THE PERCEIVED QUALITY OF VOWELS SHOWING FORMANT UNDERSHOOT

Shinichi Tokuma

Department of Phonetics and Linguistics
University College London

Submitted to the University of London
for the degree of Doctor of Philosophy

1996

To Won

Ａは黒、Ｅは白、Ｉは赤、Ｕ緑、Ｏ青よ、母音らよ、
何時の日にかわれ語らばや人知れぬ君らが生い立ちを。

アルトゥール・ランボー　「母音」
堀口大學　訳

A noir, E blanc, I rouge, U vert, O bleu : voyelles,
Je dirai quelque jour vos naissances latentes :

"Voyelles"
par Arthur Rimbaud

Acknowledgements

This PhD dissertation has needed a great deal of assistance on the way to its completion. Regrettably only a few of the people who helped me can be named here.

First of all, I would like to express my sincerest gratitude to my supervisor, Dr Mark Huckvale, University College London. I am extremely grateful for the UNIX / DOS programs which he wrote to control the perceptual experiments for this dissertation.

I am also indebted to other members of staff at Wolfson House; in particular, to Professor Emeritus Adrian Fourcin for his support in the first year of my PhD research, to Professor Stuart Rosen and Dr Andy Faulkner for their statistical support, and to Dave Cushing, John Deeks, and Andrew Simpson for their comments on the grid-test scheme, the quality of the synthesised /CVC/ used in the experiments, and especially for their voluntary participation in the grid-matching experiments. Adrian's encouragement, the statistical help from Stuart and Andy and the patience of Dave, John and Andrew gave momentum to this PhD project.

In Japan, while I was studying for the MA in Linguistics in Sophia University, Professor Tsutomu Sugawara and Professor Keiichi Kojima taught me acoustics and experimental phonetics, which awakened my enthusiasm to pursue this field of research. Now I really appreciate their initial inspiration and their moral support for studying in UK.

Finally, since I could not obtain any grant for my PhD research, I had to rely on my own savings and the financial support from my family. My greatest and cordial thanks are due to my parents, Tominori and Emiko Tokuma, and my grandparents, Ryuichi and Tokiko Tokuma.

September 1996 S.T.

Abstract

Vowels in consonantal contexts are modified from their citation form by neighbouring consonants; in particular, their formant frequencies may not achieve their target values. This phenomenon is called acoustic vowel formant undershoot, and in this study vowels in /CVC/ environments are compared with steady state vowels to investigate the perceived vowel quality change caused by undershoot. Although the phonological perception of vowels showing formant undershoot has been frequently investigated, the phonetic vowel quality change caused by the formant undershoot has received little previous attention.

The main study in this thesis uses a perceptual task, whereby listeners match constant /CVC/ stimuli of /bVb/ or /dVd/ to variable /#V#/ stimuli, using a schematic grid on a PC screen. The grid represents an acoustic vowel diagram, and the subjects change the F1/F2 frequencies of /#V#/ by moving a mouse. The main results of the study show that while subjects referred to the trajectory peak of the /CVC/ stimuli in vowel quality perception, their performance was also affected by the formant trajectory range of the stimuli. When the formant trajectory range was small, they selected a value between the edge and peak frequencies, while they selected a value outside the trajectory range when it was large. Some influences of low F1 frequency values on F2 matching were also found. These demonstrated phenomena are incompatible with existing vowel perception theories, and three modifications to these theories are suggested to explain these results: (1) to incorporate a psychoacoustic process of undershoot compensation; (2) to incorporate the influence of phonological processes on low-level matching; (3) to exploit a single auditory model with multiple stages, which are either open to general auditory processes or specific to speech perception. Proposals are made for further experimental work to choose between these possibilities.

Errata

Page	Line	
21	3	"simple" ---> "a simple"
23	7	"lexicon" ---> "the lexicon"
45	8	"15" ---> "Fifteen"
51	22	"difference" ---> "differences"
56	18	"VISC" ---> "vowel inherent spectral change"
62	16	"and" ---> "and he"
62	21	"three of them" ---> "three of them were"
64	Footnote 16	"experimented on" ---> "presented to"
65	5	"15" ---> "Fifteen"
72	18	"12 subjects" ---> "Twelve subjects"
80	19	"Bark scale" ---> "the Bark scale"
93	20	"test this validity" ---> "make a test of validity"
121	19	"its matched F2 showed F1 matched to" ---> "that its matched F2 also corresponded to"
121	19	"as in /æ/ in both consonantal context and durational environment" --->"in both consonantal context and durational environment as in /æ/"
124	1	"and four vowels" ---> "and the four vowels were"
138	6	"type, and displayed" ---> "type. The results are displayed"
155	4	"A and B, for " ---> "A and B, and for"
183	2	"method to investigate" ---> "method for investigating"
185	17	"order of AB matters" ---> "number of AB permutations"
187	10	"produce the factor [subject] as a main factor nor its interaction by [F2-value]" ---> "show the factor [subject] nor the [subject]*[F2-value] interaction as significant"
187	12	"the interaction of the factor [subject] by [F2-value]" ---> "a significant interaction between [subject] and [F2-value]"
201	20	"dynamicity" ---> "dynamic nature"
206	24	"despite that" ---> "despite the fact that"

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

> The trouble began with the spectrogram and with the recognition that as the cliche has it, phones are not strung together like a beads on a string, but rather ... like a line of eggs passed between rollers.
>
> (Studdert-Kennedy 1981:4)

Anyone who examines the acoustics of spontaneous speech for the first time will notice that its physical character as shown in a sound spectrogram is not what is expected from its orthography or even its phonetic transcription. The changes in phonetic segments in context are such that one is little aided by knowing their acoustic realisations of the phonetic segments uttered in isolation. This acoustic variation of each segment is conditioned by a combination of linguistic and non-linguistic factors such as phonological rules, human articulatory constraints imposed by speech organs, coarticulation with neighbouring sounds, speech-rate, speech-style and so on.

Despite this variability, a listener can readily decode the message from the 'altered' speech signal which contains these modified phonological segments. This process --- how listeners reconstruct the underlying phonological segments from acoustically variable speech --- has been a central problem for speech perception. In this, vowels are no exception, and the apparent variation of the acoustic structure of vowels, realised as changes of duration, F0, intensity and spectrum, is no hindrance to their phonological recognition. Two spectrograms of Figure 1-1 illustrate how vowels in /hV/ context (where they are not influenced by coarticulation with consonants) differ from those in /CVC/. The vowels in spectrograms are both British English /i/, although their durational and spectral difference is prominent.

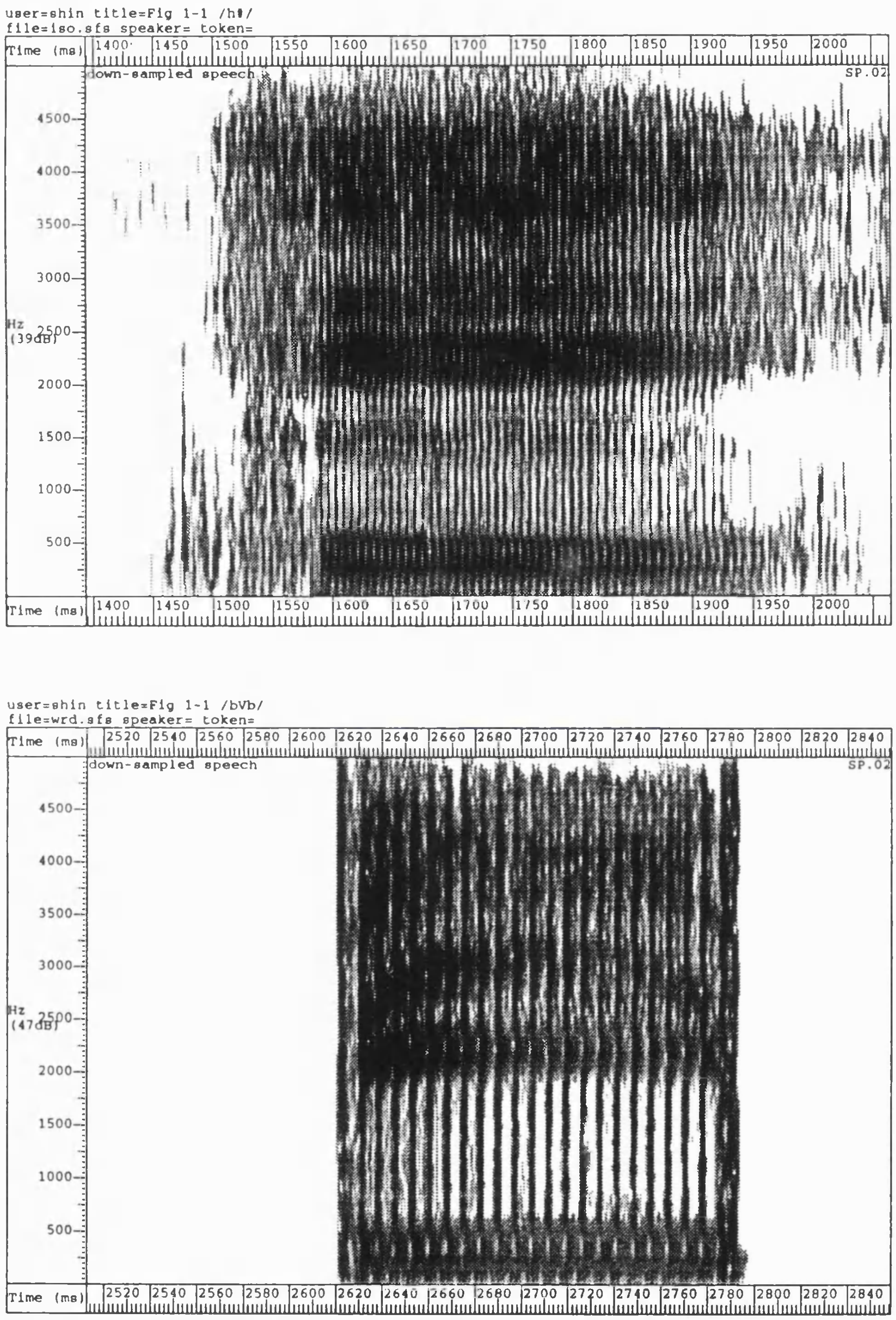


Figure 1-1: Example spectrograms. While a vowel /i:/ under no coarticulatory influence has flat formant trajectories as shown in the upper spectrogram of /hi:/, it shows dynamic formant trajectories when it is coarticulated by the neighbouring consonants, as shown in the lower spectrogram of /bi:b/.

There is, however, an interesting aspect to this phenomenon. Do the phonological segments that are acoustically realised with substantial modification give a different phonetic impression to the listener than those produced in isolation? Specifically, with regard to vowels, how does a listener evaluate the quality of a vowel if its formant frequencies shift considerably from the values obtained when it is uttered in isolation? Suppose a vowel is flanked by two consonants forming a /CVC/ syllable, and its formant trajectory is quasi-parabola shaped, then when listeners evaluate its vowel quality, will they do so with reference to the peak frequency of the parabola, to its trajectory shape, or to some other psychoacoustic parameters?

This problem is beyond the realm of studies on phonological perception, since a given phonological segment of a language permits innumerable acoustic variants within a category defined in terms of its phonological function and these variants can give different auditory impressions.

In this regard, the current study focuses on the perceived phonetic quality of vowels with modified formants. Since it is impossible to investigate all the acoustic parameters involved in the dynamics of vowel formants and to factor out their individual perceptual role, this study concentrates on how the coarticulation of a vowel with neighbouring segments shows different vowel qualities. The study uses a synthetic /CVC/ whose vowel formant values are found from acoustic analysis.

This study deals with formant shifts in vowels, but the field lacks agreement in terminology, although it has been one of the major topics in phonetics. 'Undershoot', 'reduction' and 'coarticulation' are the labels that are frequently used, but their definitions vary from one paper to another. Therefore the issue of terminology for the phenomenon needs to be addressed first.

The task of subjects in an experiment of phonological perception is to identify a sound as a phonological segment in a language --- a task readily manageable for native speakers. On the other hand, an experiment on the perceived phonetic quality of vowels causes difficulties for subjects, since the judgements are fine ones and the number of the stimuli needs to be large to enumerate all the possible combinations of vowel quality. Furthermore there are many different phonetic segments with different acoustic properties in the space between two discrete phonological segments. The current study must address the problem of how to conduct vowel quality matching experiments.

It is commonly acknowledged that low level psychoacoustic processes, phonetic processes and high level lexical/syntactic processes play a role in phonological perception, although it is controversial which of these is more dominant and how they interact. A number of hypotheses have been proposed to explain the phonological perception process, and the results of this study may contribute to its elucidation: in particular, by determining how the psychoacoustic and phonetic processes of vowel quality evaluation are involved in the identification of phonological segments in coarticulated speech.

Moreover, other related studies on phonological perception of a vowel in a consonantal context must be discussed to examine the relationship between the phonological perception of a vowel in /CVC/ and its quality evaluation. A few studies have investigated the phonetic quality perception of a vowel whose formant frequencies shift from those obtained from /#V#/, and their experimental scheme and results will give some useful guidance on the direction of this study. This study should discuss these previous works.

To summarise, the main objectives in this study are:

1) to develop adequate and explanatory terminology to describe vowel formant shifts

2) to review previous studies in order to find their implications for the current

study

3) to devise an appropriate experimental scheme to examine vowel quality perception

4) to discover by experiments how listeners evaluate the quality of vowels in /CVC/ showing formant shift

5) to investigate to what extent the phonetic processes of vowel quality evaluation contribute to the identification of phonological segments in coarticulated speech

The structure of the dissertation follows the main issues above:

Chapter 2 deals with the terminology of vowel formant shifts to give appropriate labels to the phenomenon. It discusses the adequacy of three well-known labels: 'undershoot', 'reduction' and 'coarticulation'.

Previous studies on vowel perception are discussed in Chapter 3, under three headings: acoustic studies, phonological perception studies and phonetic perception studies. Each section relates previous studies to the topic of this dissertation.

Chapter 4 discusses an experimental scheme to investigate the phonetic perception of vowels by subjects, which leads to the design of the main experiment. This is necessary because labelling or identification tests which are normally employed to investigate phonological perception will not be appropriate for a vowel quality study.

Using the design validated by the results of Chapter 4, Chapter 5 presents the main experiment to investigate listeners' strategy in the perception of vowel quality with dynamic formant trajectories. Some criticisms of the results are addressed in Chapter 6, and they are investigated by three supplementary experiments.

Finally a reassessment of the results and their implications for phonological or phonetic / psychoacoustic vowel perception are made in Chapter 7.

Chapter 2: Terminology: Reduction? Undershoot? Coarticulation?

2.1 Introduction

It is recognised that vowel formant shift is one of the major issues in phonetics but there is no unanimous description of this phenomenon although it seems superficially to have simple explanation. The difficulty of description is due to the variety of explanations made at different levels of linguistic, and phonetic or physiological analysis. There are three well-known labels to this phenomenon: 'undershoot', 'reduction' and 'coarticulation'. This chapter presents a discussion of the definitions and arguments given by previous researchers and tries to draw a systematic and explanatory picture.

2.2 Description 1: Undershoot

Surprisingly, Lindblom (1963), in one of the earliest detailed studies, does not define the terms 'reduction'/centralisation/'undershoot' in his paper, although he repeatedly mentions 'undershoot'. However, it seems to be the case that he uses the term 'undershoot' to mean failure in reaching the *acoustic* target, as well as the *articulatory* one;

> In the acoustic domain, [articulatory undershoot] is paralleled by undershoot in the formant frequencies relative to the bull's eye formant pattern.
> (Lindblom 1963:1779)

In another section, Lindblom attempts to separate 'centralisation' from the idea of undershoot, claiming, "[the term centralisation] should imply nothing about the dynamics of vowel articulation." (Lindblom 1963:1781) However he does not try to differentiate between undershoot and reduction. Although the title of his paper is "A spectrographic study of vowel reduction", the term ' vowel target undershoot' is used interchangeably with 'vowel reduction'.

Later, faced with counterexamples to his old version of undershoot theory, Lindblom (1990a) revised his strongly duration-dependent version of vowel undershoot, deriving what seems a more reasonable and less ambitious conclusion, which is discussed in Chapter 3. However, there again he does not distinguish the term 'reduction' from 'undershoot'. For example, Section 4.5 of Lindblom (1990a) has a heading of "Vowel reduction in clear speech", but what he refers to in that section is 'undershoot' in clear speech without using the term 'reduction'.

This line of argument is also observed in Moon (1991), who defines 'reduction', without giving the definition of 'undershoot', in the following way:

> In many situations, this contextual influence is so strong that vowels are reduced which means that the degree of phonetic contrast within a vowel system is decreased. (Moon 1991:15)

Like Lindblom, he does not try to differentiate the terms 'reduction' and 'undershoot', implying that they are synonymous. For example, Moon (1991:19) states, "the degree of undershoot will depend on the duration of the applied force (cf. "duration-dependent" reduction) ... " It is obvious that his argument is based upon the "duration-dependent" theory of Lindblom (1963), where the phenomenon is referred to as "duration-dependent undershoot." Or, another example, "... It seems that the presence of reduction (undershoot) is speaker-dependent." (Moon 1991:29). Furthermore, discussing Lindblom (1963), he proposes, " [Lindblom in 1963 version] attributed the observed formant shifts, or reduction phenomena, to contextual assimilation." (Moon 1991:18)

To summarise, Lindblom and Moon describe formant shift as undershoot. Although they do not distinguish it from reduction in a strict sense, one could observe that their definition of undershoot / reduction is different from those introduced in 2.3.

2.3 Description 2: Reduction

Unlike Lindblom and Moon, there are some who prefer solely the term 'reduction'. For example, Delattre (1969) adopts this term. However he does not give any clear definition of reduction. What Delattre meant by reduction would probably be the change from a full vowel to its weak counterparts in the lexical entry, not at the phonetic or acoustic level. This view is supported by his example chosen to explain reduction, which was the alternation of /i/ and schwa in "competing" and "competition" and by the fact that the materials in the experiment focused on strong-weak vowel alternation in lexicon.

Perhaps due to his focus on the higher levels, he does not distinguish between the acoustic and articulatory aspects of reduction and he considered them only as data supporting his argument. It could also be observed that he did not discriminate between factors causing reduction but belonging to different levels: lexical, phonetic, inter-linguistic (e.g. rhythm pattern difference between languages, like syllable-timed and stress-timed rhythm), lexical/syntactic stress assignment.

A more comprehensive work on vowel 'reduction' is Koopmans-van Beinum (1980). She defines reduction as "the deviation in the acoustic space from an ideal position or target towards a central point in the vowel diagram in combination with a decrease in the vowel duration." (Koopmans-van Beinum 1980:9-10) She also refers to coarticulation as "the acoustic variation caused by differences in the neighbouring speech sounds", claiming that this notion of 'coarticulation' is the result of a phonological way of thinking. However she does not explicitly mention why it is phonological.

After the acoustic analyses she reviews the terms 'reduction', 'centralisation' and 'neutralisation' and she argues that 'reduction' is an "inadequate" term since further specification is required for this term to be compatible with her results. She rejects the term 'centralisation' since it is also "used for the description of certain articulatory movement" and it does not "do justice to the alterations in vowel duration." (Koopmans-van Beinum 1980:73). She rejects the term 'neutralisation', claiming the notion of a neutral vowel is not clear. Finally she proposes the term 'contrast reduction', asserting, "it is applicable to the acoustic contrasts of vowels in the formant field as well as to vowel duration." However, if this argument is pursued, this term itself is inadequate since the notion of acoustic contrast has not been defined, still less phonetic contrast in the acoustic/articulatory vowel system, and therefore there is no positive benefit in introducing or inventing a new term.

By contrast, Koopmans-van Beinum does not use the term 'undershoot'. In an earlier chapter, when discussing Lindblom (1963), she writes, "for a given consonantal environment the deviation of a vowel from its target position may be calculated as a function of the vowel duration" (Koopmans-van Beinum 1980:10), without any reference to Lindblom's use of the term 'undershoot'.

Another comprehensive series of studies on reduction are carried out by van Bergem. Van Bergem (1989), in his earlier work, describes two levels of reduction: linguistic vowel reduction and articulatory vowel reduction, as follows:

> The loss of vowel quality due to a low articulatory effort and dependent on the way the articulators interact is called articulatory reduction. The substitution of a full vowel by a schwa, a long vowel by a short one, or a diphthong by a monophthong is called linguistic reduction.
> (van Bergem 1989:104)

It should not be overlooked that the method which van Bergem (1989) uses to investigate articulatory reduction is to observe the F1/F2 acoustic diagrams, which presupposes one-to-one mapping between articulatory/acoustic vowel diagrams. Thus at best he could only investigate articulatory reduction by inference from the acoustic evidence.

In a more recent paper van Bergem(1991a) develops his argument, defining two levels of reduction, namely acoustic and lexical. He defines acoustic reduction as "a loss of vowel quality due to relaxed articulation in less informative parts of an utterance" (van Bergem 1991a:[10-1]), and lexical reduction as "the substitution of one vowel by another that is easier to pronounce." (van Bergem 1991a:[10-1]) In addition van Bergem (1991a) proposes a hypothesis that lexical vowel reduction is "a systematic expansion of acoustic vowel reduction" (although its origin can be recognised in van Bergem (1989:103), as "linguistic reduction is a consequence of articulatory reduction."). His point of view stems from the following assumptions:

> 1) acoustic reduction is a consequence of a speaker's striving for articulatory economy
>
> and
>
> 2) "In general, unstressed vowels usually lack a clear quality due to acoustic reduction." [10-3],
>
> therefore
>
> 3)"It is only a small step to replace such a [reduced] vowel ... with a schwa and to make this substitution a permanent part of the lexical system"
> (van Bergem 1991a:[10-3]).

But this line of argument is inadequate by itself, because it cannot explain the problem of his examples: why does lexical substitution of a schwa for the full vowel /æ/ occur in *mariner* -> *marine* but not in *capture* -> *captivity*? Although he insists that lexical reduction occurs only in words whose lexically reduced forms are sufficiently "unique to be discriminated" (van Bergem 1991a:[10-3]), this additional criterion of uniqueness is vague. He also refers to pronunciation in an English language dictionary to obtain a lexical form. This clearly brings about some problems: if a dictionary has two entries of "official pronunciation" (van Bergem 1991a) like *cyclic* /'sɪklɪk, 'saɪklɪk/, then, when one considers the process of lexical reduction in the process of *cycle* -> *cyclic* with reference to an English language dictionary, it will certainly require supplementary explanation to judge whether it should be considered to be a process of lexical reduction or not.

The separation that van Bergem claims between two levels of reduction, lexical and phonetic (acoustic) seems too obvious to emphasise here, but confusion does occur in some research. For example, Keating & Huffman (1984:196) state, "One of the reasons we expect Japanese vowels not to vary much acoustically in prose is that there is no reported vowel reduction ...", although it is recognised that Japanese does not have lexical reduction but does show the shrinkage of acoustic vowel space in spontaneous speech. Their mistaken statement could be assumed to have stemmed from the confusion of the levels mentioned above.

This separation of two notions is also proposed by Fourakis (1991). There he separates the terms phonological and phonetic reduction. According to him, phonological reduction is, "the phonological process whose application causes unstressed vowels to be realized as schwas ... " (Fourakis 1991:1816). He uses an example of the alternation of the second vowel of 'telegraphic' and 'telegraphy', derived from the same morpheme {tele}, stating;

> In telegraphy, the second vowel ... is realized as a front, mid, lax vowel [ε]. In telegraphic, this vowel is realized as a schwa, the result of the application of the phonological process of vowel reduction. (Fourakis 1991:1816)

For phonetic reduction, he quotes the definition by Miller (1981):

> Vowel reduction refers to the tendency for the obtained formant frequencies of a vowel to fall short of the idealized target values for that vowel --- those values that would be obtained if the vowel were produced in isolation --- resulting in an overall shrinkage of the vowel space.
> (Fourakis (1991:1816), taken from Miller (1981:42))

Furthermore he describes vowel centralisation and vowel neutralisation as "a movement toward a schwalike formant pattern" (Fourakis 1991:1816), and proposes:

> the terms vowel reduction, vowel neutralisation, and vowel centralization will be treated as synonymous, while Lindblom [1963 version]'s vowel reduction by formant undershoot will be referred to as formant undershoot.
> (Fourakis 1991:1817)

He does not reveal why 'undershoot' would not be treated as synonymous with the others (i.e. reduction, centralisation, neutralisation), or why these three could be treated as synonymous.

