Edwards, we notice two phonetic factors involved in the $[\text{ç}]-[\text{ç}]$ variability: one is vowel devoicing; and the other is that the following consonant is mostly limited to the alveolar stop $[t]$; the fluctuation does occur in other environments such as in $'[\text{çibatçi}]-+[\text{çibatçi}]$ ‘charcoal’ brazier’ and $[\text{çi,b'i'ja}]-+[\text{çi,b'i'ja}]$ ‘place name in Tokyo’.

There is a potential condition for the $[\text{ç}]-[\text{ç}]$ variability to take place before $[t]$ in particular. Figure 7-13 demonstrates the spectrogram with LX contour and EPG patterns of $[\text{çi, to}]$ ‘man or person’. The contact increases postero-anteriorily for the production of the first mora $[çj]$. It is difficult to distinguish the palatal gesture for $[ç]$ from that for $[i]$, although the contact degree, as seen in chapter 5, is greater in the consonant. Vowel devoicing is evident from the spectrogram and LX contour. Because of the carryover effects of the preceding palatal gesture, the amount of contact for $[t]$ substantially increases particularly in the posterior regions. This contact decreases during the latter half of the complete closure, in order to prepare for the following vowel $[o]$. Based on the articulatory and coarticulatory characteristics of $[ç]$ and vowel devoicing, it is possible to speculate how $[ç]$ is derived from $[ç]$ in front of the alveolar consonant.
Given that the tip/blade gesture for [t] is activated earlier with the laryngeal gesture remaining unactivated (i.e. voiceless), the transitional narrow constriction emerges in the anterior regions in the course of attaining the complete closure for [t]. Consequently it produces the friction noise that is likely to be perceived as [ç]. This explanation is compatible with the results of our EPG experiments. On the one hand, the activity of the tongue tip/blade during the production of [ç] is unconstrained relative to the tongue dorsum: the interdependency between the two components is largely linear (see chapter 5). This gives much freedom for the tongue tip/blade to coarticulate with the following alveolar consonant. Furthermore, when the intermediate vowel is devoiced, the two consonantal gestures, namely [ç] and [t], tend to be executed with minimum temporal loss. On the other hand, the tongue tip/blade necessarily passes through the constriction ‘place’ for [ç]. Even if [ç] is successfully
produced as shown in Figure 7-13, the transitional movement involves a pattern similar to that of [ɛ]. As discussed above, devoicing increases the degree of coproduction of the two consonantal gestures. In addition, the timing of the following stop gesture, relative to the preceding fricative gesture, varies considerably (see Figure 7-8). Furthermore, the articulatory movement of [t] is affected in its positioning, because the consonant [ç] (and devoiced [j]) require the raising gesture of the tongue dorsum. This is clearly shown in the EPG pattern in Figure 7-14 below. The lack of the frontmost contact (row 1) implies that the tip is directed downwards. This suggests that the centre of the articulatory gravity for [t] is slightly further back in the production of [çj,to]. It seems that these are sufficient conditions for the [ç]-[ɛ] variability.

The phenomenon of [ç]-[ɛ] variability cannot be accounted for as long as vowel devoicing is seen simply as a biomechanical (mis)alignment of the laryngeal gesture. In (21a,b) the overall process of gestural overlap between supralaryngeal gestures and that between laryngeal gestures is modelled as changes in the gestural scores.

**Figure 7-14:** EPG Pattern of [çj,to]
Another phonetic process involving vowel devoicing is moraic consonant formation found in connected speech (examples are taken from Amanuma et al. (1978); phonetic transcriptions are mine).


(22b) [e,k'-'ika]—[e,k'ka] ‘liquefaction’, [o,k'ikae]—[ok,kae] ‘exchange’, ['ik'ik'i]—['ikk'ik'i] ‘coming and going’, [ka,k'i'kesu]—[kak,kesu] ‘vanished’, [na,kuku]—[nak,ko] ‘crying child’

The pronunciations in (22a) are largely lexicalised and could be explained by morphological process, rather than phonetic implementation. In contrast, the phonetic forms in (22b) show that the CVC syllable with vowel devoicing tends to be realised as the CC sequence without a vowel in the medial position. Although Amanuma et al. does not admit this possibility, similar examples are easily found in everyday conversations: [ta,ico'kuk'in]—[ta,ico'kokk'in] ‘retirement money’, [sa,ŋ'kaku'kee]—[sa,ŋ'kakkee] ‘triangle’, [de,ŋ'ki'ka]—[de,ŋ'kk'ka] ‘electronic section’; and [ne,tsu'sama]—[ne,(t)s'sama] ‘antifebrile’, [wu,su'sum'ire]—[wu,s'sum'ire] ‘light violet (colour)’, [ja,siu 'su'ru'me]—[ja,s'su'ru'me] ‘cheap dried squid’, [a,siu 'su'n'uni ja,ri'masuy]—[a,s'su'n'uni ja,ri'masuy] ‘I do it first thing tomorrow’, [ka,ci i'baru]—[ka,ci 'baru] ‘unwilling to loan someone’, [ni,ci'ci'z'z'u'ku]—[ni,ci'ci'z'z'u'ku] ‘place name, west Shinjuku’.
The characteristics of this phenomenon are summarised by three points: (i) the consonants before and after the potentially devoiced vowel must be homorganic, i.e. heterorganic clusters are phonotactically impermissible; (ii) the phenomenon mostly occurs when the consonants are voiceless velar [k]; and (iii) the syllable structure is converted from CVC to CC. For this characteristic it is worth noting that, while the phonetic process decreases the number of the syllables of a given word by one, the number of the moras remains the same. This is illustrated in (23) using the example of the word [sa,ŋ'kakwee] ‘triangle’.