Despite Fourakis' definitions, his paper shows a lack of consistency in the usage of the notion 'phonetic vowel reduction'. First of all, he declares that he will take the definition of Miller (1981), which is evidently composed of two components: 1) "the missing of the (acoustic) target" , and 2) "the shrinkage of (acoustic) vowel space". Later, he redefines it, referring to, "phonetic vowel reduction, defined as a shift of formant pattern of vowels toward a neutral vowel" (Fourakis 1991:1825). Clearly here the second component 2) of the definition is deliberately neglected while a new qualifier, 3) "shift toward a neutral vowel", is introduced. Furthermore, in conclusion, he asserts, " ... phonetic vowel reduction and shrinkage of overall vowel space seem to be dependent on different factors." (Fourakis 1991:1826). It seems that there he supposes that the notion of phonetic vowel reduction does not contain the component 2) 'the shrinkage of vowel space', and possibly that these two notions should be treated separately.

If it were the case that 3) is the necessary and sufficient condition of component 2), then the definition consisting of components 1) and 2) would be equivalent to that consisting of 1) and 3), and therefore Fourakis' definition would be consistent and he would not confront this contradiction. However, as is commonly known, vowel space shrinkage does not always imply a shift to a "neutral position", which he defines to by an acoustic analysis as the articulatory resting configuration derived from the vocal-tract tube model by Fant (1970), with resonance peaks of 500, 1500, 2500 Hz.

Then with respect to the legitimacy of phonetic vowel reduction, component 1) should be consistent and distinctive from the other notions, (i.e. phonological reduction and undershoot), but Fourakis also asserts:

> In fact, some vowels, even though shorter in duration, were further away from the neutral point in the faster or unstressed conditions than in the slow-stressed. This could be due to formant undershoot having a reverse effect, that is, resulting in formant values that are further away from the neutral point formants than the target ... Phonetic vowel reduction, when it does occur, and when it is associated with shortened durations due to increase in tempo and shifting of stress, must be distinguished from formant undershoot. (Fourakis 1991:1825)

Reading this, one gets the impression that 'formant undershoot' means just missing the target with or without the shift to the centre, while phonetic vowel reduction does not give priority to the process of the component 1) missing of the target, but to 3) shift toward the neutral vowel, to which he switched during the course of his argument. It might be guessed that Fourakis' own idea of "phonetic vowel reduction" would be the one consisting of components 1) and 3).

It must be added that van Bergem's recent paper (1993b) mentions Fourakis' separation of phonetic/phonological reduction. He regards the term phonetic reduction as being "vague", insisting that "the term acoustic vowel reduction more accurately describes the change in vowel quality." (1993b:3) , although the term acoustic vowel reduction does not seem to be sufficiently clear, as will be discussed in 2.4. Van Bergem also mentions that in his paper, whose main concern is a vowel reduction in Dutch, the term 'lexical reduction' is preferred to 'phonological reduction', since he claims that in Dutch "there is no general phonological rule which changes a vowel phoneme into schwa" and the term lexical reduction is adopted which indicates that "a schwa replacement is specific to particular words rather than to an entire sound system." (1993b:3)

2.4 Description 3: Coarticulation

It must be added that some have adopted explanations at lower levels in the speech production process or its acoustic output. Nord states after acoustic analyses:

> This difference [of degree of vowel reduction according to word-initial/final syllable positions] could be expressed as a difference in coarticulation: the unstressed [non-final] vowels coarticulate with the consonantal frame, ... while the unstressed final vowels are reduced towards a more neutral space in the vowel plane. (Nord 1986:25)

What is notable is his attribution of the explanation to the level of coarticulation or "contextual influence," as he puts it. This attribution, together with the hypothesis above that all vowel reduction process as can be formalised without reference to the phonological level, forms the salient part of Nord's proposals. The adequacy of his proposal is discussed more in detail in 2.5.

2.5 Discussion

Here the relation between the three notions of coarticulation, reduction and undershoot is discussed.

As we have seen, Nord (1986) formalises the process of vowel formant shift only by reference to coarticulation. It is generally accepted that coarticulation triggers, say, vowel target undershoot, acoustically and articulatorily, but normally coarticulation is considered to be no more than one of the factors triggering undershoot, which include the effect of fast speech rate or speech style (spontaneous speech against reading text, for example), etc. Therefore, since coarticulation is a sufficient but not a necessary condition of vowel formant change, it is not appropriate to attempt an explanation of this phenomenon solely at the level of coarticulation. Some might insist on the following argument:

> a) coarticulation causes vowel formant shifts not only in unstressed positions but also in stressed positions, at whatever speech rate or in whatever speech style speakers make an utterance
>
> and
>
> b) these other factors (i.e. speech rate and style etc.) just control the degree of

vowel shift

therefore, in conclusion

c) every formant shift may be described by reference to the degree of coarticulation.

Although this sounds reasonable, if this point of view is to be pursued, then 1) coarticulation has become one of the parameters of reduction, not the cause of it, and 2) one is obliged to describe what controls the degree of coarticulation. Some aspects of the vowel formant shift can be described only by coarticulation as Nord (1986) proposes in his conclusion, but it seems that by no means all of these aspects can be ascribed to the coarticulatory effect.

For 'reduction', as van Bergem (1989/1991a) and Fourakis (1991) suggested, the lexical reduction process and the phonetic (acoustic) reduction process must be kept separate, for the purpose of avoiding confused arguments like that of Keating & Huffman (1984) mentioned in 2.3. The term 'reduction' has been defined so loosely that without the separation of two levels or its redefinition, one could not get rid of the danger of intermingling these two processes.

In addition to the necessity of separation of the two levels, the notion of phonetic reduction presupposes either the notion of 'phonetic vowel contrast' as in Nord (1986) or Koopmans-van Beinum (1980:73), or shrinkage of the vowel space, as in Fourakis (1991), taken from Miller (1981). The validity of the notion of phonetic vowel contrast depends, however, on what the term 'contrast' implies.

Ordinarily, a system is defined as a set of opposing elements, each possessing its own properties which help distinguish it from the others, and a contrast of phonetic elements is obtained by the opposition of those properties in a system. Therefore the validity of the notion of phonetic vowel contrast is based on that of the notion of a phonetic vowel system. In this case, then, does the phonetic vowel system or the phonetic vowel contrast really exist? The so-called phonetic vowel system shown in acoustic vowel diagrams is not a sufficient proof, because it does not demonstrate opposition between the elements, and their relative position in an acoustic phonetic diagram varies according to the scale chosen, (i.e. linear, logarithmic, mel, Bark ...) and there might be some scales where each element is sufficiently remote while in others it is overlapped partially. The more fundamental problem is that, unlike phonemes, phonetic segments are not defined by means of abstract opposition or distinctiveness, and therefore phonetic segments do not possess abstract distinctiveness. Even if one attempts to obtain distinctiveness of phonetic segments on the basis of absolute physical value differences, it is next to impossible since it is the case that "while intrasystemic distinctiveness among speech sounds remains unchanged their absolute physical characteristics can be selected in a variety of ways." (Lindblom 1990b:137, Figure 1). Due to this deficiency of opposition among elements, it is difficult to retain the notion of a phonetic vowel system or a phonetic vowel contrast.

The essential predicament of the notion of vowel space shrinkage is closely related to that of phonetic vowel contrast: the shrinkage is dependent on the way to scale an acoustic or articulatory diagram. In this sense van Bergem's definition of acoustic/articulatory reduction avoids the problems by concentrating on loss of vowel quality, without mentioning vowel space shrinkage and loss of phonetic contrast. However, van Bergem does not clarify what he means by 'vowel quality' in his paper, and the term itself is quite vague.

Finally the term 'undershoot', although sometimes used interchangeably with 'reduction', does not appear to be a better term unless it refers only to the missing of the acoustic/articulatory target. Some might insist that the term 'undershoot' is not suitable because it does not refer to durational change, as Koopmans-van Beinum (1980) does when she rejects the notion of centralisation. However this thesis looks only at vowel shifts and not the durational factor, so that does not affect adoption of the term 'undershoot' here.

In conclusion, three terms to define vowel formant shift have been investigated and it is suggested that the term 'undershoot' is, if not optimal, the safest choice, since the other two terms: coarticulation/reduction require further specification. It would be also recommended that the term 'reduction' should be used solely for a phonological substitution of a full vowel with a schwa.

Chapter 3: Previous Studies: how vowel undershoot has been investigated

3.1 Introduction

As mentioned in Chapter 1, vowel formant undershoot has been one of the critical issues among phoneticians and speech scientists although most of these studies have shed light mainly on one aspect: how listeners reconstruct vowels as phonological segments from coarticulated speech waveforms, which differ from the isolated forms and do not provide any well-defined boundaries. In other words, researchers have been mainly concerned with how one can identify these reduced quality vowels.

In the following sections, a detailed survey of vowel undershoot studies is presented. Section 3.2 describes acoustic studies on how and to what extent the formant frequencies undershoot the acoustic target in various circumstances. Section 3.3 deals with the relevant categorical perception of vowels with formant undershoot: how vowels showing formant undershoot are phonologically labelled. Section 3.4 discusses the psychoacoustic / phonetic quality perception of a vowel when its formantsshow undershoot, which is the main focus of this study. Finally, Section 3.5 explains the motivation for the experiments in the next chapter.

3.2 Acoustic analysis of vowel formant undershoot

3.2.1 Stevens & House (1963)

One of the earliest comprehensive investigations of acoustic vowel formant undershoot phenomena is by Stevens & House (1963). They first cast doubt on earlier studies pointing out that these studies are "inconclusive" since they do not take into account 1) the variance of formant frequencies between speakers, 2) the variance caused by coarticulation with neighbouring segments, 3) the dynamic nature of vowel articulation: casting doubt on "one-point" investigation, and 4) the lack of a "generally accepted procedure" of formant measurement.

Materials in their experiment were /həCVC/ where the initial and final consonants of /CVC/ were identical. These /CVC/ syllables had options of eight vowels and fourteen consonants (excluding nasals) from American English. They were uttered without a frame sentence by three male speakers. Note that the materials used were produced only once, although the investigators insisted, "the recorded utterances [=words] were evaluated by three phonetically sophisticated listeners, and the utterances that were judged to be unacceptable samples of the phonemes in question were re-recorded" (p.112)

In order to certify the validity of their data, they also investigated vowels in the null contexts: /h_d/ and /#_#/, uttered by the same three speakers and claimed that the results obtained from their subjects were in agreement with those obtained in other established studies like Peterson & Barney (1952) and Tiffany (1959).

When analysing the results, they first responded to the point 1) cited above: the speaker variance. They took X-ray photos of the vocal tract of the subjects, measured the rough vocal tract length, and concluded that the differences in percentage of vocal tract length were "roughly compatible to the percentage differences in formant frequencies" (p.118) for null context vowels.

The variance of formant frequencies caused by coarticulation with neighbouring segments, stated above as 2), was ascribed by them to two general observations: a) consonantal influences on formant frequencies differ from one vowel to another ("some vowel articulations are more stable than others"), and b) consonantal contexts shift formants systemically, depending on the features of the consonant: [place], [manner], [voice].

Mentioning the effect of voicing on formant shift, Stevens & House stated that F1 values were generally lower in the voiced environment and that F2 values were higher for front vowels in the voiceless environment. However, in Figure 7 of Stevens & House (1963:120), these shifts by voicing contrast did not seem to be large.

In conclusion they claimed that the modification of formant frequencies by the consonantal context was systematic and they attributed the cause of the modification to "the inherent dynamic properties of the articulatory structures and of the neuromuscular system that controls them." (p.122) In fact they made an assumption about the articulatory aspect from acoustic spectrograms, since they do not investigate this phenomenon articulatorily. They also labelled the formant shift as (articulatory) undershoot, stating that during the production of a syllable the response given by the articulatory muscular system, controlled by neural signals, was an undershoot of a target vowel configuration "presumably ... corresponding to a null context." (p.123)

In summary, Stevens & House (1963) inquired into the acoustic vowel formant undershoot caused by various consonantal influences, by analysing acoustically bisyllabic words with all consonantal variations that were obtained from three subjects. They tried to overcome the variances caused by different speakers / consonantal environment / measurement position of /CVC/ trajectories. They claimed that these variances were systematic, by supplying an analogical explanation from articulatory behaviours.

3.2.2 Lindblom (1963)

Lindblom (1963) observed the formant undershoot of eight Swedish vowels in a different consonantal/stress environment. He claimed that the past theories (i.e. those until early

1960's) focused solely on static aspects of spoken language, and therefore he was concerned with the dynamic aspects of vowel articulation through a spectrographic analysis of vowel undershoot in Swedish.

Lindblom (1963) selected eight Swedish vowels in three consonantal frames: /b_b/, /d_d/ and /g_g/, which were embedded in either initial or final position of a frame sentence. These /CVC/ syllables were also recorded as stressed or unstressed. All materials were produced by one subject, who was asked to read those sentences synchronising to the isochronous beats that could be heard through headphones. Lindblom made a spectrographic analysis, measuring the first three formant frequencies "at a point in time where the first derivatives [of the F1/2/3 trajectories] equal zero." (p. 1775) i.e. a point where the formant trajectory changes its direction (henceforth called the turning sample point). In addition, he measured the onset/offset points of /CVC/ trajectories for F2/F3, and the duration of vowel segment as well.

From the measurements, he found that the formant frequencies of the turning sample points were subject to change as duration decreased, and proposed two exponential functions[1] to estimate the F1/F2 frequencies at these turning sample points. Target values in these formulae were found not by a separate acoustic analysis, but by referring to a linear function that was calculated by approximating the turning point frequencies of all /CVC/ context over all duration patterns.

Considering all his results, he suggested, "timing is the primary variable in determining the reduction of sounds ... ." To verify this, he carried out another experiment, where

[1] The function for F2 is:[F2 turning point frequency] = $\alpha * e^{-\beta\gamma}$ + [F2 target value], where $\alpha = k *$ ([F2onset]-[F2target]), (k is a constant determined by consonantal context) β = constant chosen for each consonantal context, independently of the vowel , γ = duration of the vowel segment (in ms). The function for F1 is almost identical, except that in the formula above he replaced the onset values in 'α' with the value of 375 Hz and introduced another criterion that the F1 turning point frequency is identical with the F1 target in all contexts, provided that F1 target is less than 375 Hz.

the subject repeated the /CVC/ syllables that were used in the main experiment, but with a variety of speech rates, cued by beats through headphones. The result was that the productions of the subject diverged from the target as the speaking rate increased, and Lindblom concluded, "Duration seems to be the main determinant of the reduction." (p.1780) This is a strongly duration-dependent undershoot theory, which has provoked many subsequent arguments. In fact, later, faced with counterexamples, Lindblom (1990a) revised his strong duration-dependent version of vowel undershoot, deriving what seems a more reasonable and less ambitious conclusion: that duration is just one of the factors which determine the degree of undershoot, and that "speakers have a choice" (i.e. undershoot is not obligatory but determined by the meeting point of economy of articulatory efforts and sufficient contrast in the sound system.)

Finally it must be mentioned that, like Stevens & House (1963), Lindblom (1963) was ready to provide an entire explanation for an acoustic phenomenon in terms of articulatory behaviour. He hypothesised:

> Provided that the neural events corresponding to the phonemes actually stay invariant, the speech organs fail, as a result of the physiological limitations, to reach the positions that they assume when the vowel is pronounced under steady-state conditions. In the acoustic domain, this is paralleled by undershoot in the formant frequencies relative to the bull's-eye formant pattern. (p.1779)

However, his assumption that a speaker's control of his speech organs in vowel articulation is associated with "neural events that are in a one-to-one correspondence with linguistic categories" (p.1778) was provided without any physiological evidence of his own, or from any studies concerning this phenomenon, and therefore his argument seems unsatisfactory.

3.2.3. Koopmans-van Beinum (1980)

Koopmans-van Beinum (1980) investigated formant undershoot in various speech conditions: in different consonantal environments, at different levels, and with different stress assignments. She pointed out the lack of previous research on vowel 'reduction' in a natural condition.

The solution adopted by her was to set up various speech conditions: 1) vowels in isolation, 2) in isolated words, 3) in read texts, 4) in retold stories, and 5) in free conversation, and then to carry out vowel identification tests as well as acoustic analysis. Conditions 3-5 had two options: in stressed and unstressed positions, and therefore the number of conditions was eight. Her vowel identification test is discussed later, in 3.3.5. She tried to enumerate all possible consonantal contexts in Dutch, when the vowels occurred in those contexts, although the distribution among consonants was inevitably uneven. Since her pilot studies revealed only a small variance in isolated vowels in comparison with those in running speech, she obtained only three repetitions of isolated vowels (condition 1), and five repetitions of vowels in words uttered in isolation (condition 2), while she obtained ten repetitions in the other conditions (conditions 3-5).

The parameters which she considered were F0, duration and formant frequencies (specifically F1/F2). In formant analysis, to measure a degree of acoustic vowel space shrinkage, she introduced a reference point called the "centroid"[2]. She defined the centroid as "a point in the formant field representing the mean F1 and mean F2 values" (p.55) of all the vowels in each speech condition and speaker.

---

[2] She tried to demonstrate its adequacy as the reference point in the following way. First she calculated and plotted on an F1 / F2 plane "grand centroids", which were the mean (F1,F2) per speaker over all speech conditions across all vowels in all consonantal contexts. She also calculated the regression line on an F1 / F2 plane from mean F1 / F2 across each speech condition and for each speaker. She then attempted to display that on the plane the regression line was drawn fairly close to these four grand centroids. Then she insisted, "Although the material consists of only four speakers, and of course more evidence is needed to prove the universality of this regression line, it is not unlikely that these vowel centroids indicate some neutral position of the articulatory organs." (pp. 55-56) However she did not state why the linear relationship between F1/F2 in the formant diagram enabled her to conclude that the centroid (or even the grand centroid) indicates some "anatomically" (p.56) neutral position without physiological or medical investigations on each subject. All that she did was to measure the height of each subject, which she supposed could serve an index to the vocal tract length of each speaker. It seems that she strongly wished to associate, with the articulatory neutral position, the concept "centroid", which she imported from statistics, but the explanation appears weak.

As a next step, she calculated the parameter "Acoustic System Contrast" (henceforth ASC), a mean square distance to the centroid in log Hz in each speech condition. Based upon these ASC values, she made a further statistical analysis on the ranking of the five speech conditions: a) vowels in isolation, b) in isolated words, c) in read texts (stressed / unstressed position), d) in retold stories (stressed / unstressed position) and e) in free conversation (stressed / unstressed position). She discovered that, although the degree of the decrease of the ASC differed from one speaker to another, the pattern of ASC decrease shared a common tendency among the speakers, in the order of;

[max ASC] [min ASC]

(a) > (b) > (c:strs) > (c:unstrs) > (d:strs) > (e:strs) > (d:unstrs) = (e:unstrs)

(NB. strs = in a stressed position ; unstrs = in an unstressed position)

(NB.2 (d:unstrs) = (e:unstrs) means that there was no ASC difference between two conditions)

She also computed a correlation between the three acoustic parameters (i.e. F1/F2/duration) in pairs, and refuted Lindblom's strong duration-dependent undershoot theory (See 3.2.2.) on the basis of a 'not remarkable' correlation between duration and formant shift. She maintained, "Since two different systems (i.e. durational decrease and ASC decrease) are combined, a lower correlation [between duration and formant shift] over the eight speech conditions is the result" (p.67), asserting, "vowel duration [is] mainly determined by the condition stressed-unstressed." (p.67)

At the end of the acoustic analysis, she concluded;

> With respect to the durational reduction [,] stress turned out to be the main determinant, while on the basis of the formant shift we succeeded in devising a measure of acoustic system contrast for which freedom of word choice and of the sentence structure plays an important role and to a minor extent, stress. (p.68)

Based on these observations obtained from acoustic analysis, Koopmans-van Beinum also devised a vowel identification test as a perceptual test, which will be discussed in 3.3.5.

To summarise, Koopmans-van Beinum (1980) investigated vowel formant undershoot in different speech conditions, introducing the parameter Acoustic System Contrast, the sum of the distances to the centroid, which was the mean position of all vowels in the F1/F2 diagram. She concluded that the speech condition contributed more to the formant undershoot than stress, rejecting the strong duration-dependent undershoot theory proposed by Lindblom.

3.2.4 Fourakis (1991)

As discussed in Chapter 2, the separation of phonological reduction from phonetic reduction was not thoroughly argued in the field of phonetics and phonology until Fourakis (1991) and works by van Bergem. Fourakis (1991) is discussed here, while the works by van Bergem are reviewed in the following section.

Fourakis (1991) first declared that he would look into 'phonetic' vowel reduction. He stated that his experiment "aimed to determine the extent to which phonetic vowel reduction, brought about by destressing and increased tempo, affects the nine monophthongal, nonretroflex vowels of American English." (p.1817), although in conclusion (p.1826) he modified his argument, placing more emphasis on which factor (stress/speech-rate/coarticulation) contributed most to phonetic vowel reduction and "shrinkage of the overall vowel system".

His experimental procedure was as follows; nine nonretroflex monophthongs in American English were embedded in two /CVC/ contexts: /h_d/ and /b_d/. These /CVC/ words were located after a dummy syllable (written as "kay", probably pronounced like the second syllable of "decay") and they were put in a carrier sentence, "I will say kayCVC again." Stress was assigned either on the target /CVC/, or on the dummy syllable (to make the target /CVC/ unstressed). These carrier sentences were read four times by four male and four female speakers of General American English.

In spectral analysis, Fourakis (1991) adopted Miller's (1989) auditory-perceptual theory of vowel recognition, which proposes that a vowel is represented by a path in auditory-perceptual three-dimensional space[3]. The portion corresponding to the quasi-steady part of the /CVC/ syllables was selected by observing the clustering points along the path, and the F0/F1/F2/F3 average values over the time corresponding to that quasi-steady part were calculated. He also measured the durations of the nuclei of the target /CVC/ as well as that of the target /CVC/ itself and the whole carrier sentence.

The results of his experiment show the following points:

> (1) duration: a change in stress had a slightly greater effect on vowel (nuclei of /CVC/) duration than a change in tempo, and the difference, although small, was statistically significant.
>
> (2) formant frequency: Fourakis calculated the mean Euclidean distance in the auditory space of Miller (1989) between each point representing a quasi-steady part of /CVC/ and the 'neutral reference point' obtained with reference to Fant (1970): (F1,F2,F3) = (500 Hz, 1500 Hz, 2500 Hz) and F0 (male) = 133 Hz. The results of ANOVA indicated that only marginally[4] significant effects on the mean distances were made by context and stress, and that no significant effects were made by tempo. He insisted, "even though subjects increased their tempo and shifted stress away from the target syllable, they still produced vowels with distinctive formant patterns." (p .1824)

[3] Fourakis (1991) stated that the three axes of the three dimensions are defined by the following calibration: x = log (SF3 / SF2); y = log (SF1 / SR); z = log (SF2 / SF1). SF1, SF2 and SF3 are the first three significant prominences in the short-term spectrum of a vowel utterance; and SR stands for sensory reference, defined as; $SR = 168(GMF0 / 168)^{1/3}$ (GMF0 is a geometric mean of the speaker's fundamental frequency)

[4] Fourakis (1991) defined that F-values were significant if $p < 0.01$, and marginally significant if $0.01 < p < 0.05$

> (3) vowel space: Fourakis also calculated the area consisting of the nine vowels in each condition on a two dimensional projection of Miller's three dimensional auditory space, claiming, "Although individual vowel formant patterns were not affected, at a statistically significant level, by the change of tempo and stress, it was still possible that the vowel space by the nine vowels in each condition differed ... in terms of area." With regard to these obtained areas, he asserted[5], "It can be seen that, even though the previous analysis had indicated no statistically significant changes in vowel formant patterns, the overall area was obviously affected when subjects changed tempo and/or stress. (Fourakis 1991:1824)

Fourakis (1991:1826) concluded that 'phonetic vowel reduction' (presumably shift to the neutral reference point in vowel space, as discussed in Chapter 2) and "the shrinkage of the overall vowel space" seemed to be dependent on different factors; the former, he insisted, was susceptible to the coarticulatory effect, but not to the tempo/stress effect, while the latter was "affected by tempo and stress, but not by context". This observation assumes that the process of the vowel space shrinkage is independent of the shift to the neutral reference point in vowel space.

3.2.5 van Bergem (1988/1989/1993b)

In a series of his papers, van Bergem conducted a number of studies on vowel reduction in Dutch, proposing a separation of linguistic and phonetic types of vowel reduction.

Van Bergem (1988) carried out two acoustic experiments; in the first, he investigated the formants and duration of /CVC/ syllables in the following contexts: 1) embedded in a frame sentence in a stressed position and 2) in isolation. They were uttered by van Bergem himself, only once per token. He compared formant values and duration between

---

[5] He made this remark simply on the basis of the comparison between the raw numbers without any statistical analysis. He further stated that the mean areas for each context across four conditions (two tempi & two stress patterns) were equal and therefore concluded later that context did not affect the overall vowel space (p.1826).

the two conditions, claiming that there was a significant difference in duration between the two conditions but the mean distance towards the centroid in the F1/F2 diagram did not show a significant difference. In the second experiment, a free conversation of one subject was recorded (condition 1) and part of it was transcribed into a text. Then the subject was asked to read the text (condition 2), and also read some words that were isolated from it (condition 3). An analysis was made of duration and formant frequencies across three vowels /i ɔ a/. The result of this second experiment was that there were significant differences in duration across three conditions. It also demonstrated that the difference in the mean distance towards the centroid in F1/F2 diagram was significant across three conditions for /i ɔ/ but not /a/. From the results of these experiments, van Bergem concluded that vowel reduction is observed only in some vowels and the speech condition has some effect on reduction although stress assignment does not.

In 1989, van Bergem published another paper on the distinction between articulatory and linguistic reduction, based on an acoustic experiment which is almost identical to the first one of van Bergem (1988), except that the speaker (himself) of the utterances attempted "to pronounce the sentences with a total relaxation of the articulators as in 'normal' speech, whereas the isolated CəCVCə were pronounced with a maximal articulatory effort." (1989:98). The comparison between two conditions showed the existence of a reduction effect on the utterance in the frame sentence, which was contradictory to his previous paper. He attributes this contradiction to the difficulty with the experimental method, stating, "[the discrepancy between the results of two studies] illustrates that it is far from being easy to study vowel reduction in strictly controlled experiments with nonsense words" (1989:100)

Confronting this problem, van Bergem (1993b) devised an alternative methodology. (Part of its results is published as van Bergem (1991a/1991b/1993a))[6].

In van Bergem (1993b), he noted three parameters that are supposed to affect spectral characteristics and duration of a vowel: sentence accent, word stress and word class (function/content word). Then, using the Dutch lexical database, he searched for triplets of Dutch words which share a common /CVC/ syllable at different points of their structure: (1)in a monosyllabic function word, (2) in an unstressed syllable in a content word, and (3) in a stressed syllable in a content word. He discovered 33 triplets in Dutch that fit the criteria above, although as a result of using natural words, the environment was not controlled[7].

These triplets of words were embedded in the middle of a carrier sentence, so that "each test syllable occurred in about the same position for all experimental conditions". In addition the examined /CVC/ syllables in isolation were recorded. The sentence accent had two patterns: on the examined word, or elsewhere in the sentence. Since the sentence accent cannot be placed on a function word (because he maintains that in normal speech "accented function words are rather exceptional" (p.4) in Dutch), he "attempted to create the same rhythmical pattern" (= stress pattern) using a function word preceded / followed by a content word.

These processes mentioned above produced seven conditions for each examined /CVC/:

---

[6] Chapter 2 of his PhD dissertation, van Bergem (1995), concerns the acoustic study of vowel reduction and in fact it is a reproduction of van Bergem (1993b). Therefore a separate discussion of van Bergem (1995) is not made here.

[7] The number of the syllables of the 'content' words used in (2) and (3) was either two or three, and the location of the examined /CVC/ in these content words was either word-initial, or word-medial, or word-final. Sometimes it is the case that the examined /CVC/ ("**van**") was located word-initially in condition (2) ("**van**dalen") but word-medially in condition (3) ("ha**van**na"). Neither were the coarticulatory environment surrounding the examined /CVC/ controlled; nor the type or frequency of the consonants appearing in the CVCs.

1) stressed syllable in an accented content word
2) stressed syllable in an unaccented content word
3) unstressed syllable in an accented content word
4) unstressed syllable in an unaccented content word
5) function word followed / preceded by an accented content word
6) function word followed / preceded by an unaccented content word
7) syllable spoken in isolation

15 Dutch speakers participated in the acoustic experiment. The "appropriate" accent/stress assignment pattern of the recordings was guided by the presentation of a pre-recorded question which affected their focus assignment and stress pattern.

F1/F2 diagrams of the mean formant values in mel scale were plotted for each speech condition, across all consonantal contexts and subjects. Van Bergem showed that the shifts due to context observed in every vowel were in the direction from the "target" position (calculated from /CVC/ spoken in isolation) to the "schwa", whose formant values he calculated by sampling the first 20 "schwas" from each speaker. These "schwas" used in this acoustic analysis were phonological / lexical, since he definitely stated, "the schwa [which he sampled] is a frequently occurring phoneme in Dutch." (p. 9)

Euclidean distances in the mel-scaled F1/F2 diagram from each vowel in each condition to its 'ideal' counterpart (i.e. vowels in /CVC/ in isolation) were also calculated, leading to the following observations;

1) Repeated measures analysis of variance showed that not only the factors "accent"/"condition" but also the factor "speaker" were significant. (Note, however, that he calculated the mean values across speakers in the procedure

above[8].)

2) Word stress/Word class have a greater effect on the spectral quality of vowels than the sentence accent, although this observation was made in reference to the line chart which indicates a larger "jump-up" in distance from the stressed token to the unstressed token than from the accented token to the unaccented token.

3) The absence of a significant difference in mel distance between two conditions of function words (i.e. preceded/followed by an accented content word, or not) indicates that, whether the neighbouring content word has received a sentence accent or not, they are not influenced in terms of the vowel quality.

It is worth mentioning the fact that he treated the function word with a neighbouring accented content word as "accented" when carrying out statistical analyses. As was discussed above, since the other "accented" words received accent on themselves, not the neighbouring words, this treatment seems to lack consistency.

As the next step of his formant analysis, he looked into the change in trajectories according to the experimental condition, by modelling the formant trajectory with a simple parabolic function:

$$F(t) = C_0 + C_1 t + C_2 t^2 \text{ (t: normalised time, } -1 <= t <= 1)$$

The coefficient C_2 serves as an index of the "amount of the (parabolic) curvature of the [formant] tracks" (p. 9). He plotted a graph of - C_2 of F1 as the ordinate and - C_2 of F2 as the abscissa, to "get a usual configuration of vowels". There he only plotted the values obtained from syllables uttered in isolation (condition 7) and the values obtained from the average of three "most reduced" conditions: 4), 5) and 6).

These parabolic curvature graphs for F1 and F2 showed a similarity to the figures

[8] He mentioned, "The factor 'speaker' which was significant ... will be discussed in section 4.3." (p. 9) However, section 4.3 does not present the support to calculate the mean values across the significantly different factor "speaker". It only showed that there were two speakers whose values differed from the rest.

obtained by plotting mel F1/F2 values, and van Bergem maintained that the similarity of these graphs supported the observation made on the F1/F2 frequency shift of the /CVC/ nuclei.

The experiment also showed that the mean duration changes across conditions had the same ranking (stressed syllable > unstressed syllable > function word) in both accent environments.

3.2.6 Discussion of acoustic studies

This section discusses how the studies in 3.2.1-3.2.5 are related to the topic of the current study which is concerned with the phonetic / psychoacoustic quality evaluation of a dynamic vowel.

Initially, it seemed that the main attention of researchers was the coarticulatory aspect of vowel formant undershoot. Then Lindblom (1963) proposed a theory of strong duration-dependent formant undershoot, where duration seemed to be the main determinant. As a consequence, the studies that followed Lindblom (1963) focused upon two points: the validity of his theory; or the investigation of other parameters involved in acoustic undershoot, such as word-stress, speech rate, lexical word-class (content or function word), pragmatic factors (text reading, reproduction of the stories known to the speaker, or free conversation).

The validity of the strong duration-dependent formant undershoot model was questioned by Lindblom himself (1990a), resulting in a more moderate version of the duration-dependent undershoot theory, which claimed that undershoot is not obligatory. Thus this study does not pursue the strong duration-dependent undershoot model in the stimulus synthesis of the perceptual experiment, dealing with duration and formant undershoot as independent acoustic parameters.

The second point, the interaction between these various parameters, may be of much interest if one can investigate how formant undershoot caused by the interaction gives a different phonetic vowel quality. However, it is difficult to create well-controlled stimuli for a perceptual experiment that investigates the effect of interaction, as is observed in the inadequate experimental design used by van Bergem, Fourakis, and Koopmans-van Beinum. Therefore, in this study, it would be best to concentrate on the parameters of formant undershoot due to coarticulation with neighbouring consonants and leave other parameters for future research.

3.3 Categorical perception of vowels with undershoot : how listeners reconstruct phonological segments

3.3.1 Lindblom & Studdert-Kennedy (1967) :perceptual overshoot

In 3.2 it was mentioned that a vowel is subject to variation in its formant structure due to various linguistic/extra-linguistic factors, resulting in acoustic undershoot, which, however, does not cause a failure in the reconstruction of phonological segments or the decoding of the linguistic message by the listener. One explanation for this phenomenon is to suppose that the listener, with the assistance of the acoustic attributes available in the acoustically underdetermined vowel, is able to reconstruct the target information which has been 'lost' in the course of speech production due to articulatory undershoot. This process is called perceptual overshoot, and was proposed by Lindblom & Studdert-Kennedy (1967).

The experimental procedure used by Lindblom and Studdert-Kennedy was the vowel identification test: a forced choice paradigm where the subjects had to decide whether a vowel (either in /jVj/ or /wVw/) was /ɪ/ or /ʊ/. They synthesised the material /jVj/ and /wVw/ with 20 ms steady-state offset/onset corresponding to the approximants surrounding the test vowel. The test vowel itself had a parabola trajectory for F2/F3, whose peak frequencies had 20 equally-stepped options between /ɪ/ and /ʊ/ (Note that they changed both F2 and F3 peaks simultaneously in the same direction by one step.) They also used the /#V#/ environment. They did not homogenise the number of the stimuli presented to each listener, although they claimed that the "total number of responses to each individual stimulus obtained from each listener was at least 15." (1967:833)

The result of this experiment indicated a strong perceptual overshoot. Comparing the perceptual /ɪ/ - /ʊ/ boundary shifts between /jVj/, /wVw/ and /#V#/ contexts, they remarked that, in comparison to /#V#/, the categorical boundary shifted to the lower F2/F3 frequency in the /wVw/ context (i.e. more stimuli were perceived as /ɪ/ than in /#V#/), while it shifted to the higher F2/F3 frequency in the /jVj/ context (i.e. more stimuli are perceived as /ʊ/ than in /#V#/). They concluded, "Boundary shifts in the [w] context occurred in such a direction as to compensate for formant-frequency undershoot in the vowels. Vowel recognition thus compensated for vowel production." (1967:843) In shorter tokens (i.e. duration= 100ms) they reported on the same tendency.

This hypothesis by Lindblom and Studdert-Kennedy, which is now a standard explanation for coarticulated vowel perception, has met with recent criticism. For example, van Son (1993), comparing the data of Lindblom & Studdert-Kennedy (1967) with his own, commented, "With our relatively small excursion sizes (of F2 trajectories) we already induced a sizable amount of diphthong responses. It is to be expected that the stimuli of Lindblom and Studdert-Kennedy induced an even stronger perception of diphthongs than our own" (van Son 1993:44), because their stimuli had a larger F2 excursion size. He eventually refuted the perceptual overshoot hypothesis; "When we consider the fact that their tokens strongly resemble glides or diphthongs (or even triphthongs), we might conclude instead, that they have only showed perceptual-overshoot [i.e. compensation towards the target frequency] for glides and diphthongs." (1993: 46-47)

Moreover, Rosner & Pickering (1994) reported on experiments that were similar to those of Lindblom & Studdert-Kennedy but provided contradictory results. They referred to Williams, Verbrugge & Studdert-Kennedy (1983), who modelled three-component sine-wave speech stimuli on the /ɪ/ - /ʊ/ and the /wɪw/ - /wʊw/ stimuli used by Lindblom & Studdert-Kennedy (1967). 'Phonetic' judgement[9] of the sine-wave speech continua gave the results similar to Lindblom & Studdert-Kennedy's, while no difference emerged between the two sine-wave continua when subjects had to judge the "pitch" of each stimulus as high or low. Subsequently Williams (1987) had subjects make 'phonetic' and pitch judgements of the /wɪw/ - /wʊw/ stimuli relative to isolated forms of /ɪ/ - /ʊ/, and the results showed that 'phonetic' and pitch judgements yielded the opposite response boundary shift. Rosner & Pickering also pointed out that the results of Centmayer (1975) on German vowels and of Nearey (1989) on Western Canadian English did not show as prominent boundary shifts as observed in Lindblom & Studdert-Kennedy: the boundary movements showed less compensation for undershoot than the perceptual overshoot hypothesis predicts.

The following sections discuss the theoretical alternatives to this perceptual overshoot hypothesis. These alternatives are supposed to concern more specific dynamic acoustic cues that may influence vowel perception.

3.3.2 Strange, Jenkins et al: Dynamic theory

Jenkins (1987) stated that it has been generally agreed since Joos (1948) that the frequencies of F1 and F2 are a necessary and sufficient cue for vowel identification. When listeners identify vowels whose formant trajectories miss the target frequencies

---

[9] The term 'phonetic' used in the quotations of Williams' papers in Rosner & Pickering is ambiguous with regard to whether it refers to the linguistic distinction of one phonological segment from another (=phonological), or it simply denotes the 'speech' sounds in general (i.e. not artificial and produced by human beings). Hence this study refers to the term 'phonetic' in single quotation marks when it discusses Williams' papers.

because of coarticulation, the traditional explanation is that listeners compensate to recover the 'target' formant values obtained for the vowel when spoken in isolation.

Strange (1987/1989a) stated that she, Jenkins, and their colleagues, initially in favour of this perceptual overshoot theory, came to harbour suspicion of it on the basis of the following examples: 1) the problem of coarticulatory variance: vowels coarticulated with consonants in /CVC/ syllables are identified relatively accurately and "in no study were isolated vowels perceived *more* accurately than coarticulated vowels, as would be predicted from target vowel theory." (Strange 1987:551); 2) the lack of acoustic invariance across speakers: target formant frequencies are not invariant because of vocal tract differences according to gender and age or even differences among those who belong to the same gender / age categories (See Strange 1989a: 2081-2082), but listeners are still able to identify vowels.

Having carried out a series of experiments utilising coarticulated /CVC/ syllables for identification tests (Strange, Jenkins & Johnson (1983), Jenkins, Strange & Edman (1983), Verbrugge & Rakerd (1986); these syllables were presented within a frame sentence in Strange (1989b), and Jenkins, Strange & Miranda (1994)), they presented "a dynamic specification model of vowel perception", which relies not on any particular spectral cross-section with some 'target' formant frequencies, but on the dynamic information available in the time-varying signals corresponding to vowel segments. Then as a source of that dynamic information, they proposed three hypothetical parameters that specify coarticulated vowels in /CVC/: (a) vowel intrinsic duration, or the durational difference among vowels intrinsic to them, (b) dynamic formant transitions into and out of the quasi-steady nucleus[10], and (c) the quasi-steady state nucleus in the vowel of /CVC/. (Strange (1989a) makes a revision of (c) by emphasising the existence of "vowel-

[10] The "dynamic formant transitions" do not necessarily mean that the listeners identify coarticulated vowels in reference to "formant track slopes" as mentioned in van Son (1993:71). At least, Strange and her followers did not claim that formant track slopes are used to identify the vowel in /CVC/. They only suggested the importance of onglide/offglide of a /CVC/ syllable, without specifying any particular parameter.

inherent spectral change" or diphthongisation in the /CVC/ nucleus. This "vowel-inherent spectral change" proposed by Andruski and Nearey is discussed in 3.3.3) None of these three hypothetical parameters are claimed to be "adequately captured in any single spectral cross section" (Strange 1989a: 2084), and information about the dynamic spectral structure is required for coarticulated vowel perception.

The verification of these hypothetical parameters was performed by creating /CVC/ stimuli which factored out the perceptual attributes brought about by one or more of the three parameters. Strange (1989b) gave special priority to the role that parameter (b), transitions into and out of the quasi-steady part of a vowel, plays in coarticulated vowel perception, specifically in /CVC/ syllables. She claimed:

> Since vowel-inherent spectral change patterns of [hybrid silent-centre syllables] are disrupted due to speaker differences, while temporal trajectory structure is relatively undisturbed, this result [of Strange (1989b)] argues that it is the temporal trajectory structure that provided perceptually critical information. (Strange 1989b: 2152)

Thus this strong version of dynamic theory asserts that information supplied by the transitions alone is sufficient to identify a coarticulated vowel. (See Andruski & Nearey (1992) for the opposite opinion.). It should also be noted that this is not entirely a novel idea. Verbrugge & Rakerd (1986) also suggested in conclusion, "[the acoustic information to identify /CVC/] is defined sufficiently by a relation between the initial and final regions of the syllable [i.e. the initial/final transitions of /CVC/]." (Verbrugge & Rakerd 1986:56)

This dynamic theory of vowel perception has been criticised in several respects. For example, Fox (1989) pointed out its lack of a perceptual process model, stating:

> [the dynamic theory] is relatively vague in terms of specific perceptual processes involved in identifying vowels from dynamic acoustic changes or the particular kinds of acoustic information which may be crucial to the listener's judgments. (Fox 1989:99)

Andruski & Nearey (1992) also criticised this theory, insisting that it assumes "identification rates may be as good or better for natural vowels in consonant context than for natural isolated vowels" (Andruski & Nearey 1992:403) despite the fact that vowels in isolation are more separated in acoustic vowel diagrams than those in /CVC/. Nearey (1989) and Andruski & Nearey (1992) reported that (1) some studies showed a higher score for isolated vowels; (2) any perceptual advantage for vowels in /CVC/ may have resulted from the choice of responses used in perceptual experiments, where for isolated-vowel stimuli there was no single orthographic form; (3) although some dynamic theory studies showed a small advantage for coarticulated vowels, one cannot exclude the possibility that speakers might not have been practised in the production of lax vowel stimuli in isolation since lax vowels do not occur in open syllables in English. In a more recent paper, Jenkins, Strange & Miranda (1994) introduced a procedure called "task-familiarisation procedure", which selected, out of 106, 76 subjects who could manage the experimental task, and this could contradict the criticism (2) above. However it must be remarked that this dynamic theory has not managed to reply to the first crucial criticism by Fox yet, and Jenkins, Strange & Miranda (1994) in the endnote, admitted this;

> One of the reviewers suggested that the term "dynamic information" should be more carefully specified ... We agree in principle that such specification is desirable and, indeed, this is one of our goals. The present study, however, does not separate these kinds of time-varying information. (Jenkins, Strange & Miranda 1994:1042, Endnote 1.)

This lack of parameter specification remains one of the crucial counter-arguments against the dynamic theory of vowel perception.

Overall, the dynamic theory of vowel perception advocates three dynamic parameters for the identification of coarticulated vowels in /CVC/ syllables. Some supporters of the theory propose a stronger version, which asserts the sufficiency of the information provided by vowel-inherent duration and transitions alone.

3.3.3 Andruski & Nearey (1992): vowel inherent spectral change

English monophthongs, especially when they are uttered in isolation, are generally considered to possess a quasi-stable formant structure. However, it is also claimed that a closer observation proves that the majority of American English monophthongs show formant fluctuation even without any coarticulatory effect. This phenomenon was originally reported only as an oddity of impressionistic phonetics, but recently some researchers have shed light on the importance of this acoustic information in vowel identification.

Nearey & Assmann (1986) called this formant fluctuation 'vowel inherent spectral change', which they defined as "the relatively slowly varying changes in formant frequencies associated with vowels themselves, even in the absence of consonantal context" (Nearey & Assmann 1986:1297). From the result of their perceptual experiment of windowed isolated vowels, they reported that this spectral change was a major perceptual cue for vowel identification. In addition, after a preliminary spectrographic analysis of /bVb/ syllables, Nearey & Assmann (1986) insisted that this vowel inherent spectral change persisted in /bVb/ syllables. They also asserted that this could account for the high identification score of 'silent-centre' syllables used by Strange and her colleagues, since the spectral change which specifies the isolated vowels may survive in the initial and final transitions of /CVC/. This hypothesis was verified in detail by Andruski & Nearey (1992). They made an acoustic analysis of F1 and F2 frequencies of Canadian English monophthongs produced in isolation or in /bVb/ context without a frame sentence, measuring the formant values of two parts, the nucleus and offglide of

/bVb/. (Note that they observed the formant frequency of offglide of /bVb/ at a point of the vocalic part 40 ms before the closure of the final /b/, and those of the nucleus of /bVb/ at a point 40 ms after the release of the initial /b/.) By carrying out t-tests on formant value differences and investigating the formant shift vectors in the F1/F2 diagram from the nucleus to the offglide between each condition, they claimed, "formant movement which is very similar to that found in isolated vowels persists in /bVb/ context." (Andruski & Nearey 1992:395)

Andruski & Nearey (1992) also carried out two perceptual experiments and took a further step to contradict the dynamic theory of vowel perception by Strange et al., whose claim included the perceptual dominance of transitions in coarticulated vowels in /CVC/. They concluded:

> Although no conclusions are drawn about other contexts, for speakers of Western Canadian English coarticulatory cues appear to play at best a minor role in the perception of vowels in /bVb/ context, while vowel-inherent factors dominate listeners' perception. (Andruski & Nearey 1992:390)

It must be noted that this theory of vowel inherent spectral change has a problem as well. One motivation of this theory is the observation by Nearey & Assmann (1986:1305) that judging from spectrograms of /#V#/ and /bVb/, formant patterns similar to /#V#/ are present in /bVb/ context. However, we might ask whether it is really possible to judge patterns of two spectrograms to be similar on the basis of unquantified observations.

An improvement can be recognised in Andruski & Nearey (1992), who scrutinised the vowel inherent spectral change in /bVb/ by a formant frequency analysis of its nuclear and off-glide parts. One could claim that the validity of their conclusion holds on the assumption that what Nearey & Assmann (1986) originally meant by vowel inherent spectral change is diphthongisation and that formant movements of diphthongs can be described by specifying the first and second targets in the dual target hypothesis. Nearey

& Assmann described this as, "two explicit vowel targets, one near the onset and one near the offset of the vowel, are critical [to specify it] ." (Nearey & Assmann 1986:1303) However, the adequacy of this assumption is dependent on the choice of target measurement points in dynamic speech signals. In Andruski & Nearey (1992) the acoustic parameters are the formant frequencies of 'nucleus' and 'offglide', but they admitted, "There exists no well-motivated theory for the selection of optimal time points corresponding to initial and final vowel targets in /bVb/ syllables" (Andruski & Nearey 1992:393), asserting that their choice of parameters was just a working hypothesis. As seen previously, they defined the nucleus as the point 40 ms after the release of /b/ in /bVb/, and the offglide as the point 40 ms before the closure of /b/ in /bVb/. This 40 ms used here, however, was based upon their assumption that consonant-dominated transitions take place largely within 40 ms, which is "guided by choices frequently made for the synthesis of stop + vowel stimuli" (Andruski & Nearey 1992:393). While 40 ms may be prevalent in synthesised perceptual stimuli, this does not logically support the appropriateness of the point for acoustic measurement of normal production, and it certainly gives an impression of being an arbitrary choice. It must be noted as well that Jenkins, Strange & Miranda (1994) reported on the result of their acoustic analysis that the hybrid silent-centre /CVC/ syllables have a distinctly different VISC pattern from the (one-speaker) silent-centre /CVC/, but they were both similarly well identified.

In summary, Nearey and others proposed that vowel inherent spectral change is one important cue in vowel identification. Recently Andruski & Nearey (1992:404) proposed that "listeners may rely on the vowel-inherent cues, rather than coarticulatory cues [supported by Strange et al.], to identify silent center vowels [in /bVb/]" and they claimed that this could be generalised to other contexts.

3.3.4 Huang (1991) and DiBenedetto (1989): F1 averaging vs F1 peak location

In the previous sections, 3.3.1-3.3.3, three hypotheses on perception of vowels with formant undershoot were observed, all of which take into consideration the dynamic formant trajectories. Recently there has been an upsurge of works investigating F1 dynamic trajectories to factor out the role of the F1 dynamic trajectory in vowel perception. This section focuses on two works related to this topic: Huang (1991) and DiBenedetto (1989).