(23) [s a  ka  k u  k e  e ] → [s a  ŋ  'k a  k  k e  e ]

The word [sa,ŋ'kakwee] originally consists of four syllables. The scenario based on the segmental analysis requires two steps. The devoiced vowel [u] in the third syllable must be deleted at a higher level and the remaining onset consonant [k] is re-syllabified to the coda position of the second syllable [ka]: it becomes a closed syllable [kak]. Although this scenario is satisfactory from the descriptive point of view, it is weak and implausible in that re-syllabification merely creates a permissible syllable structure; it does not prove the production of a phonetically long consonant: we can find a link between devoicing and re-syllabification, but not a direct link between the two operations and the realisation of a moraic consonant. The question is how vowel devoicing produces a long consonant. The answer that a re-syllabified coda consonant forms a geminate with the onset consonant of the following syllable is phonologically correct, but phonetically unsatisfactory. In addition, the moraic consonant formation is not always applied categorically; we must describe gradient aspects of the phenomenon.

Instead of the two rules, devoicing and re-syllabification, we propose that the moraic consonant formation in connected speech is characterised as systematic changes in the gestural parameters involved. The fluctuation between [sa,ŋ'kakwee] and [sa,ŋ'kakkee], for instance, can be modelled as the gradual modification of the tongue dorsum gesture and the laryngeal gesture. This is represented in hypothesised gestural scores and articulatory trajectories in (23). The basic score is shown in (23a). Given that the control parameters are specified for each gesture separately, there are two opening gestures of the glottis corresponding to the voiceless velar stops. When the high vowel [u] undergoes devoicing the sequence [ku] is produced by a single glottal opening gesture (e.g. Yoshioka et al. 1982b). This process is represented in the gestural score shown in (23b). Notice here that the lingual
gestures associated with voiceless velars tend to be coarticulated within the minimum temporal interval. Because of this tendency, the three underlyingly independent gestures of the tongue dorsum coalesce (or re-organised, Browman & Goldstein, 1991) into one single dorsum movement by lengthening the complete articulatory closure for the first velar stop. The illustration (23c) describes the realisation of a moraic obstruent: the lip gesture [+compressed] for [u] may or may not be retained. A similar explanation is possible for the cases with the fricatives: [a,su 'sunuji ja,tiimasu]→ [a,s'sunuji ja,tiimasu] 'I do it first thing tomorrow' and [ni,ci'çiiz'ukui]→ [ni,c'çiiz'ukui] 'place name, west Shinjuku'.

7.4.3 Summary
Vowel devoicing in Japanese has been characterised in terms of the systematic articulatory processes of laryngeal and lingual gestures. We discussed a gestural overlap analysis (Jun & Beckman, 1993) by reviewing previous experimental research into the laryngeal activity during the production of the voiceless syllables. We pointed out that the analysis was unsatisfactory in that devoicing was considered as the consequence of biomechanical
constraints; and that only the opening gesture of the glottis was taken into account. This view, that vowel devoicing is a laryngeal phenomenon, is exaggerated unnecessarily, not only in a gestural overlap analysis, but also in the modelling proposed in previous studies. Devoicing should be considered as an oral-laryngeal coarticulation.

We proposed that vowel devoicing and related phenomena were modelled in a unified way by considering the events both in the larynx and in the oral cavity. We showed the results of our experiment using the combined technique of EPG and laryngography. Major findings were: (i) when the vowel underwent devoicing, the two gestures in both fricative-stop and stop-fricative sequences were activated within a minimal temporal interval; (ii) in fricative-stop sequences the stop gesture was relatively unconstrained in timing and varied considerably; (iii) in stop-fricative sequences the releasing gesture and the (weak) aspiration prevented the two gestures from coarticulating freely; (iv) the partially devoiced sequences revealed that voicing in the vowel production required a certain amount of time for the lingual gesture to maintain a particular position and shape. Based on these results, together with previous studies of the laryngeal movement, we discussed the two related phenomena: [ç]-[ç] variability; and moraic consonant formation in connected speech. We demonstrated that these processes could be understood only if we considered vowel devoicing as an oral-laryngeal coarticulation. We generally support the gestural overlap analysis of vowel devoicing in Japanese not only as the coproduction of laryngeal gestures but also as that of lingual gestures. This view can further be explained as an example of vowel shortening, the coproduction process of articulatory gestures assumed by Fowler (1981).

7.5 Conclusion
In this chapter we discussed the four phonetic and phonological processes: /r/ variability in child and adult phonology; CyV variability in standard Japanese; de-palatalisation in the regional accents; and vowel devoicing, in particular, [ç]-[ç] variability and moraic consonant formation in connected speech. We showed how these gradient and optional phenomena were characterised in terms of the systematic articulatory controls underlying them. In contrast to our view that we have described above, the explanations might be derived from the traditional rule-based and feature-based approach in which timing control is extrinsically assigned. However, it is questionable how such an account is workable for the three seemingly unrelated processes. We conclude by comparing some aspects of our parametric analysis with those of the traditional symbol-based analysis.

There is a crucial difference between the analysis proposed above and that proposed by the traditional approach. The property of speech that we attempted to characterise is the systematic pattern of articulatory gestures in the vocal tract. In contrast, what is characterised by the traditional approach are the abstract elements that are assumed to have a link to the
sound-producing movements. Even though the two approaches do not share an identical analytical object, they share some issues in common, since they both share the same assumption that there is something invariant in the process of speech production.

There is evidence that different metrics for measuring distinctive properties are necessary. We proposed that in the production of [r] the tongue dorsum gesture should be specified as having the attribute [+suppressed] to facilitate the realisation of the tip/blade gesture. The tap contrasts with [d] in this attribute, as well as the rapid movement of the tip gesture. This proposal gave us a plausible explanation for the acquisition of the Japanese [r] by a child. However, if we assume, with Browman & Goldstein (1989, 1995), that speech production is a direct process, the units of which are articulatory gestures, then the two posited properties of [r] must be specified in the motor representation.

To the expected objection that the double specification is redundant, we reply that: (i) there is evidence that the rapidity of the tongue tip/blade is not always distinctive, for example, the /-Nr-/ sequence in [ka,nroku] ‘dignity’ where the complete occlusion by the tip/blade is as long as that in [ka,ntoku] ‘supervisor’ and [ka,ndoku] ‘documents’ (see the spectrograms and the EPG patterns in Figures 2-3 and 2-4 in chapter 2); and (ii) our specification describes the physical properties that are presumably required for the linguistically contrastive gestures. Although Browman & Goldstein (1989, 1995), as mentioned before, appear to use the metric for the distinctiveness in the same way as traditional feature-based phonological theories, its correctness at the physical level is rarely attested for lingual gestures in particular. Hence, the motor representation of the articulator in question becomes, to some extent, a sophisticated restatement of the feature specification. The validity of the gestural contents of the traditional features must be examined carefully. In addition, we should examine the compatibility of the underspecification theory with the motor representation.

The two kinds of the descriptors, as mentioned above, differ in nature. The gestures describe the abstract patterns of the movements in the vocal tract, whereas the features describe the attributes of the movements (e.g. Clements, 1992). This distinction is certainly related to the ability to explain the system of the sound structure. The advantage of the gestural analysis is that it enables us to approach speech activities in the vocal tract more directly. It is evident that the gradient and optional nature of the phonetic processes is better described by the parametric framework, rather than the rule or feature-based framework. Given the concept of coarticulation as gestural overlap, the gradient nature of synchronic sound changes and the developmental aspects of child phonology are explained as the progressive modification of the articulatory parameters involved; not a sequential shift of the

---

16 I thank John Wells for drawing my attention to this point.
17 Ladefoged (1990) criticises the same point.
sound classes. Furthermore, the language-specific articulatory settings are explained in terms of particular values of the given articulatory parameters, and not as a simple filter which modulates the values of the features. As well as a single mechanism of the changes in the magnitude and timing of gestural overlap being assumed, the concept of lexicalisation, or regularisation, is necessary to distinguish one phenomenon from the other. In this study we simply assumed that lexicalisation implies acquisition in child phonology, while it explains pronunciation habits in the synchronic sound changes of the given accents.

Finally, it is uncertain whether the phonological (i.e. cognitive) representation and the phonetic (i.e. physical) representation are isomorphic, as claimed by Browman & Goldstein (1989, 1995). We demonstrated that various phenomena in Japanese are modelled by assuming a single mechanism, the degree of gestural overlap. This does not completely prove the idea that there are no intermediate levels between the two representations; the results show articulatory evidence for the preconditions and/or actuation of the phonetic processes. Our analysis of /r/ variability in the children's speech and in the regional accents implies that the observed patterns are fundamentally relevant to the 'motor' representation. It is unrealistic to assume that: children who produce incorrect [r]s have not yet acquired phonological contrasts; and adults who use [r] for [d] have lost phonological contrasts. However, it is plausible to assume that the children are in the process of acquiring the movement pattern of [r], while the adults have the knowledge of the accent-specific tongue settings, or simply misalign the timing (in the case of the standard accent). This suggests that the motor representation is, to some extent, independent from the phonological representation. It seems reasonable to conclude that a parametric approach offers a useful technique to describe the activities in the vocal tract systematically: the cognitive aspect of the motor representation remains to be proved.
Chapter 8

Conclusions

“Speech is rather a set of movements made audible than a set of sounds produced by movements.”
R.H. Stetson (1951: 33)

The fundamental aim of research into speech production is to provide a detailed description of dynamic articulatory patterns and to uncover principles of coordinative activities in the vocal tract. The explanation of the phonological system of a language must be established from a qualitative and quantitative empirical basis. This study has attempted to achieve this aim within the limitations of the spatio-temporal aspects of lingual articulations, analysed using the EPG technique. We conclude this study by summarising the findings reported in each of the previous chapters. We then describe some implications for further research.

A parametric approach has been adopted in this study. As discussed in chapter 1, the parametric approach differs from the traditional approach in that the production of speech is seen as continuous, consisting of composite movements in the vocal tract, rather than a sequence of discrete postures of individual articulators. For the identification of the spatio-temporal aspects of lingual articulations, the two-component model of tongue physiology was assumed. The important implication of this is that it enables us to characterise the lingual articulation as coordinative patterns between the tongue tip/blade and the dorsum, rather than simply as the primary constriction between the palate and the tongue. The concepts of coarticulation and of articulatory settings (Honikman, 1964) accord well with the parameterisation of lingual activities, since the two concepts are physiological in nature.

In chapter 2 a survey of phonological and articulatory characteristics of the Japanese sound system was used to illustrate the parametric modelling, relating phonological structure to articulatory organisation. Various segmental processes were shown in terms of the systematic articulatory controls underlying them. For instance: the lip gesture of [ui] was characterised as having compressed lips; diphthongs were explained in terms of the degree of gestural overlap; and the Japanese mora was described as an ‘articulatory syllable’ (Stetson, 1951). It was understood that there are complex and close relationships between the phonological system of the language and its articulatory organisation. Thus, understanding of the physiological systems is undoubtedly necessary for the linguistic modelling of speech production and spoken language.

The specific data type that we obtained from the EPG technique enabled us to examine various spatio-temporal parameters during the production of given phonetic segments as a function of vowel context. Data of the VCV (vowel-consonant-vowel) sequences was obtained from two Japanese speakers. A particular objective was to extract
EPG correlates of articulatory parameters relevant to the realisation of ten consonantal segments \([t, d, n, r, s, c, ç, ts, tc]\); two vocoids \([i] \text{ and } [j]\); four palatal(ised) segments \([n, ç, r^i, k^i]\); and vowel devoicing. Our strategy was to infer the association of those lingual activities with factors of the articulatory control mechanisms. We shall discuss the results of the experiments in this study from four points of view: (i) the language-specific tongue settings; (ii) tongue dorsum coarticulation; (iii) intergestural coordination and timing in palatalisation; and (iv) tongue dynamics and phonology.