Huang (1991) was concerned with the dynamic information in the /CVC/ formant trajectories, but her approach was to compare the results of a Gaussian classifier with human performance for identification. In Huang (1991) she briefly referred to the F1-average processing. In the section allocated to explain previous studies, Huang mentioned her MA dissertation[11]; "Results [of Huang (1985)] were consistent with a theory of perceptual averaging of F1." (1991:23) However, in that section she did not explain what the "theory of perceptual averaging of F1" is. Later, in describing the alternative input to improve the algorithm, Huang stated, "the third strategy, suggested by evidence for perceptual averaging of F1 from previous studies, ..., was to take an average over the middle 50% of F1 as the F1 value ..." (1991:97) One could assume that the F1 average hypothesis by Huang takes the time-average F1 value from 25% to 75% durational point of the whole /CVC/, and this assumption was confirmed by her later discussions.

This F1 averaging hypothesis appears to have originated from a short report by Stevens (1959), who claimed, "... the perception of a short vowel is determined not by the frequencies reached by the formants but by some time-average values of formant frequencies for the duration of the vowel, including the transition." (Stevens 1959:138) It is inferred that Huang (1985) is the F1-specific version of Stevens (1959).

Huang (1991) referred to DiBenedetto (1987) as a source of a modified version of F1

[11] Huang (1985), her MA dissertation, seems to have dealt with the process in detail, but it is not currently available to the author.

averaging theory. The more recent study by DiBenedetto (1989) extended that modified version, asserting that the results of her two experiments indicated that the simple average value of the F1 trajectory could not explain phonological perception; and "four different stimuli characterized by different trajectories and different durations [and with different F1 average values] were perceived as similar sounds." (DiBenedetto 1989:76). DiBenedetto pointed out that more priority should be accorded to the F1 shape/F1 peak in interpreting the process of the vowel perception. She utilised stimuli with the F1 patterns of 1) a different onset/offset frequency, 2) a different location of a peak (either in the first half or the second half of the trajectory), whose values fell within the American English /i/-/ɪ/-/e/-/ɛ/ continuum, and she concluded, "stimuli with higher F1 onset frequencies and F1 maximum at the beginning of the vocalic portion characterize lower vowels [i.e. /e/ and /ɛ/] ." (DiBenedetto 1989:67).

However, it must be pointed out that in her experiment she did not factor out each of the two parameters, and therefore her conclusion should be: stimuli with higher F1 onset and/or F1 maximum at the beginning of the vocalic portion characterise lower vowels. In addition, her experiments had some more controversial points in method[12].

Finally, observing the 'general' tendency in different stimuli across three types of speakers, DiBenedetto proposed "the weighted average time formant theory", an F1 averaging theory whose weighting function provides more importance to the first half of the F1 trajectory.

---

[12] The problems of DiBenedetto's experiment are: 1) DiBenedetto examined the cross-language difference using three types of listeners (American English/Italian/Japanese) but the number of subjects of each type can hardly be considered sufficient: 4 Americans, 2 Italians and only 1 Japanese; 2) she drew a labelling function of /i/+/ɪ/ vs /e/+/ɛ/ in American English, that of /i/ vs /e/+/ɛ/ in Italian and that of /i/ vs /e/ in Japanese and she made a comparison among three languages, but since American listeners were instructed to label a stimulus as either /i/, /ɪ/, /e/ or /ɛ/, what she did was to group two different perceptual categories, /i/ and /ɪ/, into one; and 3) in one of her labelling experiments, two out of four American listeners identified one type of stimuli (with an F1 peak located on the second half of the trajectory) all as /i/ or /ɪ/, and none as /e/+/ɛ/. Consequently no labelling function was obtained for them, while the results of the other two Americans produced a proper labelling function on the same stimuli. However, she insisted that this would not cause any problem.

This theory of F1 averaging, the F1-exclusive version of the time averaging hypothesis by Stevens (1959), could present a quantified parameter of F1 perception, if the durational middle 50% is sampled to produce the average value. However, this durational middle 50% can possibly be criticised as being arbitrary, like the choice of 40 ms in Andruski & Nearey (1992). The F1 peak location theory, a specific F1 average theory by DiBenedetto (1989), claimed that the simple F1 average value was not sufficient to explain the phonological perception, but since her experiment was not controlled, her theory seems to require further convincing evidence. It would be possible to assume that, as Huang (1991:24) asserts, the results of DiBenedetto "can be accounted for by hypothesizing a weighted average of F1 in which the early portion of the vowel is given more importance than the later portion", but as will be argued later in 3.3.6., van Son (1993), after studying the perception of dichotomised /CVC/ syllables, proposed the importance of the role that the offset of /CVC/ played in perception, which is contradictory to the hypothesis of Huang and DiBenedetto, and without further persuasive and properly controlled experiments, neither of the proposals could be adopted to interpret categorical perception of vowels with dynamic formant trajectories.

In short, Huang (1991) and DiBenedetto (1989) concentrated on the effect of the dynamic F1 trajectory on vowel perception and proposed the F1 time-averaging hypothesis, an F1 specific version of the observation by Stevens (1959).

3.3.5 Koopmans-van Beinum (1980) and van Bergem (1993b) :"reduction"

As was discussed in 3.2, (acoustic) vowel reduction, the traditional way of interpreting vowel formant undershoot, is favoured by the Dutch group, and in this section, papers on perceptual aspect of vowel reduction are discussed: Koopmans-van Beinum (1980) and van Bergem (1993b).

Based upon the acoustic analysis that was mentioned in 3.2.3., Koopmans-van Beinum

(1980) carried out an identification test to investigate whether acoustic vowel 'reduction' affects the identification score and pattern. Her methodology for the perceptual experiment was not well-controlled: she recorded onto a tape the vocalic parts of /CVC/ syllables which were excised in the acoustic analysis, and presented them to 100 listeners. The actual test was an identification test, where listeners were asked to choose the answer to a stimulus written in Dutch orthography. She also provided for subjects the keywords of pronunciation in Dutch to avoid their confusion.

The results indicated that on the whole the vowel segments uttered in isolation were identified "very well", while the vowels extracted from words uttered in isolation showed a high identification score although the number of confusions increased in comparison with the vowels in isolation. The unstressed vowels from a free conversation displayed a very low identifiability.

Koopmans-van Beinum also compared the data obtained from her acoustic analysis and her perceptual studies. First she computed the correlation between duration and perceptual confusion and argued that there was no straightforward relation between the duration of a vowel and its correct identification score. Next, calculating the correlation between the identification scores of three speech conditions and the Acoustic System Contrast in these three conditions, she discovered a high correlation between them. She investigated two conditions: an unstressed vowel in a free conversation and its “reference” (vowels uttered in isolation / words that had a correct identification score of more than 98 %), and she found a high correlation between the acoustic distance and the shift of identification scores between two conditions.

In her conclusion Koopmans-van Beinum asserted that most of the identification response patterns, if they did not agree with the speaker's intention, could be accounted for on the basis of F1 and F2; if the formant frequencies of a vowel V_1 were shifted in an F1/F2 diagram into the region of another vowel V_2, V_1 was basically perceived as V_2, not as V_1. Finally she proposed that the acoustic reduction of the vowels can trigger their perceptual confusion.

Van Bergem (1993b), after his acoustic analysis, also carried out a brief vowel identification test, in order to examine the validity of the hypotheses arising from his acoustic analysis[13].

Like Koopmans-van Beinum (1980), he excised all vowel segments from three speakers' data. Note that he did not mention whether he excised the nuclei or the nuclei + transitions to consonants; and he solely stated that they were "all segmented vowels" (van Bergem 1993b:16). These stimuli were presented to 24 listeners. The task of listeners was identification. They were required to label each stimulus by pushing a keyboard with Dutch orthography.

He labelled the response an "error" if it was different from the speaker's intention. The results indicated that the error pattern across speech conditions resembled that obtained in an acoustic analysis: the error percentages increased in the order of 'stressed'-'unstressed'-'function' and the error percentages were higher for unaccented vowels than for accented vowels, and it therefore endorsed the observation in acoustic analyses discussed in 3.2.5. However, unlike Koopmans-van Beinum, he commented, "Despite the fact that the acoustic reduction can be considerable, ... listeners are yet often capable to correctly identify [vowels]." (p.18), adding that the most errors were due to the confusion of a vowel with neighbouring vowels or long/short counterparts.

It is interesting to note that although both Koopmans-van Beinum (1980) and van Bergem (1993b) presented as a stimulus a segmented vowel from an acoustically reduced environment in a natural speech condition (Speech database in van Bergem (1993b) and

---

13 Chapter 2 of van Bergem (1995) is a reproduction of van Bergem (1993b) and therefore a separate discussion of van Bergem (1995) is not made here.

free conversation in Koopmans-van Beinum (1980)), van Bergem (1993b) reported on its robustness of identification while Koopmans-van Beinum insisted on its fairly low identifiability. The language which they used (Dutch), the task which they required their listeners to do (identification; key symbols/words presented), and the way the stimuli were edited and presented (the vocalic part of /CVC/), were all identical in both experiments, and hence the reason for their diversity still remains to be explained.

Another point is that since they investigated the influence of manipulation of suprasegmental/lexical/pragmatic level (stress/sentential accent/speech style) on vowel undershoot, or what they call 'reduction', they could not properly control the parameters of the speech data in their experiment, particularly in Koopmans-van Beinum (1980). An improvement is observed in van Bergem (1993b) but as was discussed it is not satisfactory.

3.3.6 van Son (1993) : trajectory and phonetic context

The principal object of van Son (1993) was to investigate whether listeners use information from the formant track shape to decide on the vowel identity and whether vowel duration and consonantal context influence this decision and therefore devised an original and controlled experiment. He created two conditions for the identification of synthesised static/dynamic Dutch vowels: in condition i), stimulus vowels were presented in isolation, and notably, in condition ii), stimulus vowels were presented in /C_1VC_2/ (C_1 & C_2 are different, and either /n/ or /f/). However these vowels were synthesised separately from the neighbouring consonants and three of them just concatenated. The object of this quasi-natural but well-controlled /CVC/ condition was to detect the pure effect of the presence of consonantal context on vowel identifications.

In the first experiment dealing with the vowels presented in isolation, van Son constructed two stimulus types: tokens with F1 and F2 static, and tokens with either F1 or F2 dynamic. The vowels used were /i u y ɪ o ɛ ɑ œ/, whose formant values were taken from Koopmans-van Beinum (1980).

He utilised a parabola function to calculate the dynamic F1/F2 trajectories[14]. The types of tokens presented in isolation are as follows. DF1 represents the dynamic trajectory range of F1, and DF2 that of F2.

1) vowels with steady F1/F2 (static tokens)

2) vowels whose F1 or F2 is dynamic and whose trajectory ranges are; DF1= +/- 225 Hz and DF2= +/- 335 Hz (original dynamic tokens)

3) vowels whose F1 or F2 is dynamic but whose trajectory ranges are obtained by referring to the acoustic observation ("realistic" dynamic tokens[15])

4) vowels which are created by dividing original dynamic tokens into halves in the time axis (onset tokens: the first half, and offset tokens: the latter half)

The values of the dynamic trajectory range in original dynamic tokens were provided by the result of his acoustic analysis on dynamic /a u i/ tokens. To create onset / offset tokens, he only utilised the original dynamic tokens. The task of the listeners was to identify a stimulus from the choices written orthographically on an answer sheet.

The results of the static tokens showed that if the duration was not less than 25 ms, the tokens were identified correctly or as the other counterpart of the long-short pair, while if the duration was less than 25 ms, more confusions were observed.

In analysing the responses to original dynamic tokens(= token type 2), he introduced the

[14] The formula, which he applied to both F1 and F2, is as follows: for a given time t (ms), the formant frequency Fn(t) is given; $Fn(t) = Tgt - DFn \{ 4 (t/dur)^2 - 4 (t/dur) +1 \}$, where Tgt= "target"(=nuclei) frequency value (Hz), DFn=the excursion size, (the range of the dynamic trajectory range, obtained by [nucleus frequency] - [onset/offset frequency]), dur=duration of vowel (in ms).

[15] The trajectory ranges that he used are: /y/ (DF1, DF2)=(0,225); /ɪ/ (DF1, DF2)=(75, 225); /o/ (DF1, DF2)=(75, -225); /ɛ/ (DF1, DF2)= (150, 150); /ɑ/ (DF1, DF2)= (150, -75); /œ/ (DF1, DF2) =(75,75). All numbers in Hz.

index called "net shift" to quantify the changes that occurred. It was provided, with a reference to a vowel ranking obtained according to a formant frequency (like /i/-/ɪ/-/e/ ... in F1), by computing the percentage of responses to a dynamic token which shift to a certain direction in the ranking from a response to a static token[16].

Using this net shift, he observed on original dynamic tokens that 1) token duration had only a slight effect on the F1 shift index, while no effect was observed on the F2 shift index; 2) the directions of these shifts were towards the onset/offset, which was more prominent in F1 than F2; 3) the degrees of the net shift were different from those obtained from offset/onset only tokens: except for tokens with DF2=375 Hz, offglide-only tokens indicated the greatest shift, onglide-only tokens the least, and the complete (original) dynamic tokens between them.

Van Son applied this index method to the evaluation of "realistic" tokens, forming pairs of a "realistic" token and its corresponding static token. While he was unable to discover in F2 any apparent relation between trajectory size and net shift, he could find a net shift of F1 towards the lower rank tokens (i.e. shifts towards the onset/offset).

In the other experimental condition (vowels presented in /CVC/), van Son reduced the number of vowels tested, by utilising vowel types /ɑ ɪ ɛ o/ with duration 50ms/100ms only. The consonants, /n/ and /f/ were synthesised and vowels were original dynamic tokens used in the previous conditions (i.e. DF1= +/- 225 Hz, DF2= +/- 375 Hz). The

---

[16] The procedure to obtain this index is; (i) He ranked the vowels according to their F1 or F2 frequencies. This produced two rankings: one in the order of F1 (low to high; /i y u ɪ e ø o œ ɔ ɛ ɑ a/), the other in the order to F2 (low to high; /u o ɔ ɑ a œ ø y ɛ e ɪ i/). Each ranking was used on the corresponding data evaluation, i.e. F1 ranking to evaluate F1, etc. (ii) Then he created pairs of a dynamic token and a static token with the same duration and the same vowel label experimented on the same subject. (iii) Next he scored the shift of the labels from a static token (as a reference) to a dynamic token, by observing whether the label of a dynamic token was lower/higher in the ranking in (i), and assigning a positive sign if the it is higher in ranking, therefore in formant frequency. (iv) Finally he pooled these signs obtained by iii) across all subjects, and the net shift was calculated as percentage by {(number of occurrences of the majority shift)-(number of shifts in the other direction)} / (total number of pairs)

static tokens, with the duration of 50 ms/100 ms, were also presented in /CVC/. Like the first experiment, the onset/offset tokens were produced in the /CVC/ condition, by dividing the whole /CVC/ token into two at the durational middle point. All the vowels stated above were presented in isolation as well.

15 Dutch speakers participated in this experiment. Their task was an open-response paradigm; "to write down orthographically whatever they heard" (1993:77) To interpret the results, the net shift was obtained from a pair of a token presented in isolation (used in this experiment) and a token in /CVC/ or /CV#/ (onset) or /#VC/ (offset). Comparing the two experimental conditions, van Son argued that the size of the net shift in this experiment (i.e. /CVC/+/CV#/+/#VC/ conditions) was smaller than the previous experiment, but the pattern was more or less the same for both durational conditions. He also remarked that the off-glide tokens in this experiment, like those in the previous experiment, indicated a greater shift towards the offset than the tokens with complete trajectories.

He concluded, "In context, the responses to the vowel tokens were essentially the same as in isolation." (p.89), which was a response to the initial question of whether consonantal contexts would affect listeners' decisions. This conclusion holds, however, only in a special environment where the coarticulation between a consonant and a vowel would not occur, since van Son (1993) synthesised the stimuli by concatenating three acoustic segments without coarticulatory smoothing of formants.

Another point is that the reliability of his experiment rests upon the validity of the index "net shift" introduced by van Son to quantify the response pattern of subjects. As concisely mentioned in the text above (and described in the footnote in detail) the net shift concerns the number of stimuli shifts to a certain direction in a ranking, but it does not concern the magnitude of a stimuli shift in a ranking: for example, it does not take into account whether a response to a stimulus shifts two steps down to a low F1 or five steps down. This indicates that the net shift is not a sensitive analysis that can be made on data.

Like DiBenedetto and Huang, van Son (1993) seems to support the time-averaging process of dynamic formant perception:

> The "perceptual-undershoot" found in both experiments suggests some kind of averaging of the formant frequency inside the tokens. The largest shifts were found in offglide-only tokens, followed by tokens with complete formant tracks ... " (van Son 1993:91)

Furthermore, as was briefly mentioned in Section 3.3.4, van Son insisted on the importance of the offset of /CVC/ in perception, contrary to DiBenedetto: "Tokens with curved formant tracks were generally identified near their formant offset frequencies" (van Son 1993:69), or "However, compound target-models [= dual-target theory] assume that listeners use the vowel kernel or nucleus to identify it. In our study listeners used the offset part." (van Son 1993:125-126).

Here it must be remarked that van Son asserted this idea solely on the basis of the fact that the "net-shift" was the greatest in offset tokens: the fact that the net shift was greater than in original dynamic tokens and in onset tokens. Even though the offset tokens indicated greater shift in F1/F2 ranking, it is evident that the dominance of the role of the offset in vowel perception does not logically follow from it. The hypothesis that listeners use the offset of /CVC/ to perceive the vowel requires further verification, while the time averaging theory would be worth scrutinising as to whether it could be applicable to vowel quality evaluation.

3.3.7 Rosner & Pickering (1994)

After an extensive review of the literature on vowel perception and production, Rosner & Pickering (1994) put forward arguments for an auditory theory of vowel perception, a theory which aims to account for vowel perception in a formal manner. This section first reviews their general auditory model of vowel processing which uses a spectrum obtained from the temporal mid-point of a vowel. This review is then followed by a description of their discussion of the effects of coarticulation on perception, along with their proposal of an auditory perception model.

In the model used by Rosner & Pickering, special consideration is given to the spectral characteristics and the duration of a given production of a vowel. This model was designed to address the problem of categorizing vowels. They noted in their work that the model was incomplete in certain respects and that it borrowed heavily from work by numerous investigators. The first simplified version of the model was divided into five stages.

Stage I: ERB transformation
An ERB-rate auditory transform E(f) is applied to the frequency axis of the spectrum of a vowel.

Stage II: Filter bank
The resulting spectrum is then smoothed by a bank of auditory filters, where filter bandwidths increase proportionately with E(f)

Stage III: Suppression
The peaks of the excitation pattern of the spectrum were then sharpened and the valleys deepened.

Stage IV: Intensity transformation
The intensity axis is transformed by L(I) to an auditory loudness density pattern (ALDP).

Stage V: Second function and peak / shoulder picking
The output of Stage IV, the auditory loudness density pattern, goes through an integrating function (AI function), which performs a weighted integration over a local region in the spectrum. After this integration, the locations of the first two peaks, E1 and E2, are determined in ERB scale, which specify the location of the vowel in an auditory vowel space. Finally, one of the vowel prototypes is chosen

on the basis of the E1 and E2 values. The shoulder[17] picking process in the spectrum is required in order to deal with vowels where two spectral peaks appear close together.

Later Rosner and Pickering reviewed the perceptual effect of various stimulus features other than formant centre frequencies, and although they claimed that no major changes were required in the model described above, it became necessary to extend it in the following three ways: (1) E3, as well as E1 and E2, must be introduced to explain data on rhotacised and nasalised vowels, (2) duration as a perceptual cue must be considered, and (3) to account for the contrast and assimilation effect in vowel perception, the auditory vowel space should possibly refer to signal detection theory or criterion setting theory.

The first version of their model above, which relies upon a single spectrum, does not explain the perceptual salience of coarticulated vowels for which spectral peaks vary in different phonetic environments. Chapter 6 of Rosner & Pickering contains a detailed review of the previous works related to extensions of this model. After a review of the literature and of their own data, Rosner & Pickering concluded that the "nearest prototype rule [i.e. to find the nearest prototype in the auditory vowel space, as shown in Stage V above] must fail if it depends exclusively" on the spectral peaks or shoulders in Stage V that are computed around the middle of the particular realisation of a vowel. (p. 275) Therefore, in order to maintain the nearest prototype and the underlying concept of an auditory prototype, they introduced an auditory space function (ASP-function) to describe the auditory path that represents a particular production of a vowel. The function integrates over the E1/E2 values along the dynamic trajectory that comprise a

[17] Spectral shoulders are defined as irregularities where two formants merge to form a single peak in a spectrum or to form one intense formant peak whose skirt has a manifestation of a less intense formant. Both cases lacks two clear peak points in the spectrum. Rosner & Pickering (1994:133) suggested that they could apply to their theory the shoulder-picking method by Assmann & Summerfield (1989), which determines shoulders of a spectrum by finding the zero crossings of the third differential in the excitation pattern envelope.

vowel's path in the acoustic vowel space, which feeds the nearest prototype rule.

After their review of previous investigations of vowel inherent spectral change (see 3.3.3 for the description of vowel inherent spectral change) and other topics, they argued that the ASP-function should take a domain consisting of the whole of the vowel path but that it should use the weighted values along the path.

They also described the following characteristics of the ASP-function: (1) its "window should have sloping skirts, attaching less importance to information further removed in time from the current position of the centre of the window " (Rosner & Pickering 1994:330); (2) "the window must be negatively skewed, so that anticipatory information gets greater emphasis than carryover information" (Rosner & Pickering 1994:330); (3) the weighting should emphasise vowel inherent spectral change; (4) the function should generate a running average, in order to be compatible with continuous uptake of information. However, it must be noted that they stated that there are three further important properties of the ASP-function unspecified[18], and they did not present any concrete algorithm for this function.

3.3.8 Discussion of categorical perception studies

The sections 3.3.1-3.3.7 reviewed the previous studies concerning phonological perception of vowels with formant undershoot. It was seen that many alternative theories were proposed. Some of them criticised the theory of Lindblom and Studdert-Kennedy and a provided more specific interpretation of the perceptual overshoot effect.

Since the object of this thesis is to discover how dynamic formants can alter the auditory impression of a vowel, it is not necessary to be deeply involved in the nature of

[18] They are; (1) whether the same function operates on three local effective vowel indicators; (2) the window width of the ASP function; and (3) the way the ASP-function treats unvoiced consonants.

phonological perception, but it is important to consider the implication of phonological perception models on phonetic vowel quality evaluation.

Considering the arguments in the previous section, it appears that a single-target compensation process might be inadequate to explain vowel perception. However there are some recent studies which persist with the single target-based approach. For example, Miller (1989) proposes the F0 normalised target-based model where a vowel is represented by a path in auditory-perceptual three-dimensional space, whose three axes are defined by a logarithmic calibration of the ratios of the first three formant frequencies transformed by the speaker's average F0. As an acoustic reference point to measure formant frequency shift, a single acoustic nucleus measurement provides a single parameter which must contribute to vowel identification to some extent. If one can deal with a properly quantified value that represents vowel formant shift and provides a scale according to which a psychoacoustic experiment is evaluated, then there seems no reason to avoid its use.

The adoption of this single target approach does not mean that we should ignore all the dynamic-information based theories. It seems that the time averaging of a dynamic formant trajectory provides a quantified value which could be used as a model of vowel quality perception. Thus the time-averaging process can be adopted as a testable hypothesis for vowel quality evaluation.

The ASP function, proposed by Rosner & Pickering (1994), which calculates the weighted average of the values that comprise a vowel's path in the acoustic vowel space and returns a single set of values, could also provide a quantified value. Regrettably, Rosner & Pickering did not offer any concrete averaging algorithm to compute the integrated value of a vowel's path. However, as the best compromise, they presented a time integrating weighted function by Kuwabara (1985), which they claim could be transformed to this algorithm. This Kuwabara function is discussed in Chapter 4.

3.4 Phonetic/psychoacoustic perception of vowels with formant undershoot: how dynamic formant trajectory gives a different auditory impression

3.4.1 Pols, Boxelaar & Koopmans-van Beinum (1984): Matching experiment

The phonological perception of vowels with undershoot has been a major issue in phonetics; what helps listeners to identify a segment whose acoustic attributes are different from those found in isolation? However, since the experiments dealing with that issue are concerned with the listener's strategy to cope with the variance of the acoustic realisation, they do not explore finer vowel categories than those in the phonology. Pols, Boxelaar & Koopmans-van Beinum (1984), following this argument, carried out a matching experiment on dynamic F2 trajectories that is "situated halfway between psychophysics and speech perception." (p.374)

They first emphasised the necessity of the matching process, by claiming, "... one of the drawbacks [of previous perceptual experiments] is that the listener could only specify his percept in terms of a limited number of predefined categories ... " and therefore they introduced a matching paradigm to "elicit more finely-grained judgments." (p.372)

As a stimulus, four-formant vowel-like signals were synthesised according to the following parameter specifications: 1) three formants out of four remained static during the whole duration of 200 ms while the remaining one, either F1 or F2, changed its frequency in the final 100 ms; 2) their F0 declined "continuously" (presumably linearly) from 110 to 90 Hz; and 3) the dynamic trajectory of changing F1/F2 was a quarter sine function.

Pols, Boxelaar & Koopmans-van Beinum chose an example of one trajectory pattern, whose F2 changes from the region of /ɛ/ or /œ/ to that of /ɑ/. The rest of the stimulus patterns were discussed in the final part of their text. It is assumed from the examples that the end point of the formant trajectories changed in steps of 100 Hz, although it is not stated in the text. All stimuli patterns are shown below.

F1 change

from 240 Hz to 400-600 Hz (F2 constant at 800 Hz)

F2 change

from 700 Hz to 1100-1800 Hz (F1 constant at 220 Hz)

from 1000 Hz to 1400-2200 Hz (F1 constant at 450 Hz)

from 1800 Hz to 900-1300 Hz (F1 constant at 450 Hz)

from 1000 Hz to 1400-2200 Hz (F1 constant at 650 Hz)

from 1800 Hz to 900-1300 Hz (F1 constant at 650 Hz)

For matching, each of these stimuli was presented, followed, after an interval of 500ms, by its comparison vowel. This latter vowel could be interactively modified by subjects. The comparison vowel had four static formants with 70ms duration. Subjects could vary the formant frequency by 32 logarithmic steps in frequency, although the actual step values and the range of the frequency variation was not stated in the paper. The whole 32 choices were displayed as a horizontal scale with 32 points. By moving a cursor on the scale, the subjects could alter one formant frequency of the static comparison vowel. In one session, every stimulus (i.e. a dynamic vowel) was presented twice in one set of 20 stimuli, and a subject sat in for two sessions. 12 subjects participated in this experiment.

Not all the results of the experiment were presented; Pols, Boxelaar & Koopmans-van Beinum only referred to those of the two stimulus sets: F2 from 1800 Hz to 900-1300 Hz with F1 = 650 Hz, and F2 from 1800 to 900-1300 Hz with F1 = 450 Hz. They calculated the difference between the end frequency of the dynamic trajectory and the matched frequency, and plotted the results of the two shift patterns. They demonstrated that both shift patterns shared . regular characteristics with regard to the difference between the endpoint of the trajectory and the matched frequency: if the stimulus had the smaller shift (e.g. 1800 Hz to 1300 Hz), the result indicated a tendency for the matched frequency to be lower than the trajectory endpoint, while if the stimulus had the larger sweep (e.g. 1800 Hz to 900 Hz), the result indicated a tendency for the matched frequency to be higher than the trajectory endpoint. The dissimilarity between the two stimulus sets is only that "the absolute deviation from the final F2 frequency [i.e. the difference between the matched F2 frequency and the trajectory endpoint] is smaller" (p.376) in the shift patterns of F1=450. For the other stimulus sets, they simply reported that the data showed "substantial individual differences although an overall pattern is to be seen", being "unable to evaluate all these data and their implications." (p.377)

Thus Pols, Boxelaar & Koopmans-van Beinum failed to deliver clear results from their data and claimed deficiency of data to verify whether the phenomenon reported on was caused by the underlying human perceptual mechanism or the experimentation.

3.4.2 Nord (1986): grid experiment

Nord (1986), after analysing the acoustic difference in undershoot of Swedish stressed/unstressed vowels in word-initial/word-final syllable position, set up a perceptual experiment to test "the perceptual significance of the results from the acoustic analysis by means of an interactive matching paradigm." (Nord 1986:23). Eight subjects participated in his experiment. His matching paradigm required them to listen to synthetic two-syllable words and to adjust the vowel in order to make each word sound as natural as possible. The vowel was tested in both stressed and unstressed /CVC/ syllables embedded in word-initial and word-final position. The subjects could change the quality of a tested vowel by moving a cursor around in the grid appearing on a display using a joystick. The grid was an acoustic F1/F2 vowel diagram.

He claimed that this task is quite manageable with phonetically untrained subjects, while in his actual test he used phonetically trained subjects to reduce the variability. He investigated four Swedish front vowels,but each subject had only two trials on a set of 25 words since he limited the number of words "as the test was rather exhausting." (p.30)

When interpreting the results, Nord plotted the matched results of the tested vowels on an F1/F2 diagram, and found that the result of the matching experiment showed a similar, although less obvious, tendency as the production experiment: irrespective of their duration, unstressed vowels coarticulate with their contexts more strongly than stressed ones and especially in a word final position they coarticulate with the "neutral articulatory position", which Nord claims to correspond to "a centralised schwa vowel".

The interesting aspect of this experiment is the interactive matching scheme using a grid, since this, as Nord suggested, could be "an efficient way of evaluating perceptual cues and testing theories of speech dynamics". (p.34) An interactive matching scheme could be an effective alternative to the normal identification / discrimination test in investigating phonetic / psychoacoustic properties, although a prudent implementation will be required.

3.4.3 Discussion of psychoacoustic perception studies

The previous sections 3.4.1-3.4.2 reviewed the papers on psychoacoustic/phonetic studies of dynamic vowel formant perception, and this section discusses their relevance with respect to the aims of this study.

The constraint posed by categorical testing , as Pols, Boxelaar & Koopmans-van Beinum (1984) suggested in their introduction, is a central issue in this study. An ordinary identification or labelling technique would only supply listeners with pre-defined labels (phonemes) to attach to stimuli, which could mean that they could not make a quality judgement on a token which falls perceptually between two pre-determined categories. In this respect the experiment by Pols, Boxelaar & Koopmans-van Beinum is of much interest. However, their experiment, as was discussed, could not offer any clear results, since there was an insufficient number of stimulus presentations. Furthermore they did not present the results of all tokens; they simply gave the examples of only two vowels, without referring to the results of the others. Also, since Pols, Boxelaar & Koopmans-van Beinum investigated only one-formant changes, this study should pursue the possibility of discovering how listeners evaluate the quality of a vowel stimulus whose two formants (F1 and F2) draw a dynamic trajectory.

With regard to the testing of vowel quality evaluation using a vowel whose two formants are dynamic, it has been argued that the normal categorical labelling test would not be of any help. The similarity judgement test, or the paired triad comparison test (presenting stimuli as X-A+X-B) which require a listener to judge the similarity of the stimuli, would impose a significant burden on listeners since the number of the stimuli presented could be enormous if two independent parameters, F1 and F2, are involved.

Therefore, it is necessary to adopt the grid-matching scheme from the study of Nord (1986), since this interactive mode of the experiment can give listeners visual feedback, making the task manageable. Moreover, the grid-matching scheme in Nord (1986) allows a listener to respond to two parameters simultaneously by allocating each parameter to each dimension in a grid, which saves time and effort on the part of a listener by reducing the number of stimuli, and in addition, enabling the investigation of any interaction between parameters.

3.5 Concluding Remarks

Sections 3.2 to 3.4 surveyed the history and current trends on vowel undershoot studies. This section summarises them and proposes some issues to be covered by the following experiments.

Section 3.2 reviewed the acoustic analyses made by previous studies that investigated undershoot. Various factors were observed to have an influence on undershoot: speech rate; coarticulation; word-stress; and pragmatic factors (whether the utterance is made in text reading / free conversation etc.). However the experiment of this study concentrates on the parameters of formant undershoot due to coarticulation with neighbouring consonants, in order to create well-controlled stimuli for a perceptual experiment.

Section 3.3 examined the phonological perception of vowels showing undershoot. The studies proposed alternatives to the traditional target compensation model, and suggested that temporal averaging as well as the ASP-function could provide a quantified value to model phonetic vowel quality evaluation. Other models of phonological perception of vowels with undershoot were not appropriate to incorporate in this study since they did not provide concrete algorithms.

Section 3.4 surveyed the phonetic / psychoacoustic process of vowel quality evaluation with dynamic formant trajectories and it was found that an interactive matching paradigm that contrasts vowels in a /CVC/ environment with steady state vowels could be a useful method for the investigation of the perceived vowel quality produced by dynamic formant changes.

Taking into consideration these points, the next chapter examines whether an interactive matching scheme, proposed by Nord and modified to serve the purpose of this study, could cater for the investigation of the perceived quality of vowels showing formant undershoot. A paired triad comparison experiment, which presents two stimulus pairs of X-A and X-B and asks its subjects to choose which pair is closer in vowel quality (henceforth called an XAXB experiment), is carried out, and compared with an interactive matching design.

Chapter 4: Pilot Experiments: verification of the grid matching scheme

4.1 Introduction

Chapter 3 reviewed how vowel formant undershoot has been investigated using a variety of experimental techniques. In 3.4 it was suggested that an interactive matching technique could be a suitable method for such studies, and specifically a grid matching scheme was proposed. This chapter describes the pilot studies which eventually led to the main experimental design to be presented in Chapter 5.

4.2 Pilot Experiment 1 (preliminary XAXB matching experiment)

4.2.1 Introduction

In the previous chapter, it was mentioned that the phonetic or psychoacoustic evaluation of vowel quality in undershoot vowels has not been thoroughly investigated and experimental data is rather limited. As a consequence it is impossible to see whether the phonological perceptual overshoot hypothesis (i.e., perceptual target compensation hypothesis) proposed by Lindblom & Studdert-Kennedy (1967) can be applied to the quality evaluation of a vowel with dynamic formant trajectories. Perhaps listeners can compensate for the 'target' formant frequencies which can be obtained when these vowels are uttered in isolation. Alternatively, it is also possible that listeners simply pay attention to the peak of the dynamic formant trajectories and use these peak values to determine vowel quality.

Due to the lack of the data in previous studies, it is difficult to reject claims that the XAXB matching paradigm would work satisfactorily for this kind of experiment, although it was argued in 3.4 that such a scheme would be less effective than grid matching. Considering these points, we present a preliminary XAXB test on quality evaluation of a dynamic formant trajectory designed to simulate a /CVC/ context. In other words, this preparatory experiment studied whether the vowel quality of dynamic /CVC/ stimuli, whose vowel was subject to acoustic undershoot, could be evaluated by a simple similarity judgement: whether X (=/CVC/) was closer to A or B in an X-A X-B pair presentation, where A and B were vowels in isolation with steady trajectories.

4.2.2 Materials

The materials of this experiment were constructed in the following way: first, the reference material X was synthesised. This reference material consisted of a vowel with a dynamic trajectory simulating the /CVC/ pattern according to the formula devised by Nearey (1989). The consonants /b/ and /d/ were selected, in order to investigate the effect of the F2 trajectory direction on perception. (F2 trajectory is concave in /dVd/ while it is convex in /bVb/.)

For vowels in /CVC/, of all RP short monophthongs, the four vowels /ɛ,æ,ɒ,ʊ/ were selected, because in an acoustic (F1/F2) vowel diagram they are well separated from each other. /ɪ/ was not selected since its inclusion would change the pattern of F2 trajectory in /dVd/; it would make both trajectories concave. The actual formula followed Nearey (1989), according to which the formant trajectory was synthesised as follows:

> If one defines t as relative time within /CVC/, (i.e. t = 0 at the release of initial consonant), then formant frequency Fm(t), when t is between 0 and the midpoint of /CVC/, is :

$Fm(t) = Fv + (Fi - Fv)[(t-Tv)^6/Tv^6]$ [Hz]
where,
Fi; consonantal loci [Hz]
for /bVb/, (F1, F2)= (150,700)
for /dVd/, (F1, F2)= (150,2000)
Tv: durational midpoint of a stimulus [ms]
Fv: frequencies in the steady part of nuclei [Hz]

The total duration of this /CVC/ and the formant frequencies of /CVC/ syllable nuclei were obtained from an acoustic analysis that investigated the F1/F2 frequencies of the RP short monophthongs produced by one male RP speaker in the following three conditions: (1) in isolation; (2) in /CVC/ (C=/b,d,g/) uttered in isolation; and (3) in /CVC/ embedded in two positions within a frame sentence, "The word is '___'; so repeat '___' slowly." In the condition (3) the second token was analysed since it was a repetition of the first token, behaving like a pronoun, and therefore receiving minimum sentence stress. Ten repetitions of each token were taken per condition. In the formant analysis of each vowel, either entire /CVC/ nuclei with static formants or three central sampling points in /CVC/, whichever the longer in duration, were segmented and the average F1/F2 frequencies were obtained in each token type and condition. Then all the formant values were converted into Bark scale and plotted in F1/F2 diagrams. The results (except those of /ɒ/, which does not exist in American English) are in general agreement with the observations made of American English by Stevens & House (1963)[19].

The input formant values for Nearey's formula were obtained from CVC syllables uttered in isolation (condition 2) in this pilot acoustic experiment. They are shown in Table 4-1.

	ɛ	æ	ɒ	ʊ
/bVb/	542/1806	760/1621	545/1009	491/1122
/dVd/	526/1848	733/1619	565/1135	437/1359

Table 4-1: Input peak values obtained from the pilot acoustic studies. (The values indicated are "F1/F2") All values in Hz.

The mean duration of /CVC/ in condition (2) was 121 ms, and therefore, the duration of the synthetic /CVC/ was set to 120 ms. (i.e. the durational midpoint in Nearey's formula, Tv, was 60 ms) Vowels in these /CVC/ syllables had only two formants, F1 and F2, and they were synthesised using the parallel formant JSRU synthesiser by Holmes (1982) implemented in the Speech Filing System running on a Sun SPARC workstation. In the synthesis, F0 declined linearly from 130 Hz to 100 Hz, and every 10 ms point was interpolated linearly. Voicing of the token started at 10 ms after the initial point, reaching the full amplitude at 20 ms, and it started to decline 20 ms before the final point, ceasing completely 10 ms before the end. Between these 20 ms turning points, the formant amplitude was kept constant. F1 and F2 intensities were fixed to 50 dB during the voicing excitation. To ensure the reliability of the experiment, the actual formant frequencies of the output /CVC/ syllables were confirmed by obtaining a spectral section of the durational midpoint and measuring the formant centre frequencies, using Kay Sonagraph Model 5500.

To address a potential criticism that the two-formant stimuli may not be easily identifiable or natural, three native speakers of South-East British English with phonetic training were asked to judge all types of the synthesised /CVC/ tokens by listening to them through headphones twice. The subjects all agreed that the stimuli all had "acceptable quality of synthesised speech" although two of them remarked on the unnaturalness of the /d/ in /dVd/ tokens synthesised according to Nearey's formula. Subsequently, the /dVd/ tokens in this experiment were modified by the addition of an initial intense burst and a final voiceless release, which improved the naturalness of the

/d/ segments. This modification was accepted positively in the second informal survey on the stimulus quality. Spectrograms of examples in the final version are shown in Figure 4-1.

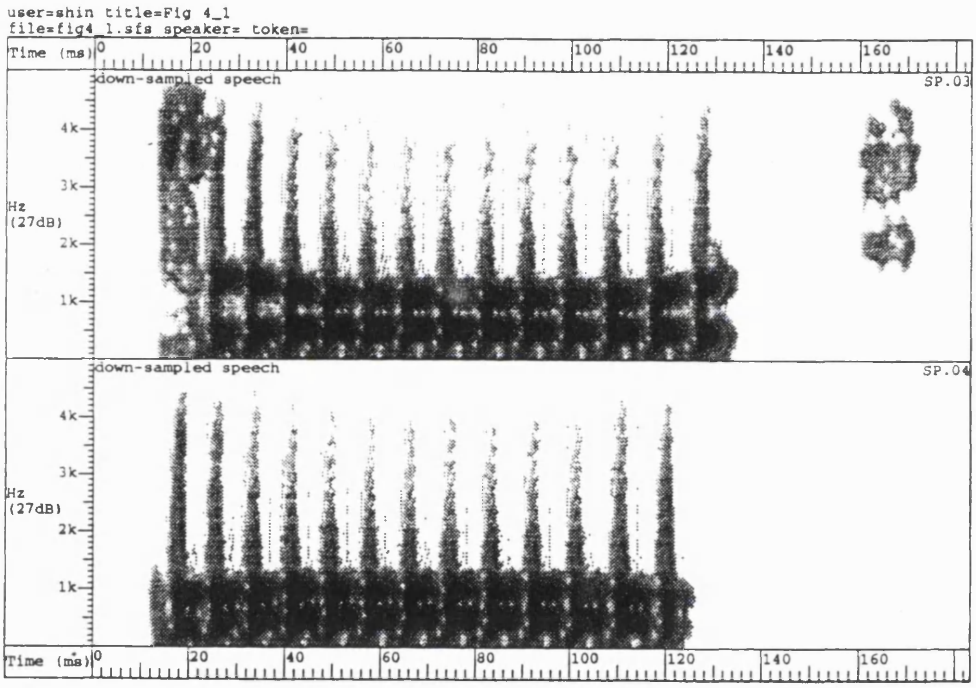


Figure 4-1: Spectrograms of /dɒd/ (above) and /bɒb/ (below) used in Pilot Experiment 1

The other type of test stimulus, A or B in XAXB test, was /#V#/, which had only two formants F1/F2 like the /CVC/ tokens, although these formants were static. Its F0 pattern was identical: linear declination from 130 Hz to 100 Hz, and its duration was 120 ms. In the initial 10 ms the voicing increased gradually and in the final 10 ms it declined to avoid providing a clipped auditory impression. In accordance with the hypothesis described above, two types of the /#V#/ tokens were synthesised: (1) vowels in isolation whose static F1/F2 frequencies were identical with the peak values of /CVC/ tokens (henceforth called peak tokens), and (2) vowels in isolation whose static F1/F2 frequencies were found from the previous acoustic observation on the same vowel types /ɛ,æ,ɒ,ʊ/ (i.e. each vowel token was synthesised using the result of the pilot acoustic study on the same vowel in isolation, produced by the identical speaker; henceforth called iso-target tokens). The values used for the iso-target tokens, which come from the results of condition (1) in the pilot acoustic experiment, are shown in Table 4-2 below.

	ɛ	æ	ɒ	ʊ
F1	677	847	629	421
F2	1945	1665	976	1066

Table 4-2: F1/F2 values for iso-target tokens. All values in Hz.

Therefore, for each vowel /ɛ,æ,ɒ,ʊ/ embedded in CVC, there were two choices as a vowel quality match: a peak token and an iso-target token. Figure 4-2 is a schematic figure of this setting.

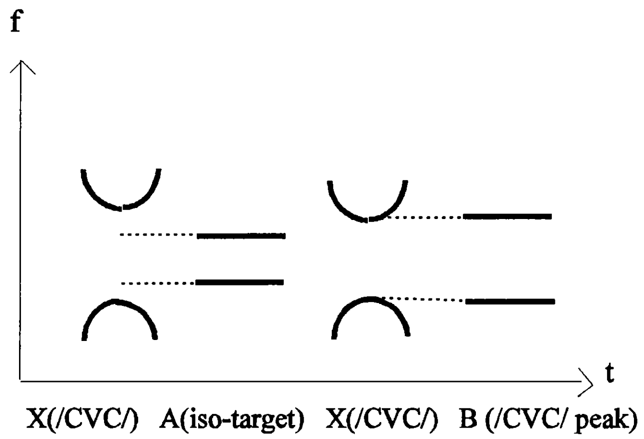


Figure 4-2: Stimulus pairs presented to subjects. They are asked to choose whether the pair XA or XB has a closer vowel quality

4.2.3 Subjects

Five speakers of South-East British English, four males and one female, participated in this experiment and none of them had a history of hearing defects nor suffered from any such defects at the time of the experiment. Three of them were researchers in the Department of Phonetics, University College London, while two were postgraduate students of the department.

4.2.4 Procedure

For each type of four vowels, two sets of two pairs, XA-XB, were created, where X was /CVC/ and A was a peak token corresponding to V while B was an iso-target token corresponding to V, or vice versa. When they were synthesised, the intervals within pairs (i.e. between X and A and between X and B) were 120 ms, and the interval between two pairs (i.e. between XA and XB) was 240 ms. The interval between each XAXB was 2 seconds. Each stimulus pair was presented a total of five times, producing a total of 80 sets of XAXB pairs (2 consonant types x 4 vowel types x 2 combinations consisting of peak tokens of either A or B in XAXB x 5 repetitions), and these 80 sets were preceded by the five trial sets which were meant to make listeners familiar with the experimental setting and the nature of the stimuli. After these five trial sets and every 10 sets, a long pause of five seconds and a beep were inserted. These sets were recorded in a random order onto a DAT tape through a low-pass filter of 5 kHz.

The task of the subjects was to listen to the X-A+X-B set of pairs and to judge which of the pairs had the closer vowel quality. They were asked to tick a box corresponding to the selected pair on an answer sheet. The test was carried out in a sound-proof room and the sound was played out through a loudspeaker at a comfortable level. Only one subject was tested at a time. The whole test took around 20 minutes, including the time for instruction and five trial sets.

4.2.5 Results and Discussion

The responses of the subjects were analysed in the following way: for each subject, the responses were accumulated and then for each /CVC/ type the number of peak responses and the number of iso-target responses were counted. These calculated numbers are presented as tables in Appendix I.

At this stage, before accumulating data across subjects, it is necessary to examine the possible interactions between subjects and conditions within this experiment, since different subjects may have reacted in different ways to the experiment. Since the variable in this case is not numerical but categorical, a formal assessment of the interactions was done with a log-linear model for the XAXB experiment. In performing this analysis, Version 6.1.3 of the SPSS for Windows statistical software package was used.

Accordingly, Model Selection Log-linear analysis was made on the data of this experiment, with factors of [vowel], [consonant], and [subject]. The optimal model that should fit the data and be substantively interpretable was selected by a procedure called backward elimination, which starts with all effects in a model then removes those whose elimination results in the least significant change in the likelihood-ratio chi-square. (See Norusis 1994:145-170)

The analysis resulted in the creation of a final model which showed the interaction of subjects: an interaction of [consonant]*[subject]*[choice of peak/iso-target token], and that of [vowel]*[subject]*[choice of peak/iso-target token]. The final model with its predicted counts is shown in Appendix II. This clearly shows that the group of Subjects 1-5 in this experiment was not homogenous, thus making it necessary to create subgroups of subjects to obtain homogenous data.

As the next step, all the data were re-examined and an analysis of these data showed that, of all Subjects 1-5, Subjects 2 and 3 appeared to behave in a different way from the others. Then separate log-linear analyses were done on two subgroups of subjects to secure homogeneity amongst these subjects, following the model-selection procedure described above. The final models obtained from this log-linear analysis, shown in detail in Appendix III, indicate that the factor [subject] was not significant in these two subgroups. The final models also suggest that each subgroup had a distinctive response pattern.

The final model of the subject group consisting of subjects 1, 4 and 5 produced the interaction of [consonant]*[choice of peak/iso-target token], and its expected counts displayed a discernible preference for peak tokens over iso-target tokens, although the preference appeared to be more prominent in /dVd/, which corresponds to the consonantal effect in the final model.

On the other hand, the subject group consisting of subjects 2 and 3 had a final model of log-linear analysis whose expected counts in peak/iso-target choices were influenced by the vowel by showing the interaction between [vowel]*[choice of peak/iso-target token], and this was confirmed by the actual data, shown in Appendix IV, where peak tokens were preferred in /ɛ/ and /æ/ but not in /ɒ/ and /ʊ/.

The results show that for the front vowels /ɛ/ and /æ/, the peak token seems to have been preferred as the choice, while in the back vowels /ɒ/ and /ʊ/, two of the five listeners showed a preference for the iso-target token, and the remaining three listeners continued to show a preference for the peak token. To examine whether these preferences are statistically significant, t-tests were made on proportions of 'peak' responses for each /CVC/ token type and for each subject group. The null hypothesis was that the proportion of 'peak' responses was 0.5 for each /CVC/ token type and for each subject group. The significance level was set to 1%.

The results of t-tests[20] indicate that none of the null hypotheses can be rejected, and the experiment does not provide support for the preference for the peak token by subjects. However, it must be emphasised that these observations are weakened by the limited number and quality of data.