We begin by considering the tongue settings of Japanese. The important consideration is the variability of the pattern of tongue-palate contact for the two speakers. Edwards (1903) observes that the tongue settings of Japanese are 'dentalised and palatalised laminal articulation'. Although the number of informants is small, it is possible to ask whether a speaker regularly uses a certain type of articulation. In chapter 4 we found that: (i) the consonants \([t, d, n]\) were mostly apicolaminal on the tongue and dentalveolar on the palate; (ii) \([r]\) involved the tip and very anterior part of the blade, and often exhibited a larger amount of tongue contact than \([d]\); and (iii) the Japanese \([n]\) was constricted slightly further forwards than the Catalan \([n]\) and was frequently accompanied by tip closure. The evidence can be interpreted as a sign of the frontal anchoring of the tongue. In general, Edward's statement has some truth. However, we also found that the production of the above consonants varies considerably; and that the realisation of the alveolar fricative \([s]\) is typically different for each speaker; apical and laminal (chapter 5). Thus, it is impossible to consider one particular articulation as typical for the language in question (Dart, 1991). Yet this does not necessarily negate the idea of articulatory settings.

The gestural analysis of the pronunciation habits in the regional accents (chapter 7) demonstrated that the phonetic forms in non-standard accents were systematic deviations from the standard accent. We showed the example of different phonetic processes in the same phonetic environment: CyV variability in standard Japanese (e.g. \(['c^w\text{u(d)}\text{ziN}\] 'husband') and de-palatalisation in south Tohoku accent (e.g. \([(c^w\text{u(d)}\text{ziN})\rightarrow[(s\text{d}iz\text{üN})]. Instead of assuming the settings as a filter, we incorporated the posited tongue settings of the south Tohoku area into the gestural specification, or the gestural score. Our analysis is a reasonable representation of the parametric nature of the variability of the tongue settings between the two particular accents. We may therefore conclude that the articulatory settings are real objects of a language's sound structure, and that they constitute an integral part of the phonetic representation.

The results of the EPG experiments were presented in chapters 4 to 6. For the consonantal segments \([t, d, n, r, s, ç, ts, tc]\), qualitative and quantitative analyses were performed to identify the basic articulatory features; the degree of interdependency between the two components of the tongue; and the degree of coarticulatory effects along with their
temporal extent. Of particular importance was the identification of the articulatory properties that are unexplained by the feature specification.

The interdependency between the two functional divisions of the tongue (or coupling) was found to be the important parameter for the coronal stop consonants \([t, d, n, \mathfrak{n}, r]\) and the fricatives and affricates \([s, \mathfrak{s}, \mathfrak{c}, ts, t\mathfrak{c}]\). The results of the experiments in chapter 4 demonstrated that the dorsum involvement decreases in the order \([\mathfrak{n}] > [t] > [d] > [r]\) (in the full-context comparison); and \([\mathfrak{n}] = [t] > [n] > [d] > [r]\) (in the partial context comparison, where \(V_2 = /a/ \text{ and } /u/\)). These consonant-specific degrees of coupling correlated with the extent to which the V-to-C coarticulatory effects are allowed, or resisted. Since these coronal stops involve the tip/blade component as the primary articulator, the different manifestations of the dorsum activity during articulatory closure are considered as self-maintained. It follows from this that a speaker actively controls, not only one tongue region that forms the primary constriction, but also the other region that supports the primary constriction.

There is evidence that the linguistically contrastive characteristics of a certain consonant are created, not only by the independent activity of one component, but also by the supportive, or non-interfering, movement of the other component. The consonant \([r]\) exhibited \(V\)-to-\(C\) coarticulatory effects substantially less than the other coronal stops: the raising gesture of the tongue dorsum appears to be tightly constrained. This implies that the degree of the \(V\)-to-\(C\) coarticulatory effects varies, not only with the involvement of the dorsum in the constriction formation, but also with the manner of articulation.

The voiceless fricatives and affricates demonstrated the coupling effects similar to the coronal stops. In the experiments reported in chapter 5 the fricatives \([s, \mathfrak{s}, \mathfrak{c}]\) generally showed higher degrees of interdependency between the tongue tip/blade and dorsum. It was found that this characteristic was closely interconnected with the groove width and constriction location. We addressed the questions of whether there was a correlation between the groove width and constriction location for \([s, \mathfrak{s}, \mathfrak{c}]\). A tolerable value of regression analysis was obtained, suggesting a linear correlation between the two parameters. However, closer examination of the contact data indicated that, while the channel width was constantly maintained within each group of fricative, it varies considerably across speakers and contexts. This is particularly noticeable for \([s, \mathfrak{s}]\). In contrast, the constriction locations of the three fricatives were preserved without fail. As the spectrogram suggested, different constriction locations are supposed to exhibit different sizes of front cavity. Therefore, it can safely be stated that the precise distinction of the groove width might not be important when contrasting the constriction location of the fricatives.

A more consistent constriction location can be related to a more stable overall tongue shape. Furthermore, this effectively limits the \(V\)-to-\(C\) coarticulatory effects upon the fricatives. The EPG evidence uncovered the fact that the three consonants exhibit
coarticulatory effects in the different regions: in the posterior region for [s]; the anterior region for [ç]; but neither region for [ç]. These coarticulatory trends support the claim that the tongue shape is important in the production of the fricatives. Consequently, the fricatives exhibited a higher interdependency between the two components of the tongue. The fricative elements of the affricates [s, ç] also exhibited the same trends as those of the corresponding independent fricatives.

One exception to this trend was the palatal fricative [ç], for which a fairly high value of linear correlation was obtained: a proportional increase was found in the amount of contact in the front and central regions. This contrasts with Recasens’ monotonic-inverse hypothesis (1997) that the degree of the coarticulatory effects decreases as the involvement of the dorsum increases. The consonant [ç], even though it involved the medio-dorsum in the narrowing constriction, exhibited significant variability in the contact pattern. One possible source of variability is the manner feature. Another possibility might be the effects of the tongue settings. Recall here that the tap showed a higher degree of coarticulatory resistance, even though the tongue dorsum was not particularly involved in the formation of the constriction. The characteristics of these two articulations should be taken seriously for modelling the lingual gestures.

Each consonant has its own distinctive values for each parameter. However, when the consonants are considered as a set, their contrasting values of the given parameters generally correlate to form a continuum. Regression analysis was used to infer the factors that underlie the unity of [t, d, n, ñ, r] in chapter 4. It was demonstrated that: (i) [t, d] and [n] were similar in closure-to-consonantal duration, but different in lingual configuration and intergestural coordination; (ii) the consonants [r], [n] and [t] are alike in their pattern of intergestural coordination, but their control parameters were different for each consonant; (iii) [d] resembles [r] in the degree of front contact and in tongue configuration, but the two consonants differ in terms of closure-to-consonantal duration and intergestural coordination pattern. These results lead us to believe that the consonants are organised systematically as a class.

The crucial characteristic of the affricates [ts, te] is that the two elements of each affricate are ‘closely knitted’ (Gimson, 1980). It is clear from the results reported in chapter 5 that the articulatory correlates of this characteristic are that: the two elements are homorganic; the preparatory activity of the fricative element occurs during the stop closure; and that the vowel-dependent coarticulatory effects are similar between the two elements when they form an affricate, rather than when they are independent segments. Thus, we should conclude that affricate articulation is a unitary process, not the compounding of the two independent articulations.

Both anticipatory and carryover effects were observed for the consonantal segments.
[t, d, n, j, r, s, c, č, ts, tč] in chapters 4 and 5. The results show that carryover effects are systematic and extensive in the sense that they influence the amount of contact in all three regions: front, central and back. In other words, the height of the preceding vowel determines the positioning of the tongue dorsum during complete occlusion or narrowing. It seems that these vowel-dependent effects contradict the common expectation that the basic syllable structure is the sequence CV, and that Japanese is a mora-timed language. In other words, the anticipatory effects of the following vowel are expected to be extensive and systematic. However, there are two difficulties. First, the preference for the CV syllable does not characterise Japanese as a mora-timed language: phonologically the CV sequence can be considered as both a syllable and a mora. Second, the anticipatory effects of the following vowel do occur: the effects are systematic but not extensive, in the sense that the changes in the degree of contact are limited to the particular region.

The EPG data for [t] revealed that carryover coarticulation forms partly a co-production strategy deliberately used by a speaker. There is a common assumption that the carryover effects are mechanical in nature. However, from the data it can be seen that during the production of the tap, there is a dorsum lowering gesture which occurs between the preceding vowel and [r] before its complete closure is attained: the other stop consonants generally keep the tongue position of the preceding vowel at the onset of the closure.

In chapter 6 the articulatory realisation of palatalisation [p, č, r, k] was investigated in detail. We discussed the inadequacy of the common definition of palatalisation which states that there are no palatalised alveolopalatals and palatais. Then we formed some specific proposals concerning the distinctive feature specifications and the definition of the phonetic process of palatalisation. We then defined palatalisation as a specialised use of the raising gesture of the tongue dorsum. Also, we argued that the issue of articulatory complexity is not a question of whether the palatal articulation is realised in a complex way or in a simplex way; but rather that what is at issue is why and how various degrees of intergestural coordination occur during the production of the palatal(ised) segments.

The EPG evidence revealed that there are two kinds of tongue dorsum raising gesture, one [i]-like and one [j]-like; and that there are two kinds of timing for the palatalisation. It was found that these two characteristics corresponded to phonological contrasts, the syllable types Ci and CyV. In contrast, the contact configuration at the MAX point was similar between the two syllables. This was confirmed by the analysis of the centre of articulatory gravity (COG). These results indicated that the spatial configurations, similar between the two syllable types, were derived from the different temporal coordinations between the two components of the tongue. We tested the degree of articulatory complexity by comparing the amount of the central region contact in [p, č, r, k] with that of [i, j]. The results indicated that all the consonants were complex in that the amounts of contact in the [p,
ç, r], k] tokens were the same as, or greater than, those of [i, j]. We also discovered three kinds of gestural overlap for the coordination between the two tongue regions involved: blending, sequencing and fronting. We reached the conclusion that, if we assume two timing mechanisms specific to the syllable type, two kinds of palatalisation were represented. We must note that there is a tendency for the alveolopalatal fricative [ç] in Ci and CyV syllables to fail to attain the posited timing distinction.

In chapter 7, based on the results of the EPG experiments reported in the previous chapters, we discussed the patterning of the articulatory gestures underlying the three different phenomena in Japanese: /r/ variability in child and adult phonology; synchronic sound changes (CyV variability and de-palatalisation); and vowel devoicing ([ç]-[ç] variability and moraic consonant formation in connected speech). A preliminary result of the EPG experiment was presented for vowel devoicing. Assuming that coarticulation is realised as overlapping and blending of articulatory gestures, we showed that the gradient and optional nature of these phenomena was consistently represented. We suggested that the motor representation and the phonological representation were, to a certain degree, independent.

The results of the explorations in this study have raised a number of issues that may require further research. It has been shown that there are language-specific trends in the realisation of the lingual consonants on the one hand, and a great deal of interspeaker variation on the other. This fact must be taken into account when the process of speech production is modelled both within and across languages. More investigation is necessary in order to explain and generalise the nature of phonetic differences and similarities in the individual languages and the idiosyncratic controls of the lingual gestures. Such a study could help to uncover the importance of these factors both in the theory of phonetics and in that of phonology.

The experimental settings in this study used a set of VCV sequences for all the target articulations, produced at the self-selected normal speaking rate. The results obtained from this basic design need to be confirmed by further experiments that take into consideration other linguistic and non-linguistic variables such as different speaking rates and the length of test words. Such extensions could help to determine the validity of the two timing mechanisms proposed for the articulatory realisation of palatalisation.