[20] For the group consisting of subjects 1, 4 and 5, the t-values, with df = 2, are;

/CVC/	bɛb	bæb	bɒb	bʊb	dɛd	dæd	dɒd	dʊd
t	1.25	0.01	2.65	0.00	5.50	1.73	2.65	2.00
prob.	p>.01	p>.01	p>.01	p>.01	p>.01	p>.01	p>.01	p>.01

For the group consisting of subjects 2 and 3, the t-values, with df=1, are;

/CVC/	bɛb	bæb	bɒb	bʊb	dɛd	dæd	dɒd	dʊd
t	4.00	1.00	1.00	1.00	5.00	1.00	2.00	0.33
prob.	p>.01	p>.01	p>.01	p>.01	p>.01	p>.01	p>.01	p>.01

The following criticisms of this experiment should also be noted:

> (i) The 'target' formant values were based on the results of independent acoustic analyses of /#V#/ rather than on some derivation from /CVC/. Although the input formant values of /CVC/ syllables and iso-target /#V#/ syllables are based on the production of only a single speaker, it is still difficult to justify a hypothesis that listeners could compensate from a vowel in a given /CVC/ to the vowel quality of a /#V#/ whose formant values refer to different realisation and contexts.
>
> (ii) The nature of this experiment forced the listener to choose one of two tokens each of which came from a different hypothesis, and therefore no 'intermediate' values which filled the gap between the 'target' value and trajectory peak value could be selected.

Point (i) above could be addressed by introducing alternative target values, which would be calculated with reference to the acoustic parameters of the CVC stimuli. To address this problem, the Kuwabara function (1985) noted in 3.3.8 could be used to calculate the target values. The Kuwabara function is designed to compensate for, "the ambiguity of these [coarticulated] vowels in the F1 / F2 plane," (Kuwabara 1985:686), by incorporating information conveyed by the surrounding segments. For a given temporal point in a formant trajectory, this function calculates the weighted sum of the formant values by applying to the trajectory a Gaussian window with its centre on that point; then it provides a compensated formant value of that particular point by adding the actual formant frequency and the weighted sum. Kuwabara applied this algorithm to the middle vowels in /CVCVCV/ (the data collected in Kuwabara & Sakai (1972)), and claimed that the function could also compensate for formant frequency shift in CV syllable sequences. Rosner & Pickering (1994) point out some methodological problems in Kuwabara's study: no description is given of the shape of the weighting window nor of any detailed data. They also claim, "[Kuwabara's] algorithm could be transformed directly to an auditory vowel space..." (p.290), thus suggesting that the present study might also benefit from the use of Kuwabara's algorithm to calculate the compensated target value.

The formula of Kuwabara's algorithm is as follows:

$$Fn(t) = fn(t) + \sum_{i=-N}^{N} \omega(iT_0) \times [fn(t) - fn(t-iT_0)] \quad (n = 1,2),$$

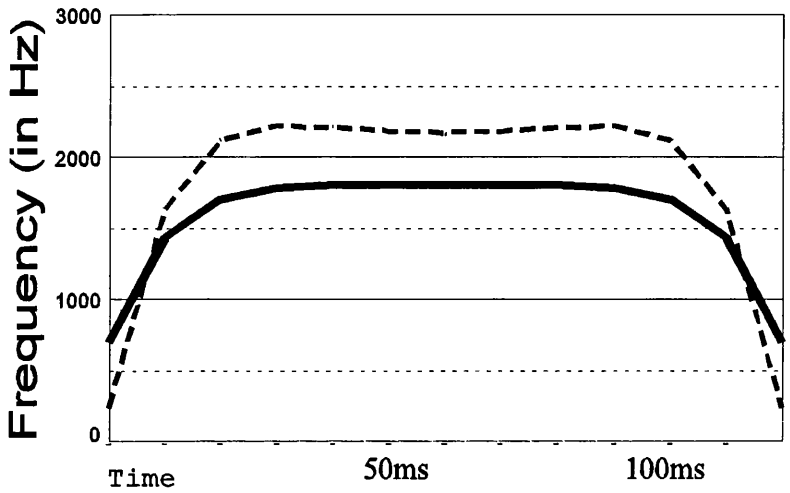
where $fn(t)$ = the original nth formant value at the time of t,
$Fn(t)$ = the compensated nth formant value at the time of t,
N = 1/2 number of the data points (i.e. $2N$ = their total number),
T_0 = sampling time interval

And the weighting function ω in time T is defined as follows;

$$\omega(T) = 7.3 \times \exp[-T^2 / 2(52.0 \times 10^{-3})^2]$$

As an example, this function was applied to the second formant of /bæb/ and /dɒd/ used in the pilot XAXB experiment, and its results are shown in Figure 4-3. Trajectories in a solid line in these figures represent the original F2 trajectories of each token, while those in a broken line stand for the trajectories compensated by the Kuwabara function.

/bæb/



/dɒd/

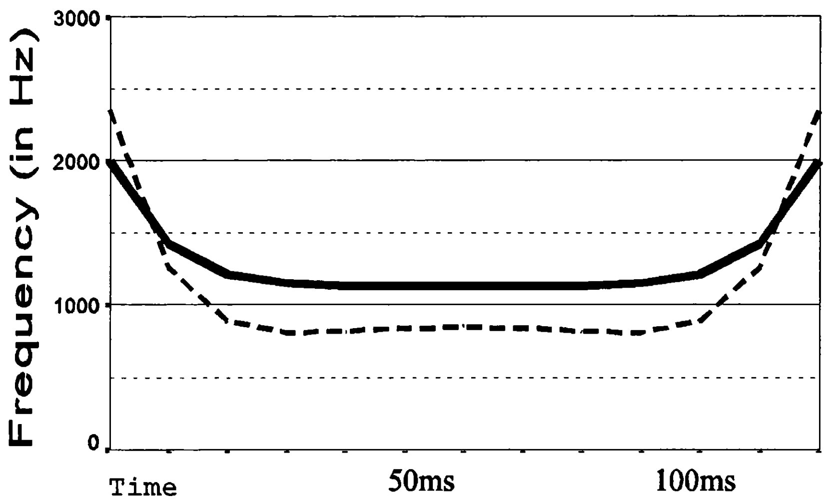


Figure 4-3: Trajectories of /bæb/ (in the spectrogram above, solid line) and /dɒd/ (in the spectrogram below, solid line) and the trajectories compensated by Kuwabara function (in both spectrograms, broken line)

The introduction of the values obtained according to this calculation addresses the criticism (i) mentioned above, since it proposes an alternative 'target' value which could explain the vowel quality perception by an auditory function that refers to the acoustic parameters of the /CVC/ trajectories.

One might assert that point (ii) regarding intermediate choices can be solved in the XAXB experiment by presenting a sufficient number of stimuli to enumerate the potential responses of a listener: perhaps by presenting vowels with some range of F1 and F2 values in equal auditory steps. Theoretically possible as this is, it would create practical problems since the number of stimuli presented in that experiment would be beyond a subject's ability to cope. Suppose that F1 and F2 take six different values. If an XAXB test is created using the same eight X token types as in Pilot Experiment 1, the number of XAXB sets presented can be calculated as follows: first, there will be eight X token types (two consonants and four vowels). Then the number of possibilities for A or B in XAXB is 36 (six F1 variations x six F2 variations), and the number of different A and B is ${}_{36}P_2 = 36 \times 35 = 1260$. Therefore, if each combination is repeated five times, as in Pilot Experiment 1, the total number of XAXB will be 1260 (AB types) x 5 (repetitions) x 8 (X types) = 50,400. With regard to test duration and, particularly, attention by listeners, this experiment would be impossible to perform.

Another alternative to investigating vowel quality through similarity judgements is to introduce an interactive matching experiment scheme, as was discussed in 3.4. If an interactive matching scheme could be introduced, it is expected that 1) subjects would make more finely-tuned judgement in terms of similarity than in a labelling test; and 2) subjects could proceed at their own pace, which would diminish their fatigue and increase the reliability of the experiment.

An interactive matching scheme equipped with a grid-display scheme (a revised version of Nord (1986)) could provide a suitable means to meet the requirements of this study. The procedure of Nord (1986) can be altered as follows:

1)a 6 x 6 grid, with a cursor in one of its blocks, is displayed on a terminal screen as shown in Figure 4-4.

2)the two dimensions of the grid represent the two parameters involved, i.e. F1/F2, and a move from one block in the grid to one of the neighbouring blocks increases/decreases one of the formants of /#V#/ by one pre-determined step (e.g. 100 Hz). See Figure 4-4.

3)/CVC/ (X in XA) stays the same while /#V#/ (A in XA) changes its formant frequencies according to the position in a grid, as Figure 4-4 indicates.

4) a pair of /#V#/ and /CVC/ is replayed upon the request of the subject to the terminal keyboard (in Nord (1986) the task of the subjects was to listen to /CVC/ and find a block where that /CVC/ sounds most "natural").

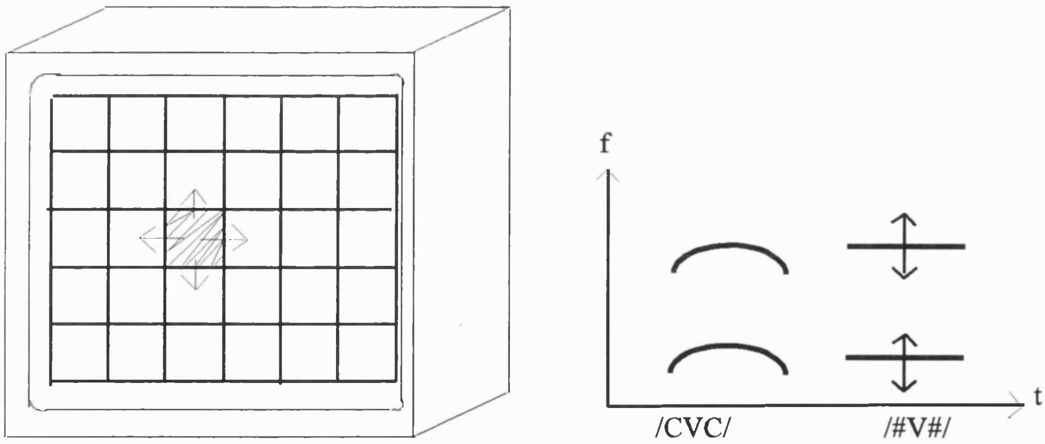


Figure 4-4: Operation of the grid. When a subject moves the cursor on the grid by one step, as shown on the left, the test /#V#/ token changes either its F1 or F2 by one step, while the reference /CVC/ token remains unchanged.

This experimental scheme is superior to XAXB because; (1) the test tokens /#V#/ change in only a single formant frequency step from one box to its neighbour, and therefore subjects can easily avoid the "implausible" matches of /CVC/ and /#V#/ (i.e. matches whose vowels of /CVC/ and /#V#/ provide an obviously different auditory impression), and this saves time and effort for subjects; (2) the subjects, once they have found some "plausible" sets of /CVC/ and /#V#/, can have them played back as many times as they like, easily and quickly, by moving the cursor around a particular region, in order to find a best match between /CVC/ and /#V#/, and this produces more reliable responses; (3) the scheme allows a listener to change F1 and F2 simultaneously, which makes possible the investigation of the interactions between the two parameters; and (4) by giving visual feedback, it reduces the monotony of the listening tests.

It must be added that this type of interactive matching design is not common in the literature and a scheme similar to this has not been found in the literature. The scheme by Nord (1986) only changed the formant values of one token to make it sound more 'natural' as was discussed in 3.4.2.

However there is a potential criticism of this interactive grid-matching scheme: whether subjects are really able to cope with this task and can really tune into the vowel quality and make judgements finer than phonological categories.

In order to test this validity, Pilot Experiment 2 was conducted which implemented this scheme but whose matching task was, not between /CVC/ and /#V#/, but between /#V#/ and /#V#/ with changes only in formant values. This matching of /#V#/ and /#V#/ was employed since this experiment concentrated on the effect of the matching scheme on vowel quality evaluation in the most fundamental condition, and if, under this adequately controlled condition, the subjects failed to match the values of /#V#/ and /#V#/, then the experimental set-up would be clearly inappropriate. Pilot

Experiment 3, which dealt with the matching between /CVC/ and /#V#/, was also performed and is discussed below.

4.3 Pilot Experiment 2 (matching between /#V#/ and /#V#/)

This section describes the detail of Pilot Experiment 2 which investigated the credibility of the grid-matching scheme with elementary material.

4.3.1 Materials

Two types of stimuli, playing a different role, were used in this experiment.

a) test tokens: short vowels in isolation, whose formant frequencies could be modified by listeners in an interactive mode by moving a cursor in a grid to another position, which will be discussed below (A or B token in Pilot Experiment 1)

b) reference tokens: RP short monophthongs /ɪ,ɛ,æ,ɒ,ʌ,ʊ/ in isolation, whose formant frequencies were not able to be changed by listeners and were identical in every position in a grid

The types of the vowels used in this experiment were increased from four to six, enumerating all RP short monophthongs to investigate whether /ɪ/ and /ʌ/ show any particular behaviour in matching. These tokens were synthesised using the JSRU synthesiser as in Pilot Experiment 1, and they had only two formants, F1/F2. The duration of the test token and the reference token was identical: 220 ms, both with linear F0 declination from 130 Hz to 100 Hz within it. Each type of six reference-test pairs had four repetitions per session, creating 6 x 4 = 24 matching trials. They were randomised in the order of presentation, preceded by one trial session with feedback to the listener before the main sessions.

The formant values of the reference /#V#/ were obtained from the results of the acoustic analysis, which were also used in Pilot Experiment 1. The values are reproduced in the Table 4-3 below.

	ɪ	ɛ	æ	ɒ	ʌ	ʊ
F1	396	677	847	629	805	421
F2	2157	1945	1665	976	1286	1066

Table 4-3: F1/F2 values for reference /#V#/ tokens. All values in Hz.

As both types of tokens simulated /#V#/, they had two steady formants. In the synthesis, every 10 ms point was interpolated linearly. Voicing reached the full amplitude 10 ms after the onset and started to decline 10 ms before the end. F1/F2 intensities, between initial and final changing 10 ms, were fixed at 50 dB.

4.3.2 Subjects

Three native speakers of South East British English participated in this experiment. They were either postgraduate students of the Department of Phonetics, University College London or departmental staff. They had no reported hearing defects.

4.3.3 Procedure

The procedure of this experiment was as follows: in a sound-proof room, individual subjects were asked to sit in front of a PC terminal. Its screen displayed a schematic grid (6 x 6 blocks) with a cursor on one of these blocks, showing the "relative" position of the test token. In fact the grid was an acoustic vowel diagram, with F1=25 Hz and F2=50 Hz step, but the subjects were not informed of this. As they moved the cursor into a neighbouring block, so there was an increase/decrease of either F1 or F2 frequency of the test token by one step, while the reference token remained unchanged. The advantage of using this grid system, as was discussed above, is that it can tell subjects intuitively how much they have changed the vowel quality from a previous position. Each time the cursor was moved, or when the space key was pressed, a pair of a test token and a reference token having an interval of 120 ms between them was replayed through a speaker. These two tokens were played in the order of 'reference'-'test'. The procedure is shown in Figure 4-5.

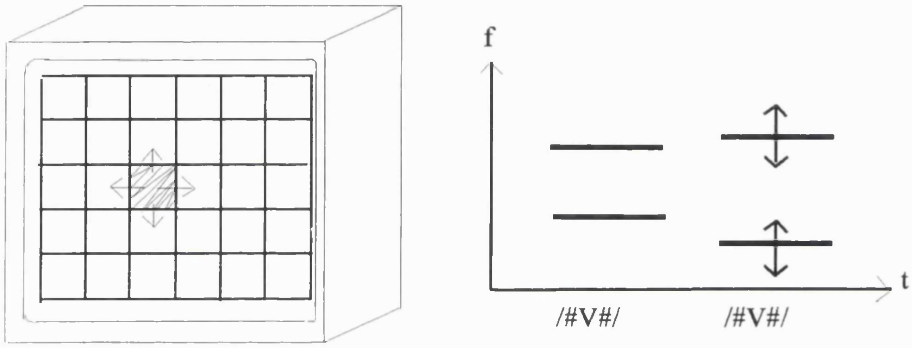


Figure 4-5: Illustration of the task in Pilot Experiment 2. As the cursor on the PC screen moves to its neighbours, so there was an increase/decrease of either F1 or F2 frequency of the test token (right) by one step, while the reference token (left) remained unchanged.

One further point on the grid arrangement must be made. If there is one particular block in the grid whose F1/F2 values are the same as the F1/F2 values of the reference /#V#/, the task of the subject would be simply to find the identical reference /#V#/ and test /#V#/, a result which would not prove the grid mechanism for matching. Therefore, the grid was arranged to make the grid values of F1/F2 all different from the reference F1/F2.

Subjects were instructed to tune into the "vowel quality" of both tokens and to discover in the grid a block where these tokens sounded closer than in any other block, i.e. to find a "closest" pair of /#V#/ and /#V#/ in vowel quality. When they found one, they pushed the "q" key on the keyboard and F1/F2 values of the 'test' token and the other parameters were stored in an ASCII data file, and the next stimulus pair and the next grid were presented.

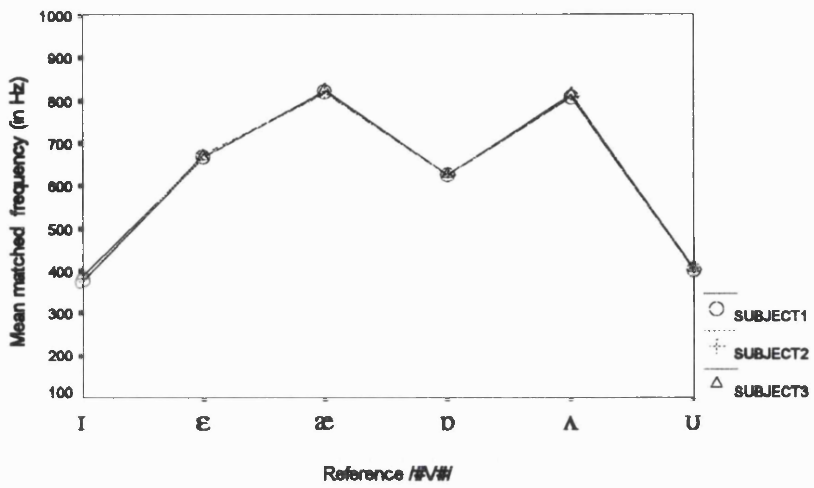
To avoid any bias effect by subjects, the allocation of F1/F2 on the two axes was changed from one stimulus pair to another: the vertical axis was F1 and the horizontal axis was F2 but the direction of F1 increment could be either "down" or "up" in the vertical axis and that of F2 increment can be either "right" or "left" in the horizontal axis. Therefore, there were four possibilities of allocation: (F1 increment, F2 increment) = (up,right), (up,left), (down,right), (down,left). All these axis allocation patterns were used in this experiment. One could have had F2 in a vertical axis and F1 in a horizontal axis, so that all axis allocation patterns could have been enumerated, but this would have doubled the number of trials.

The whole experimental process carried out on a terminal screen was programmed in C-language by Dr Mark Huckvale, University College London.

4.3.4 Results and Discussion

First, mean matched frequencies were calculated across all four trials for each subject, vowel type and formant. They are plotted in Figure 4-6 to study the general behaviour of each subject.

F1 mean matched frequency of each subject



F2 mean matched frequency of each subject

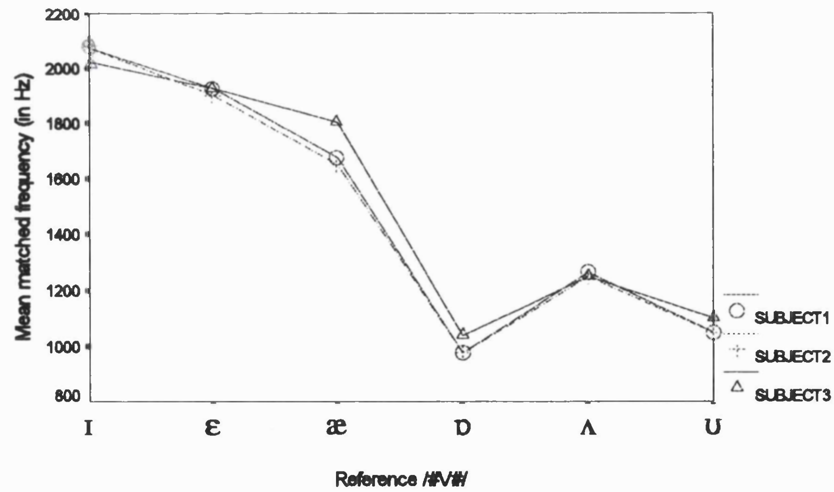


Figure 4-6: Mean matched F1 and F2 frequencies of each subject for Pilot Experiment 2.

Figure 4-6 suggests that there is little difference between the results of the three subjects. However, to ensure the homogeneity of the three subjects, Repeated Measures ANOVA was carried out for each formant number (i.e. F1 / F2), with factors of [vowel], [subject] and [trial]. This repeated measure procedure was introduced because the same variable was measured on several occasions for each subject of this experiment, which is statistically a repeated measurement design. To show how far a test /#V#/ formant value was shifted from its reference /#V#/ formant value, a shift index was obtained for each trial/each formant number, by [one matched frequency of one trial] - [its reference formant value]. This shift index was utilised instead of the actual matched formant frequency. SPSS for Windows (Version 6.1.3) was used for this statistical analysis[21] and the significance level was set to 1 %.

For F1, the result of Repeated Measures ANOVA indicated that [subject] as a main factor was not significant: $F(2,30)=2.11$, $p>.01$; and the interaction of [subject] by [vowel] was not significant: $F(10,30)=0.64$, $p>.01$. The result of F2 showed the same tendency: [subject] as a main factor was not significant: $F(2,30)=1.47$, $p>.01$; and the interaction of [subject] by [vowel] was not significant: $F(10,30)=1.75$, $p>.01$.

Then the mean of the shift index across three homogenous subjects was calculated and displayed in the Table 4-4 below.

---

[21] However, this experimental design, where all the degrees of freedom were accounted for by the factors included in this analysis, [subject], [vowel] and [trial], does not produce any residual error term which SPSS for Windows (Version 6.1.3) would normally use to make an F test on the subject factor. It would be possible to obtain a residual error term by defining each trial as a different case in SPSS, but in SPSS data structure, a Repeated Measures ANOVA model should be carried out with the multivariate set-up used in SPSS. This means that all of the measurements for each subject should be present on a single case as different variables. The factor [trial], representing the set of equivalent trials, could be excluded from the Repeated Measures ANOVA model by using a custom model option, but this does not introduce a non-zero degree of freedom for the residual error term. Therefore, we made an F test manually, using the MS for a factor or an interaction investigated and the MS error from the highest order interaction, [subject]*[vowel]*[trial] in this case.

	ɪ	ɛ	æ	ɒ	ʌ	ʊ
Mean F1 shift index	-16.8	-6.2	-20.2	-4	5.4	-18.9
Mean F2 Shift index	-18.9	-32.5	-14.2	11.5	-23.5	-3.5

Table 4-4: Mean shift index across three homogenous subjects. All values in Hz

Table 4-4 indicates that the F1/F2 values which subjects selected for the test tokens (whose formants are interactively changeable) mostly matched the F1/F2 values of the reference tokens. Besides, the raw data showed that there was only one trial (out of 90 trials) where the matched F1 and F2 values of a test token diverged from its corresponding reference token by two formant steps. There were no regular patterns of variations observed in terms of the subject, the vowel type or the formant values.

The null hypothesis that the median of the shift index was 0.0 was tested by Wilcoxon test, on the shift index (i.e. [one matched frequency of one trial] - [its reference formant value]) for each formant type and a reference vowel type. The significance level was set to .01, as in Pilot Experiment 1. Since this procedure is a multiple-paired comparison, the individual significance level was determined by the Bonferroni procedure in order to decrease the probability of Type I error. The procedure stipulates that given N tests of a family, the Type I error rate per comparison should be obtained by dividing, by the number of the tests N, the probability that at least one Type I error occurs in N tests. (Hays (1988:410)). Following this, the total significance level of .01 in this experiment should be divided by the number of comparisons, 12, in this case, (6 vowels x 2 formants) therefore the significance level of each Wilcoxon test being .01/12 = .0008. This Wilcoxon test was carried out by Minitab for MS-DOS (Version 6.1)[22]. The z values produced by Wilcoxon tests are

[22] SPSS for Windows (Version 6.1.3) is implemented with pair-wise Wilcoxon tests but not with the Wilcoxon test with a given hypothetical median, which is the case in this experiment.

displayed in Table 4-5 below.

	ɪ	ɛ	æ	ɒ	ʌ	ʊ
F1 Wilcoxon z	-2.82	-3.05	-3.05	-3.05	-0.86	-2.98
Probability	p>.0008	p>.0008	p>.0008	p>.0008	p>.0008	p>.0008
F2 Wilcoxon z	-2.52	-1.88	-0.25	-0.47	-2.58	-0.47
Probability	p>.0008	p>.0008	p>.0008	p>.0008	p>.0008	p>.0008

Table 4-5: Results of Wilcoxon tests. Their obtained z values are shown for each formant type and reference vowel type.