The main goals of this study were to determine the basic articulatory features of the lingual consonants in Japanese; to give a detailed account of the articulatory and linguistic nature of palatalisation in Japanese; and to explore the relationship between articulatory gestures and linguistic structure. It is hoped that the descriptive and experimental investigations presented in this study make modest contributions to expanding our knowledge of spoken Japanese, as well as to our understanding of language as one of our cognitive behaviours.
Bibliography


Barry, M.C. (1992) 'Palatalisation, assimilation and gestural weakening in connected speech.' Speech Communication, 11, 393-400.


Delattre, P. and D.C. Freeman. (1968) 'A Dialect Study of American R's by X-ray Motion Picture.' Linguistics, 44, 29-68.


Firth, J.R. (1948b) 'Sounds and Prosodies.' *Transactions of the Philological Society*, 127-152.


358


Houde, R.A. (1968) *A Study of Tongue Body Motion During Selected Speech Sounds*. Speech Communications Research Laboratory, Santa Barbara, California, SCRL Monograph No.2.


Kiritani, S., S. Sekimoto, and H. Imagawa. (1977b) 'Parameter Description of the Tongue Movements for Vowels.' *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, Faculty of Medicine, University of Tokyo, No. 11, 31-37.


*Phonetica*, 37, 235-252.


Technical Report No. 2 of the Rutgers Centre for Cognitive Science, Rutgers University, ms.


*Journal of Phonetics*, 12, 63-73.

*Phonetica*, 41, 125-139.

*Journal of the Acoustical Society of America*, 76(6), 1624-1635.

*Language and Speech*, 28, 97-114.


*Language and Speech*, 36 (2,3), 213-234.

*Periodic Progress Report, ESPRIT-ACCOR BRA project of the European Community*, vol. 2, 19 pages.

*Catalan Working Papers in Linguistics*, Universitat Autònoma de Barcelona, 3(2), 71-96.


*Journal of Phonetics*, 27, 143-169.


Ushijima, T. and M. Sawashima. (1972) ‘Fiberscopic Observation of Velar Movements during Speech.’ Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, Faculty of Medicine, University of Tokyo, No. 6, 25-38.


Yoshioka, H., A. Löfqvist, and R. Collier. (1982a) 'Laryngeal Adjustments in Dutch Obstruent Production.' *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, Faculty of Medicine, University of Tokyo, No. 16, 27-35.


Yoshioka, H., A. Löfqvist, H. Hirose, and R. Collier. (1986) 'How Voiceless Sound Sequence Are Organised in terms of Glottal Opening Gestures.' *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, Faculty of Medicine, University of Tokyo, No. 20, 55-67.
Appendix 1

Palatalisation in Japanese, Russian and Greek: a descriptive analysis

1. Introduction

This appendix discusses palatalisation in Japanese, Russian and Greek. Given the notion of the ‘correlation of palatalisation’ (Trubetzkoy, 1958/1969), we find interesting similarities among the languages whose family and geographical location are quite different. Palatalisation can be considered as a distinctive opposition between phonological classes. Articulatorily, palatalised and non-palatalised consonants differ from each other with or without a secondary [i]-like posture superimposed on the primary articulation. Acoustically, the two series are distinguished with or without a specific timbre. Thus, palatalisation creates a privative opposition. If this opposition covers the whole part of the consonant system in a language, the correlation is established and it is identified as the proportional opposition to secondary phonetic characteristics. It is, in principle, assumed that the correlation, if any, is organised in a series of places of articulation. However, as Trubetzkoy acknowledges, languages differ in the scope of the correlation and more significantly in the phonetic realisation of palatalisation. Trubetzkoy suggests that Japanese is similar to other languages (e.g. Russian) on the grounds that palatalisation overlays all the series of localisations, to create the combination of palatalised and non-palatalised consonants.

We may place the main focus on how many and what kinds of localisations the secondary series among languages have, since the secondary opposition identifies the proportional opposition in the consonant system. If, however, we look at the phonological contrast in the vowel context, another characteristic can be found in a group of languages that have the correlation of palatalisation in all the localisations. This can simply be examined by comparing Japanese palatalisation and its phonological contrasts with those in other languages. We develop and exemplify our theme by comparing Japanese palatalisation with that of Russian and Greek. These languages are chosen because they have a five-vowel system similar to Japanese, and also because all the localisations in their consonant system contain the secondary palatal series. The descriptive generalisations will be based on referential phonetic work and my informal observations of native speakers of each language.

In discussing the language-specific realisation of phonological contrasts, we cannot avoid facing the problem that phonemic phonology struggled with. It is the multiplicity of possible phonemicisations. Because of its prosodic nature, palatalisation is a typical example of phonemic indeterminacy (e.g. Trager, 1942). Thus, the three languages differ in their phonological representation that assigns palatality to the segment. On the one hand, for
Russian the palatalised consonant analysis, which parallels plain consonants with palatalised consonants, is commonly applied (e.g. Jones & Ward, 1969). On the other hand, the palatal glide analysis, which displays the sequence of C/j/V, has usually been adopted for the consonant system of Japanese and Greek, although all the localisations oppose in palatality (e.g. Hattori, 1950; Koutsoudas, 1962; Pring, 1950). The variability among languages depends on analytic strategies within the phonemic principle as well as practicality for a given language. Such a difference may not affect our analysis. Nevertheless, we should notice that the correlation of palatalisation is not always identifiable from the phonological inventory of a language.

2. Palatalisation in Russian

In Russian the palatalised consonant analysis yields an inventory as follows (Jones & Ward, 1969; 299: the phonetic transcription for palatalised consonants is updated).

![Tabelle](image)

Evidently palatalisation overlays all the localisations except the consonants /g, x, ʧ, j/. The palatalised allophone of velar stop and fricative phonemes /g, x/ occurs before /i, e/ such as /gipkli/ ['supple', /genli/ ['genius', /xitri/ ['cunning', /xeris/ ['sherry'] (examples from Jones & Ward (1969)). We find minimal pairs in each vowel context (examples are taken from Boyanus (1935)).