None of the Wilcoxon z in Table 4-5 is significant at the level of .0008. This provides no evidence that subjects chose anything but F1/F2 values which matched the reference F1/F2.

The subjects completed each full test within 30 minutes. In comparison, an equivalent XAXB would have had 30,240 stimulus presentations. This provides a persuasive support for adopting this grid matching scheme.

Thus it seems reasonable to apply the grid-matching scheme to /CVC/ and /#V#/ in the main experiment. One aspect learnt from this study was that some subjects could not distinguish vowel quality change between adjacent blocks. Therefore, the formant step was modified in the next /CVC/-/#V#/ experiment through the use of Bark scaling of frequency.

The next section 4.4 presents Pilot Experiment 3, whose object is to examine the validity of this grid scheme on /CVC/-/#V#/ matching and, in addition, to observe the phonetic / psychoacoustic evaluation of vowel quality with dynamic formant trajectories, which is the main object of this thesis.

4.4 Pilot Experiment 3 (matching between /CVC/ and /#V#/)

4.4.1 Materials

The test token in this Pilot Experiment 3 had identical acoustic properties as in Pilot Experiment 2: /#V#/ whose formants subjects could modify in the grid; while the reference token in this experiment was altered: RP short monophthongs in /bVb/ or /dVd/ context whose formant frequencies listeners were not able to change. They were synthesised by JSRU synthesiser following the identical procedure as in Pilot Experiments 1 and 2. The vowel was /ɪ,ɛ,æ,ɒ,ʌ,ʊ/, as in Pilot Experiment 2. The formant trajectory of the reference token /CVC/ was again calculated by Nearey's formula, as in Pilot Experiment 1, using the same input values for /ɛ,æ,ɒ,ʊ/. For /ɪ,ʌ/ the values obtained from the acoustic analysis mentioned in the preparatory experiment were used like four other vowels: the peak (F1,F2) values for /ɪ/ were (415,1985) in /bVb/ context and (403,2049) in /dVd/, while the (F1,F2) values for /ʌ/ were (656,1354) in /bVb/ context and (658,1341) in /dVd/. The input values are shown in Table 4-6.

	ɪ	ɛ	æ	ɒ	ʌ	ʊ
b_b	415/ 1985	542/ 1806	760/ 1621	545/ 1009	656/ 1354	491/ 1122
d_d	403/ 2049	526/ 1848	733/ 1619	565/ 1135	658/ 1341	437/ 1359

Table 4-6: Input peak values obtained from the pilot acoustic study, where these words are embedded in a frame sentence. The values indicated are F1/F2. All values in Hz.

The duration of each test token was 220 ms, while the reference tokens had two durations: 120 ms and 220 ms. These two durations were chosen since 120 ms was obtained from the pilot acoustic analyses on /CVC/ tokens in an unstressed position of the frame sentence, while 220 ms was the value used in the test token in Pilot

Experiment 2. Two different values were chosen to see whether a difference in duration had any effect on dynamic formant trajectory evaluation. The other procedures and parameter settings of the synthesis were identical with Pilot Experiment 2. Each pair of tokens ("reference"-"test") was run four times on the grid, and therefore the number of stimuli per subject is; 6 (vowels) x 2 (types of consonantal contexts) x 2 (duration types) x 4 (repetitions) = 96, preceded by two test matching sessions with a feedback to the subject before the main trial session. These stimuli were grouped together into stimulus blocks according to their consonantal environment and durational pattern. This process produced four blocks: 1) reference token = 120 ms /bVb/, 2) reference token = 220 ms /bVb/, 3) reference token = 120 ms /dVd/, 4) reference token = 220 ms /dVd/. Inside these blocks, the stimuluspairs were in a randomised order.

4.4.2 Subjects

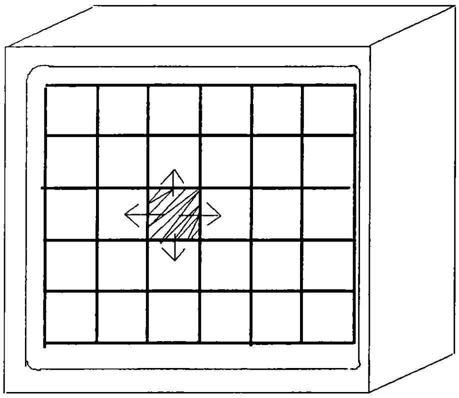
Four native speakers of South East British English, either a PhD student or one of the staff of Department of Phonetics, University College London, took part in this experiment. None of them had evidence of a hearing problem.

4.4.3 Procedure

The procedure was similar to Pilot Experiment 2: in a sound-proof room, each subject, in front of a PC terminal, was asked to match the vowel quality of a reference token (in this experiment /CVC/) and its paired test token /#V#/, by moving around a cursor over a schematic grid on the screen, listening to the /CVC/-/#V#/ (reference-test) stimuli, the latter of which changed its F1/F2 according to the position of the cursor on the grid. The same instructions as in Pilot Experiment 2 were given to subjects orally.

There were two minor aspects in this experiment which were modified from Pilot Experiment 2. One was that discussed in 4.3.4: the F1 and F2 step change on the grid was 50 Hz (instead of 25 Hz) for F1 and 0.5 Bark (instead of 50 Hz) in F2. The Bark scale was introduced to approximate the actual frequency selectivity process and it has a narrower resolution in the lower frequencies than the linear scale. In this research, Bark values were obtained by the formula in Bladon & Lindblom (1981)[23]. The other modification was the duration of the interval between the reference and the test token: it was 120 ms in Pilot Experiment 2 but 300 ms in the current experiment, since it was revealed that the 120 ms interval, being short, caused the reference-test pair, /CVC/-/#V#/ to sound like one word /CVCV/. To prevent this, the interval between the reference-test was lengthened and it was found by a brief survey of several listeners that 300 ms was sufficient to make them sound separate from each other. The set-up is shown in Figure 4-7. An example of the actual arrangements of the frequencies on the grid is shown in Figure 4-8. The 120 ms and 220 ms reference /CVC/ tokens shared the identical grid pattern. Note that as in Pilot Experiment 2, no block in a grid coincided with the precise F1/F2 values of the /CVC/ trajectory peaks.

[23] The formula devised by Bladon & Lindblom is: up to 3000 Hz, the critical band number z (Bark) for a given frequency is provided by: $z = 7 * \ln \{(f/650)+[(f/650)^2+1]^{1/2}\}$ (Bark)



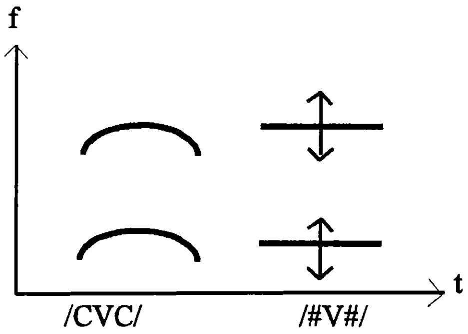


Figure 4-7: Illustration of the task in Pilot Experiment 3. As the cursor on the PC screen moves to its neighbours, so there was an increase/decrease of either F1 or F2 frequency of the test /#V#/ token (right in the spectrogram) by one step, while the reference /CVC/ token (left in the spectrogram) remained unchanged.

250 800	250 875	250 955	250 1041	250 1131	250 1227
300 800	300 875	300 955	300 1041	300 1131	300 1227
350 800	350 875	350 955	350 1041	350 1131	350 1227
400 800	400 875	400 955	400 1041	400 1131	400 1227
450 800	450 875	450 955	450 1041	450 1131	450 1227
500 800	500 875	500 955	500 1041	500 1131	500 1227

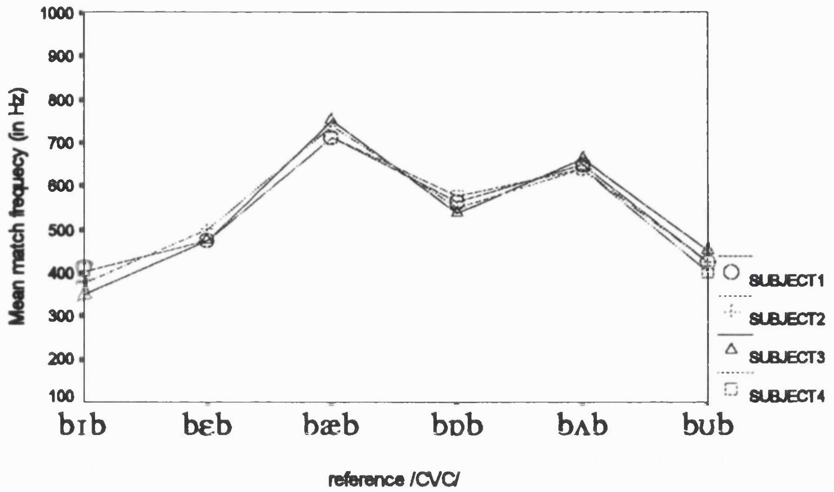
Figure 4-8: Example of grid arrangement (for 120ms / 220ms /bʊb/). In each block, the number above represents the F1 value of the test /#V#/ token and the number below its F2 value. All values in Hz.

The subjects had two stimulus blocks of 48 runs on one day and the two further blocks several days later, since having all stimulus blocks on one day would have been exhausting. The total duration of the testing differed from one subject to another, the longest being 1.5 hours for one day session.

4.4.4 Results and Discussion

The results obtained were analysed as follows: as in Pilot Experiment 2, mean matched frequencies were calculated across all four trials for each subject, /CVC/ reference token type and formant. They are plotted in Figures 4-9 to 4-12 to study the general behaviour of each subject. Figures 4-9 to 4-12 do not show whether the matching pattern of four subjects is identical or not. Hence Repeated Measures ANOVA was made on the shift index (i.e. [matched formant value of a test token] - [peak formant value of its corresponding target CVC token]), with the factors of [subject] (4 levels), [consonant] (2 levels), [vowel] (6 levels), [duration] (2 levels) and [trial] (4 levels). The highest order interaction, employed as a denominator of F-ratio, was therefore [subject]*[consonant]*[vowel]*[duration]*[trial]. The significance level of the ANOVA was set to 1% and its F-ratios were calculated manually, as in Pilot Experiment 2.

120ms /bVb/F1



220ms /bVb/ F1

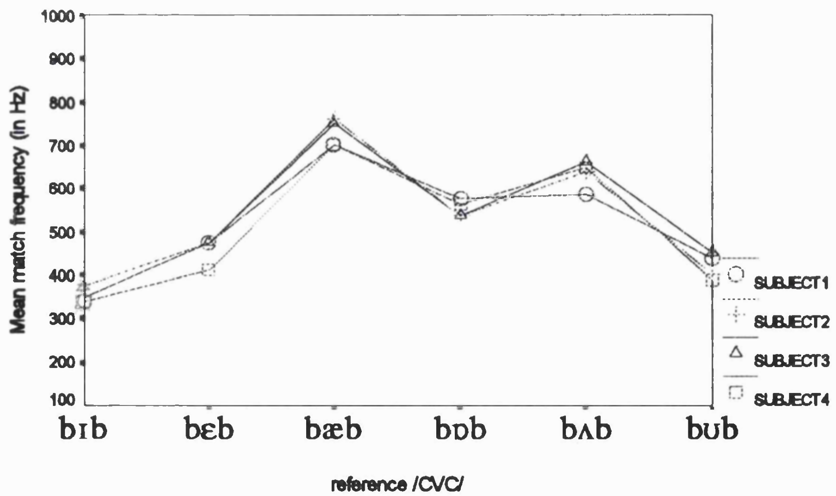
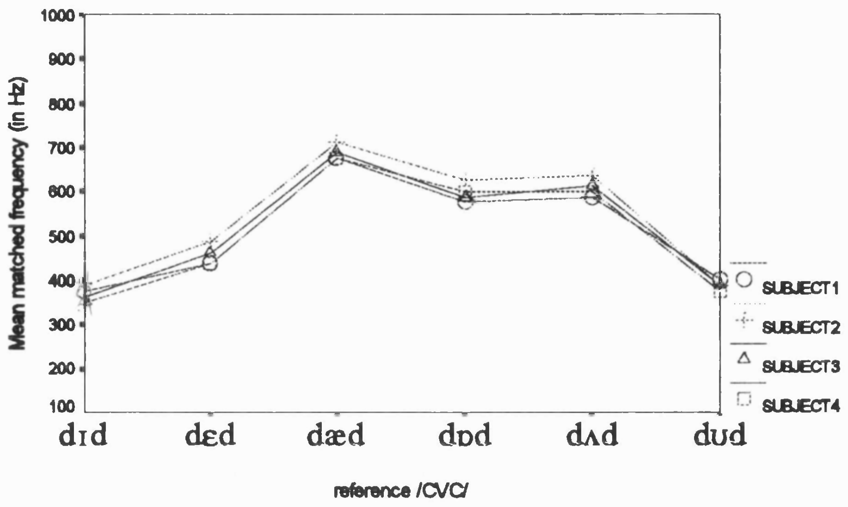


Figure 4-9: Mean matched frequency of each subject: /bVb/ F1

120ms /dVd/F1



220ms /dVd/ F1

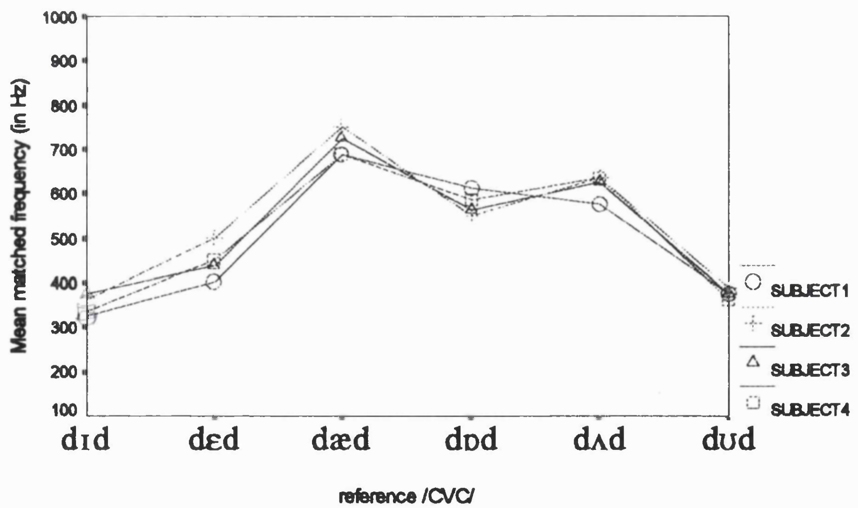
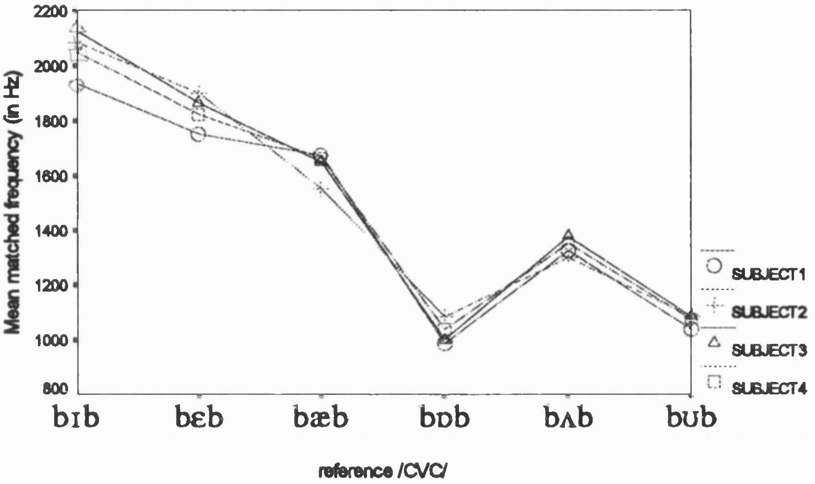


Figure 4-10: Mean matched frequency of each subject: /dVd/ F1

120ms /bVb/F2



220ms /bVb/ F2

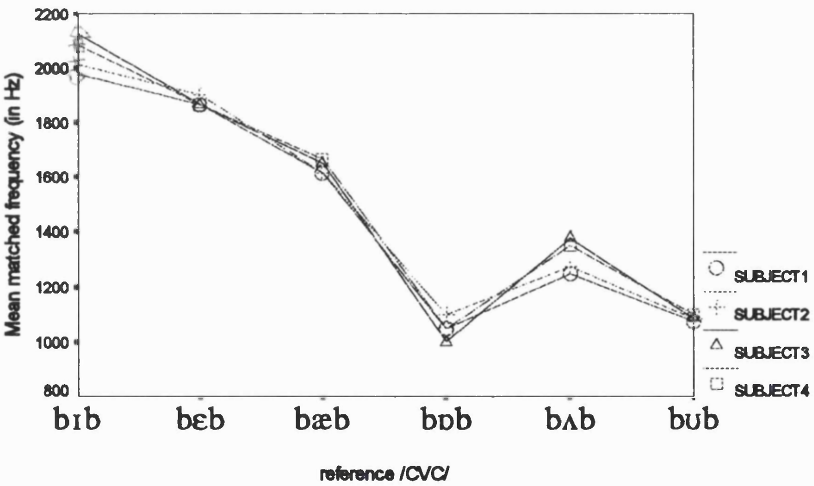
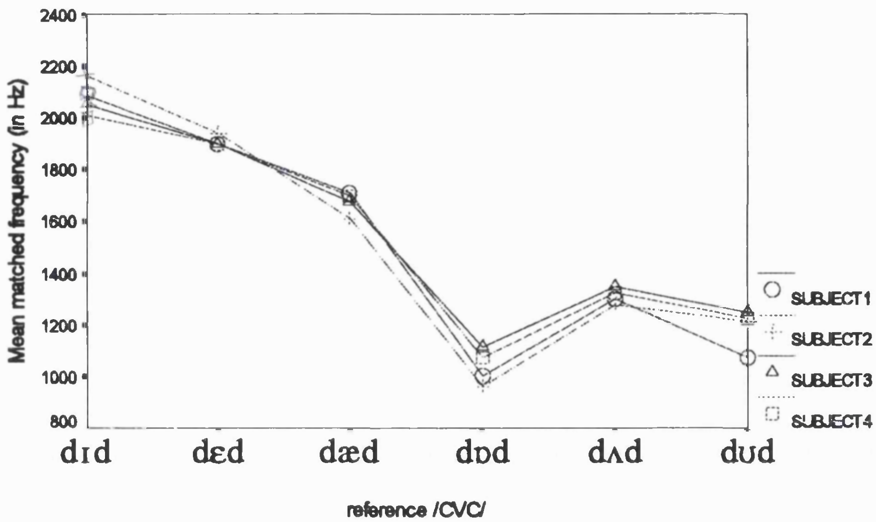


Figure 4-11: Mean matched frequency of each subject: /bVb/ F2

120ms /dVd/F2



220ms /dVd/ F2

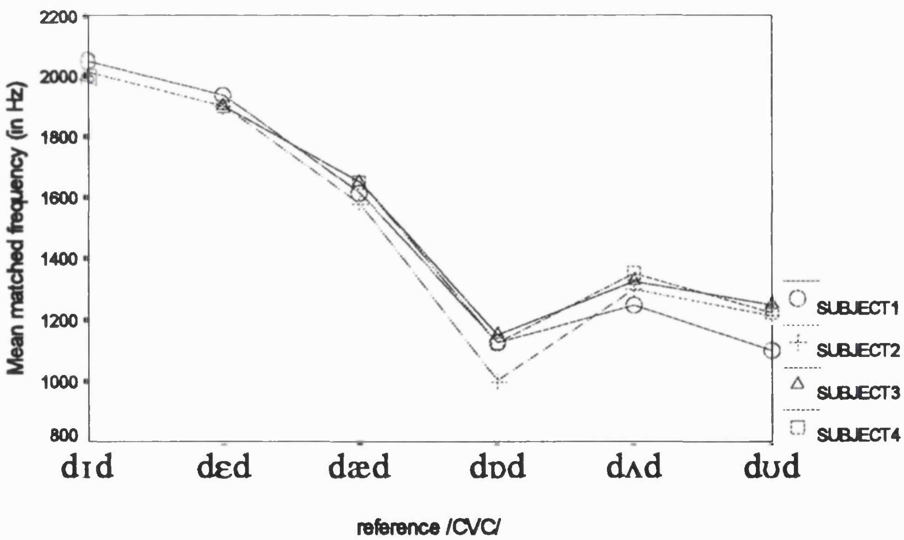


Figure 4-12: Mean matched frequency of each subject: /dVd/ F2

For F1, the results of the Repeated Measures ANOVA indicated that [subject] as a main factor was not significant: $F(3,45) = 0.89$, $p > .01$; and none of the interactions of [subject] by [vowel], [subject] by [consonant], and [subject] by [duration] was significant: respectively, $F(15,45) = 3.17$, $p > .01$; $F(3,45) = 2.41$, $p > .01$; and $F(3,45) = 0.94$, $p > .01$. The results for F2 indicated an identical tendency: [subject] as a main factor was not significant: $F(3,45) = 2.40$, $p > .01$; and none of the interactions of [subject] by [vowel], [subject] by [consonant], and [subject] by [duration] was significant: respectively, $F(15,45) = 1.46$, $p > .01$; $F(3,45) = 1.06$, $p > .01$; and $F(3,45) = 0.85$, $p > .01$. With regard to the influence of each factor, in F1 the factor [vowel] was significant ($F(5,45)=39.96$, $p<.01$) but not the factors [consonant] ($F(1,45)= 0.94$, $p>.01$) and [duration] ($F(1,45)=1.27$, $p>.01$). In F2, the factors [vowel] and [consonant] were significant ($F(5,45)=26.89$, $p<.01$; $F(1,45)= 32.51$, $p<.01$), but not the factor [duration] ($F(1,45)=2.31$, $p>.01$).

Since it was shown that the four subjects in this experiment can be treated as a homogenous group, the shift index of F1 and F2 was pooled according to a reference /CVC/ type, and the mean was calculated, and listed in Table 4-7 below. The mean values are plotted in Figures 4-13 to 4-20. The two durational conditions are plotted separately.

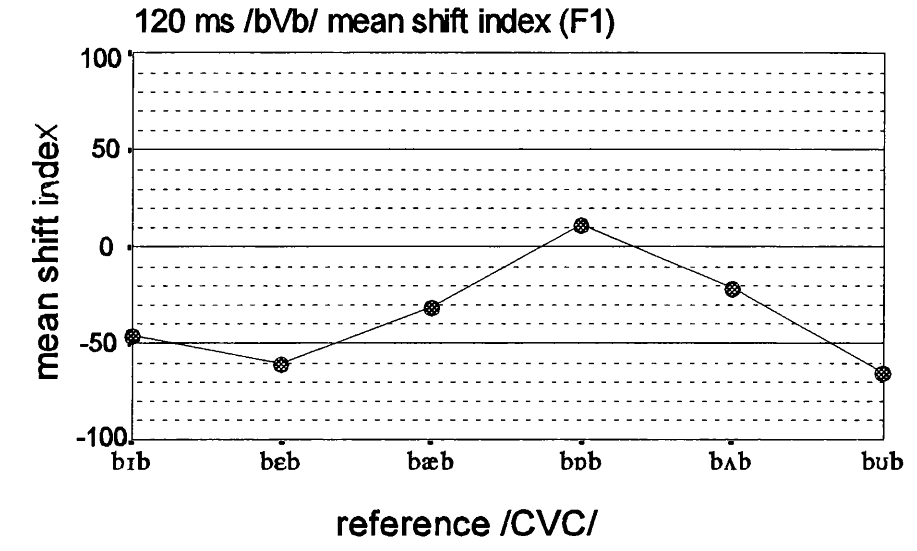


Figure 4-13: F1 mean shift index of 120ms /bVb/ token.