![Minimalpaare](image)

Jones & Ward (1969; 82) mention that 'palatalised consonants in Russian are frequently followed by a slight 'glide', like a j-sound, and known as an 'off-glide' and 'between a vowel and a following palatalised consonant there is often an 'on-glide', like a
faint suggestion of an i-like vowel glide from the vowel into the consonant'. The presence of the [i]-like glide before and after a palatalised consonant is the characteristic sound quality in Russian (Boyanus, 1935; 41f.). Hattori (1984; 110) describes such a j-like off-glide as particularly long in comparison with that in Japanese. Comparing the pronunciation of /m^asa/ [m^asA] 'meat' in Russian with that of /mjaku/ [m^akul] 'pulse' in Japanese, Hattori mentions that 'it is no exaggeration to transcribe the former as [m^jasA] with an additional symbol [j] to suggest the phonetic quality of such an off-glide' (my translation)\(^1\).

From an acoustic point of view, Sepp (1987) compares palatalisation in Russian with that in Estonian where palatalisation occurs only in /t, s, n, l/ immediately after the vowel in the primary-stressed syllable. Assuming that the slight rise of the second formant frequency (F2) is the main and sufficient acoustic parameter for palatalised consonants as distinguished from non-palatalised consonants, Sepp examines how the point of maximum F2 is manifested in VCV sequences, where the intervocalic consonant is a palatalised [sj]. Sepp (1987; 37) found that in Russian the maximum effect appears at the release of the consonant and the transition to the following vowel. On the contrary, in Estonian the peak occurs at the transition from the preceding vowel to the consonant. This leads Sepp to the idea that the time-varying parameter in question can be considered as prepalatalised (or early palatalisation) and postpalatalised (or late palatalisation).

From an articulatory point of view, the raising or fronting gesture of the tongue dorsum for palatalisation may affect the point of articulation of contextual neighbours. Jones & Ward (1969; 188ff.) symbolise different allophones of a vowel to represent the effect: for instance, /v^esV/ [v^es^] 'all' as opposed to /v^es/ [v^es] 'weight'; another examples are /n^it^/ [n^it^] 'thread', /d^ad^a/ [d^æd^o] 'uncle', /t^ot^a/ [t^æt^o] 'aunt', /t^ol^/ [t^ol^] 'tulle'.

A voiced palatal approximant also reveals a language-specific characteristic. Boyanus (1938; 25) states that, while the approximant may be recognised as a voiced sound, '[i]n Russian this sound is louder and is spoken with more friction than in English'. Words containing /j/ are /majak/ [majak] 'lighthouse', /jes^t/ [jes^t] 'to eat and is', /jest/ [jest] 'eats', and /juk/ [juk] 'south'. The Russian /j/ occurs after a palatalised consonant, which is phonotactically constrained: for instance, /p^ju/ [p^ju] 'I drink', /stat^ja/ [stat^ja] 'article' (Boyanus, 1938, 27).

3. Palatalisation in Greek

Palatalisation in modern standard Greek is quite different from Japanese and Russian. We have the consonant inventory of Greek below (Maddieson, 1984).

---

\(^1\) When I analysed the pronunciation of Russian in the course on Phonological Analysis (1997-1998, University College, London) run by John Wells, I had a similar experience to that described by Hattori (1984). Impressionistically the off-glide in question sounds to me like a high front vowel [i].
Not all the researchers accept the inventory above. There are three controversial points in phonemicisation: (i) the voiced plosive series may be analysed as pre-nasal allophones of the voiceless series (Holton et al., 1997); (ii) two affricates /ts, dz/ may be analysed as a cluster of two independent phonemes (Holton et al., 1997; Koutsoudas, 1962; Sotiropoulos, 1972); and (iii) a palatal glide may be, not an independent phoneme, but an allophone of a high front vowel /i/ (Holton et al., 1997; Pring, 1950). Within an SPF framework, Newton (1972) eliminates these problematic points from the inventory to suggest more abstract underlying segments.

Among the three issues above the third one is particularly important for our discussion. If we accept the inventory in (3) by Maddieson (1984), which is similar to Koutsoudas (1962) and Sotiropoulos (1972) except admitting two affricates, then we can represent palatalised consonants in Greek with the /CjV/ formula, i.e. using the palatal glide analysis. Since all the localisations of Greek non-vocoids have the palatalised counterparts, this treatment has several advantages. It avoids making complex phonotactic statements (Koutsoudas, 1962, 6) and assuming an abstract derivational process for palatalised

---

2) Koutsoudas (1962) and Householder (1964) are aware that the palatal glide analysis raises the problem of interpreting /γ/. There occurs phonemic overlapping between the allophone of /γ/ before front vowels and that of /j/. Thus, there are two possible phonemicisations.

/γi/ or /γi/ [ji] ‘earth’ - 
/γeros/ or /γeros/ [jeros] ‘old man’ -

Before non-front vowels /a, o, u/, the two phonemes contrast each other, these are the cases where different phonemic representations are necessary. We follow the proposal of Koutsoudas (1962) that /γ/ before non-front vowels is represented by a sequence of /γj/.

/γjorvos/ [vro’s] ‘George’ - /vorγos/ [vorγos] ‘fast (masc.)’
/γa’ju/ ['aju] (‘aju’) ‘of the saint (gen.)’ - /pavu/ ['pavu] ‘of the ice’

(All the examples are from my Greek informant)

Householder (1964; 25) mentions that ‘it [a palatal approximant] cannot be a proper phoneme, though it could be a feature or prosody.’
consonants. Furthermore, it enables us to compare the phenomenon cross-linguistically since the same analysis has been taken for palatalisation in Japanese.

The distribution of palatalised and non-palatalised consonants in Greek is asymmetrical. From the description by Koutsoudas (1962; 6), we have three distinct groups of phonological contrasts in palatalisation: (i) non-velars except /t, d, s, θ, n, l, r/; (ii) /t, d, s, z, θ, n, l, r/; and (iii) velars /k, g, x, y/. Whereas palatalised consonants contrast in initial and medial position, in final position only non-palatalised consonants occur (ibid.). Minimal pairs of words containing selected consonants from each group exemplify the phonological contrast. These examples are gathered from my informal observation and interview with one native speaker of modern standard (Athenian) Greek. The phonetic transcriptions are normalised from my impressionistic observations.

As an example of the non-velars except /t, d, s, θ, n, l, r/ (the group (i)), the minimal pairs containing /p/ and /pj/ are illustrated in (4) below.

(4) /pV/ - /pjV/: non-velars except /t, d, s, θ, n, l, r/
/pj/ [pj] '(he) says (subj.)' - /pj/ [pçi] 'drink'
/pes/ [pes] 'say (imper.)' - /pjes/ [pçes] 'drink (imper)'
/'pano/ ['pano] 'on' - /'pjano/ ['pçano] 'I hold'
/po/ [po] 'say (subj.)' - /pj/ [pço] 'I drink'
/'kapu/ ['kapu] 'somewhere' - /'kapçu/ ['kapçu] 'someone’s'

This group of consonants displays the same contrast pattern as it is in Russian. The palatalised series on the right-hand side do not yield the expected phonetic realisations: [pçi], [pço], [pçan], or a bilabial consonant with a simultaneous secondary palatal articulation such as [p^a]. Instead, there is an audible friction between a consonant and a vowel in a /pjV/ sequence. We represent it as a voiceless palatal fricative [ç].

The phonetic realisation of /j/ described above is fairly similar to that in the second group of consonants /t, d, s, θ, n, l, r/ (the group (ii)). Yet, one phonological contrast is lost before the front vowel /i/: the sequence of /ti/ does not have its pair, /tji/.

---

3) Koutsoudas (1962; 6) does not include /n/ and /l/ in the group (ii), but he notes ‘[n] and [l] before /i/ vary freely with their palatalised counterparts’. In other words, palatalisation involving /n/ and /l/ are neutralised before a high front vowel. As we will see below, the same is true for palatalised series with /s/. Notice here that not all the alveolar consonants show such a free variation between palatalised and non-palatalised forms.

4) I thank my Greek informants, George Petropoulos and Ioanna Deimezi, for their patient advice on the pronunciation of Greek.

The phonemic combination of Greek /kj/ is realised as a palatal plosive [c]. Hattori (1984; 84) notes that palatalised consonants found in /kja, kjo, kju/ and /gja, gjo, gju/ in Japanese can be transcribed as [ca, co, cu] and [ja, jo, ju] respectively.

4. Palatalisation and the Types of Phonological Contrasts
We have seen the palatalisation phenomenon in Japanese, Russian and Greek. These languages, as mentioned before, share two characteristics in common: the five-vowel system /i, e, a, o, u/ and the correlation of palatalisation over all the localisations. Trubetzkoy’s conceptualisation allows us to identify similarities between Japanese, Russian and Greek, on the grounds that they organise the correlation of palatalisation in their consonant systems. We have examined Trubetzkoy’s idea a little further, asking how the phonetic property exhibits phonological contrasts in the three languages and what types of contrasts may result. Our

5) George Petropoulos (p.c. March 2001)
observations can be summarised with the vowel diagram below. The shading shows the environment for contrastive palatalisation.

(7) Russian

(8) Japanese

(9) Greek
(9a) Non-velars (except /t, d, s, z, θ, n, l, r/)

(9b) /t, d, s, z, θ, n, l, r/

(9c) Velars /k, g, x, ɣ/

We can now propose an answer to the question that was posed at the beginning of this appendix. Japanese, Russian and Greek exhibit the three different, but interrelated, manifestations of contrastive palatalisation. In Russian the opposition between palatalised and non-palatalised consonants is retained before all the vowels, as seen in (7), while in Japanese it is restricted only to the non-front vowel contexts, as in (8). Assuming that Russian and Japanese are a full contrast and a partial contrast system respectively, we can say that the Greek system merges the two types of phonological contrasts. As diagrammed in (9), the opposition of the secondary palatal articulation varies with the places of articulation: labials (including labio-dentals) are fully contrastive as in (9a), while alveolars and velars are partially contrastive as in (9b,c). The context of the contrastive palatalisation is quite different from that of Russian, despite the fact that the correlation of palatalisation overlays all the localisations in each language. This contrasts with Trubetzkoy’s observation. It appears that, when both the correlation and phonological contrasts are taken into account, the palatalisation system of Japanese is more similar to that of Greek.

---

6) John Wells first suggested this idea to me.
Having examined cross-language differences in phonological contrasts, there is one further point that needs to be clarified for the correlation of palatalisation. From the phonological inventories and representations assumed, we notice that Japanese, Russian and Greek have a semivowel (or semivowels). In Japanese there are /j/ and /w/, while in Russian and Greek there is only one semivowel /j/\(^7\). The phonetic realisation of /j/ varies language-specifically. Since a semivowel /j/ is closely related to the palatalisation phenomenon, it is reasonable to assume that there would be co-occurrence restrictions between /j/ and the correlation of palatalisation. This question is examined by Maddieson (1980; 118), suggesting that there be some tendency for an association between the occurrence of palatal consonants and /j/ in a given phonological inventory. From a sample of 317 languages in the UPSID inventories, Maddieson found that while 'there are three languages in the survey which have palatalised consonants but no /j/ phoneme,' 'true palatalised consonants [i.e. phonemicised as elements in the inventory like Russian]...occur in about 10% of the language in the survey' (Maddieson, 1980; 117). As mentioned before, the correlation of palatalisation is not always reflected in the inventory structure. The palatalised consonant analysis is applied for some languages, while the palatal glide analysis is taken for others. Furthermore, it may be the case that a language is analysed by the three possible analyses including the palatalised vowel analysis. To account for the relationship between the occurrence of /j/ and the correlation of palatalisation, languages like Japanese and Greek should be included. Our limited sample generally agrees with Maddieson's observation.

\(^7\) Maddieson (1980, 113) reports that: (i) 71% of the surveyed languages have both /j/ and /w/; and (ii) 14% languages have only /j/. Some languages (10%) have no semivowel phoneme.