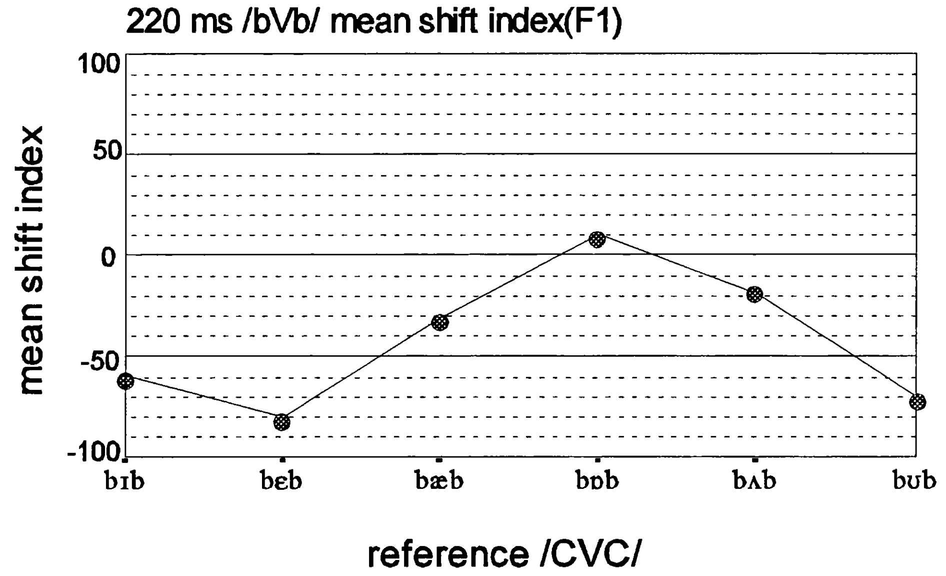


Figure 4-14: F1 mean shift index of 220ms /bVb/ token.

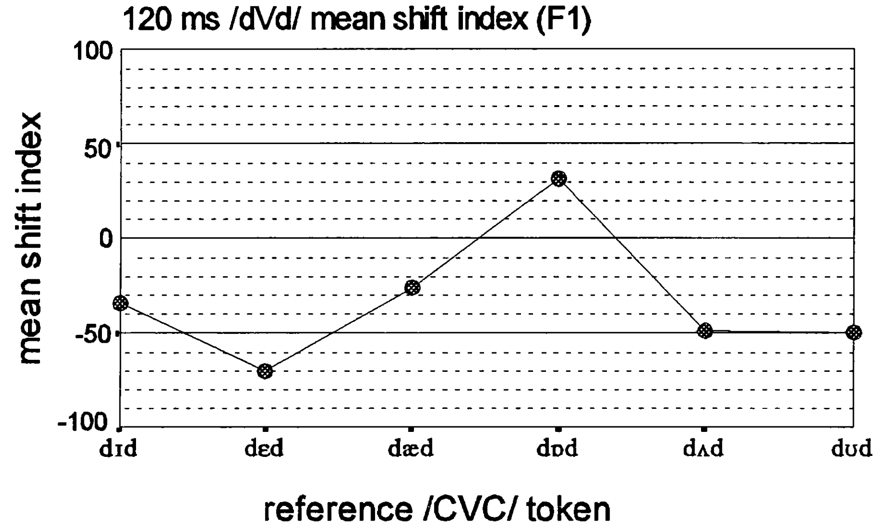


Figure 4-15: F1 mean shift index of 120ms /dVd/ token.

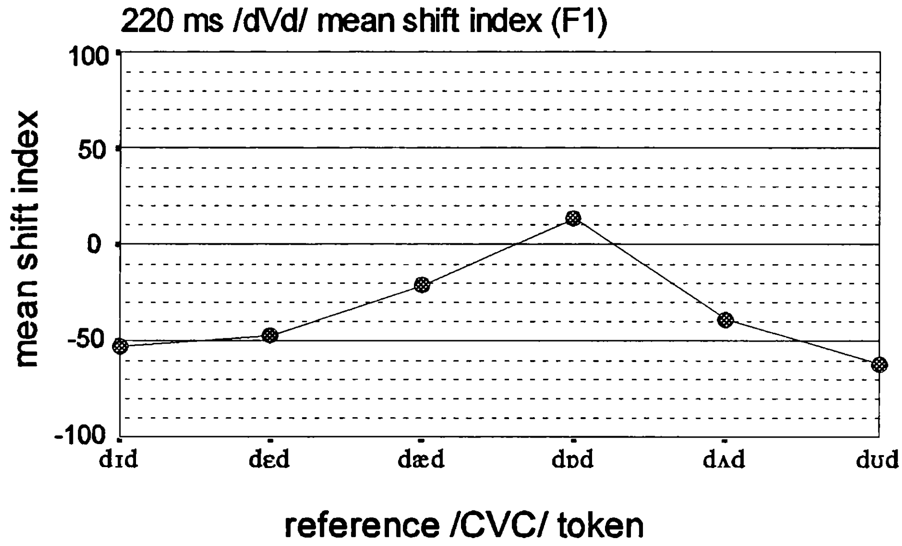


Figure 4-16: F1 mean shift index of 220ms /dVd/ token.

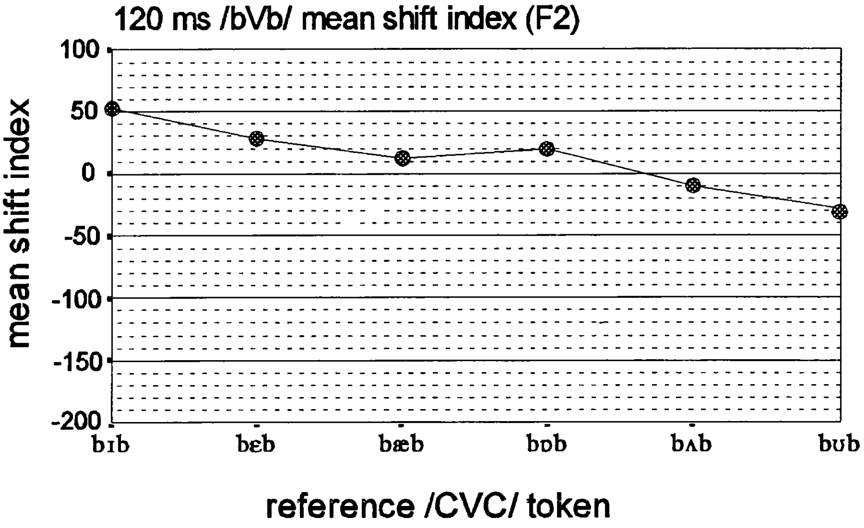


Figure 4-17: F2 mean shift index of 120ms /bVb/ token.

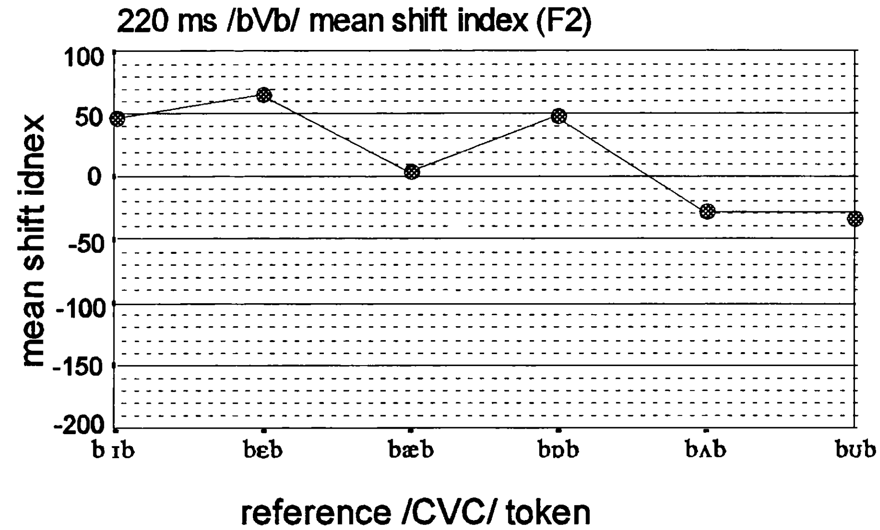


Figure 4-18: F2 mean shift index of 220ms /bVb/ token.

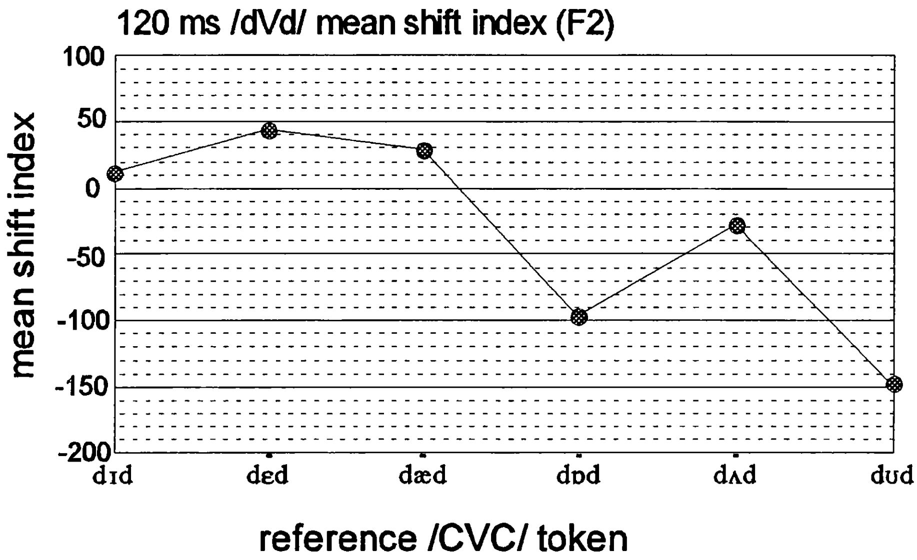


Figure 4-19: F2 mean shift index of 120ms /dVd/ token.

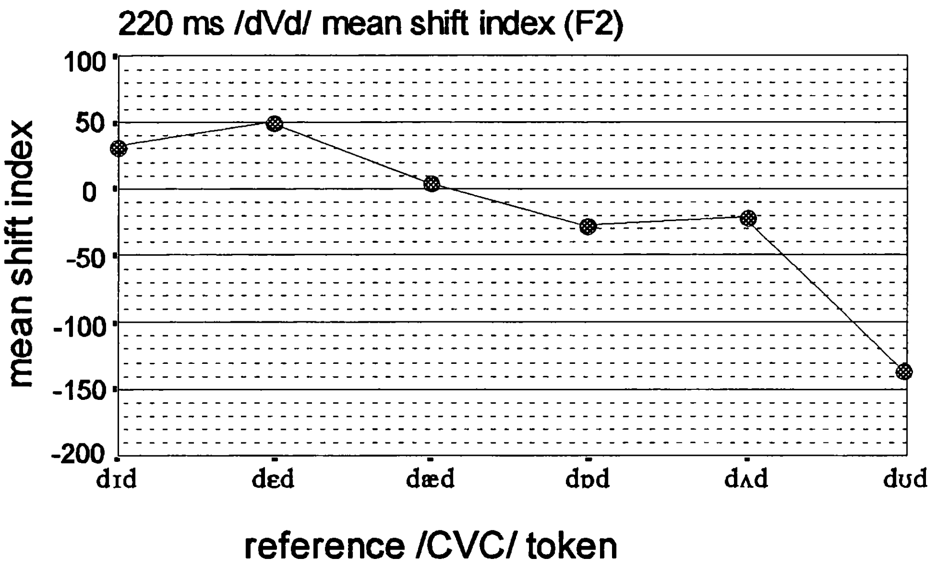


Figure 4-20: F2 mean shift index of 220ms /dVd/ token.

120ms /bVb/

	bɪb	bɛb	bæb	bɒb	bʌb	bʊb
F1	-46.25	-60.75	-31.88	11.25	-21.63	-66.00
F2	52.50	28.38	11.62	19.13	-10.25	-28.25

220ms /bVb/

	bɪb	bɛb	bæb	bɒb	bʌb	bʊb
F1	-61.88	-82.62	-32.88	8.12	-21.63	-72.25
F2	46.25	65.88	4.00	47.25	-29.00	-28.25

120ms /dVd/

	dɪd	dɛd	dæd	dɒd	dʌd	dʊd
F1	-34.25	-69.75	-25.50	31.88	-48.62	-49.50
F2	10.38	42.63	27.88	-97.50	-28.50	-149.63

220ms /dVd/

	dɪd	dɛd	dæd	dɒd	dʌd	dʊd
F1	-53.00	-79.13	-20.50	13.13	-39.25	-62.00
F2	-27.13	48.88	2.87	-35.00	-22.25	-137.12

Table 4-7: Mean shift index across four homogenous subjects, for each consonantal context, vowel, duration type and formant number, in Pilot Experiment 3. All values in Hz.

Figures 4-13 to 4-16 together with Table 4-8 may be studied to investigate F1 matching. Figure 4-13 shows the results of the block with 120 ms /bVb/ as reference tokens. One can observe that the mean matched frequencies of /bɪb/, /bɛb/, /bʊb/ were lower than their trajectory peaks, while in /bɒb/ and /bʌb/ the mean shift index was closer to the peak. The same tendency can be seen in Figure 4-14, for 220ms /bVb/ as a reference token: the shift index for /bɪb/, /bɛb/ and /bʊb/ was much lower than zero, corresponding to the F1 trajectory peak. On the other hand, the mean shift index of /bæb/, /bɒb/ and /bʌb/ was close to zero.

Figure 4-15, with 120 ms /dVd/ as a reference token, shows the mean shift index considerably lower than zero, the trajectory peak, in /dɪd/, /dɛd/, /dʌd/ and /dʊd/ context. In /dæd/ and /dɒd/, the mean shift index was close to zero, the F1 trajectory peak. It was the same in Figure 4-16, for 220ms /dVd/ as a reference.

As in Pilot Experiment 2, Wilcoxon tests were employed to investigate whether these shifts are significant. The null hypothesis is that the median of the shift index was 0.0 for each formant type and a reference token type. Following the Bonferroni procedure, the significance level of each Wilcoxon test was set to .01/48= .0002 (2 consonants x 6 vowels x 2 durations x 2 formants) This Wilcoxon test was carried out by Minitab MS-DOS Version, as in Pilot Experiment 2.

The z values produced by Wilcoxon tests are displayed in Appendix V. No cases are statistically significant at the level of .0002, the 1 % significancelevel modified by the Bonferroni procedure, showing that the results fail to refute the null hypothesis in all cases. This indicates that none of the divergences of the matched formant value from the /CVC/ trajectory peak was statistically significant, although the existence of a consistent difference pattern observed in F1 suggests an effect of the range of the F1 movement, i.e. the difference of the peak and the onset/offset in the trajectory, (the trajectory range henceforth): /ɪ/, /ɛ/ and /ʊ/, whose trajectory range was comparably small, showed lower match frequencies than the rest, and it seems to hold in both contexts. The apparent failure of the shift index to be statistically significant could also be ascribed to the robust nature of the Wilcoxon test and the small number of the cases. It must be noted, however, that this interpretation cannot be evaluated further owing to the data deficiency. Further verification with a increased number of cases is required.

Figures 4-17 to 4-20 and Table 4-7 show F2 matching results. Figure 4-17 and Table 4-7 display the results of the block with 120 ms /bVb/ as reference token. There, mean shift indices slightly higher than zero with a large frequency range are observed in

/bɪb/ and /bɛb/, while means lower than the peak of the trajectory with a small frequency range are observed in /bʊb/. A similar pattern is observed in Figure 4-18, with 220 ms /bVb/ as a reference token. These results also suggest the possible involvement of the trajectory range in the matching process, although the result of Wilcoxon tests in Appendix V shows that none of the divergences of the matched formant value from the reference /CVC/ peak was significant. This hypothesis needs further empirical evidence to test its validity.

Figure 4-19, representing the results of 120 ms /dVd/ as a reference token, reveals that in /dɒd/ and /dʊd/ the mean shift index was lower than zero, the /CVC/ trajectory peak, and in /dɛd/ the matched values were higher than the peak, while in the other tokens the mean shift index was closer to the /CVC/ peak. In Figure 4-20 for 220 ms /CVC/ tokens, a similar pattern is observed, except for /dɒd/, where the mean shift index is close to zero, the /dɒd/ peak.

Despite Wilcoxon tests which show that none of the divergences of the matched formant value from the reference /CVC/ peak was significant, the results of Pilot Experiment 2, empirically insufficient as they may be, could suggest a hypothesis of low F2 matching in /dɒd/ and /dʊd/: low F1 values of /CVC/ could affect the F2 matching process by lowering it from the F2 /CVC/ trajectory peak. This could be explained by three factors: 1) the trajectory shape of F2 is concave; 2) the two formants are relatively close[24]; and 3) F1 is in a low frequency region. It is only /dɒd/

[24] For reference, the Bark distance of F1peak - F2 peak is;

	ɪ	ɛ	æ	ɒ	ʌ	ʊ
bVb	8.84	6.90	4.54	3.22	4.14	4.31
dVd	8.95	7.19	4.72	3.76	4.07	5.97

Notice that the distance of two formant peaks is small in /ɒ,ʌ,ʊ/

and /dʊd/ that satisfy all three factors. The third factor is required to exclude /dʌd/, which satisfies the factors 1) and 2) but does not show any F2 lowering. However it should be emphasised that these hypotheses could not be tested without more extensive data.

To conclude, Pilot Experiment 3, which employed an interactive grid-matching system using /CVC/-/#V#/, was shown to reduce the time and effort of subjects normally inevitable in a matching experiment. The results suggest that in F1 matching, the trajectory range may have some effect, and, in F2 matching, the location of F1 in frequency range, together with the relative distance of two formants, could trigger a lower F2 match, although these observations are not based upon statistically sufficient data.

4.5 Discussion of the Pilot Experiments and the main experimental design

In 4.3 and 4.4. Pilot Experiments 2 and 3 confirmed that the grid matching scheme can be managed by the subjects and the introduction of this scheme has a potential not present in the ordinary similarity judgement experiment. There are, however, some remaining problems. This section 4.5 discusses these and proposes a framework for the main experiment.

First of all, the number of subjects participating in Pilot Experiments 2 and 3 was not sufficient. Therefore, the main experiment should examine more subjects so as to increase the power of the statistical tests, such as the F-tests used in the Repeated Measures ANOVA, by increasing their degrees of freedom.

The second problem is to do with the burden of the test on the subjects. The grid matching scheme enables a listener to perform a large number of vowel quality judgements which could never be achieved by normal experimental designs. However, even though the whole session was divided over two days, each subject had to spend up to 1.5 hours listening attentively per day. Therefore in the main experiment some improvements need to be incorporated in two aspects: a revision of the software / hardware used in the main experiment, and a re-organisation of types / numbers of stimuli.

During Pilot Experiment 3, some subjects remarked that sometimes the program was running at such a slow speed that there was a noticeable delay on playback; at worst, it took 2-3 seconds to have a stimulus pair replayed after moving a cursor to another block; or sometimes even if they tried to move a cursor, it was 'frozen' on one point for several seconds before it started to move. The problem was due to the overloading of the Sun SPARC workstation which controlled the test terminal. Since it is a multi-user system, its operating load varied, which affected the testing. One subject in Pilot Experiment 3 reported that the delay affected his concentration.

Therefore, the program needed to be revised and designed to operate in a non multi-tasking environment. The details of this implementation will be discussed in Chapter 5.

Another point to be improved was the cursor operation. As was argued in 4.2.5, the virtue of introducing the interactive grid matching method is that subjects can reduce the number of very different quality matches of /CVC/ and /#V#/ presented, by concentrating on a narrow region on a grid. However, the pilot program used arrow-keys to change the cursor position on a grid, and it took a considerable amount of time to traverse the grid. It would be better if a subject could have "hopped" from one block in a grid to another in a remote position without passing through intermediate blocks. For that purpose, the program should use a mouse so that a subject could change the position of a cursor simply by clicking on the grid.

There was also some reorganisation of the stimuli: Pilot Experiment 3 had two durational settings for /CVC/ reference tokens, but the results showed that the duration did not seem to have a crucial effect in F1 matching, and in F2 matching it only triggered F2 lowering: i.e. the matched F2 frequencies became lower than the /CVC/ trajectory peak values in some tokens and shorter duration seems to encourage this effect. In the light of this observation and the potential effect of low F1 on F2 matching, it seemed sensible that in the following main experiment, the durational effect could be set aside, using only one duration, in order to control all other parameters, and in order to reduce the number of the stimulus blocks presented to subjects. Since the low F1 effect on F2 matching seems to have been more prominent in a token with shorter duration, the token with 120 ms duration was adopted.

Another point concerns the vowels used: in Pilot Experiment 1, four British English monophthongs /ɛ,æ,ɒ,ʊ/ were used because in a traditional vowel quadrilateral or an acoustic vowel diagram they are well separated from each other without any overlaps. In contrast, Pilot Experiments 2 and 3 investigated all six British English monophthongs. The results of Pilot Experiments 1 and 2 suggested that the /ɪ/ did not show any particular behaviour and was quite similar to /ɛ/ in terms of F1 and F2 matching regardless of the consonantal and durational context. With regard to /ʌ/, it was observed in Pilot Experiment 3 that its matched F1 was close to the trajectory peak as in /ɒ/ and /æ/, and its matched F2 showed F1 matched to the trajectory peak as in /æ/ in both consonantal and durational environments. Thus, it was decided that the main experiment should concentrate on the original four vowels.

Finally in Pilot Experiment 3, each /CVC/-/#V#/ stimulus block was repeated four times and from a statistical point of view their number was not satisfactory. Since the number of vowels used in the main experiment can be reduced from six to four, it is possible to increase the number of repetitions from four to six, and this leads to 4 [vowel type] x 2 [consonants in /CVC/] x 6 [repetitions]= 48 matching sessions per subject, which is considerably reduced in comparison with Pilot Experiment 3 (6 [vowel type] x 2 [consonants] x 4 [repetitions] x 2 [duration types] = 96). In response to the criticism that six repetitions are still too small to produce a reliable data, it should be emphasised that in the grid-matching pilot experiments 13-24 playbacks per one /CVC/-/#V#/ matching were performed by subjects to make a decision, and one subject used 60 playbacks to judge one matching pair. Thus for six repetitions, the total number of /CVC/-/#V#/ playbacks per /CVC/ type would be between 78-144 presentations. Furthermore, even if the same number of repetitions were used in a normal XAXB test, more would be wasted because the grid scheme allows subjects to concentrate on plausible matches.

In summary, this chapter discussed whether an interactive matching experiment was appropriate to the aims of this study: to investigate how a listener would evaluate the quality of a vowel with dynamic formant trajectories. Pilot Experiments 1-3 revealed that 1) a subject could manage the task required by this interactive grid matching scheme; 2) the number of stimuli judged by a subject was more than an ordinary XAXB test could provide; and 3) the results showed a possibility of peak matching of /CVC/ trajectory both in F1/F2, with an influence possibly by the trajectory range in F1 and a low frequency F1 effect on F2. The main experiment in Chapter 5 studies whether the observations gained in the Pilot Experiments hold true, with a statistically adequate number of subjects and trials, while maintaining subject motivation.

Chapter 5: Main Experiment: testing vowel quality perception

5.1 Introduction

This chapter is concerned with how the main experiment addresses these aims:

> 1) to study how listeners evaluate the quality of a vowel with dynamic formant trajectories in a simulated /CVC/ context.
>
> 2) to enquire into whether these listeners, in judging the quality of a dynamic vowel, would follow some hypotheses proposed in the field of categorical perception / single-formant perception like the temporal averaging or the target compensation by a Kuwabara function, or whether they would pursue an entirely novel perceptual strategy.
>
> 3) in connection with 2), to investigate whether the initial observations of Pilot Experiment 3 are confirmed in a larger experiment: (a) the effect of the /CVC/ formant trajectory on F1 matching, and (b) the effect of low F1 values on F2 matching.

As Section 4.5 argued, the grid matching experiment scheme is employed with some modification in response to the experience of Pilot Experiments 2-3.

5.2 Materials

As in Pilot Experiment 3, two types of tokens, the reference /CVC/ token and the test /#V#/ token, were created. The structure of these tokens was identical with that of the previous experiment: subjects could change the formant frequency of the test /#V#/ by moving a cursor in a grid while they could not change that of the reference /CVC/. Both tokens had only two formants. The consonants of the reference /CVC/ were

/d_d/ or /b_b/, and four vowels /ɛ,æ,ɒ,ʊ/. Therefore the number of the reference token types was 2 (consonantal frame) x 4 (vowel types) = 8 (reference token types). Each reference token type (i.e. each /bɛb/, /bæb/ ... /dʊd/) had six repetitions, producing 8 x 6 = 48 matching sessions. The order of presentation of these matching sessions was randomised.

The formant trajectories were calculated according to the formula by Nearey by exactly the same procedure described in 4.2.2. F0 values, intensity specification, and modification to enhance the identifiability of /d/ consonants also followed the parameter settings of Pilot Experiments 2-3. The duration of the reference /CVC/ token was 120 ms. The test /#V#/ token had a duration of 220 ms.

5.3 Subjects

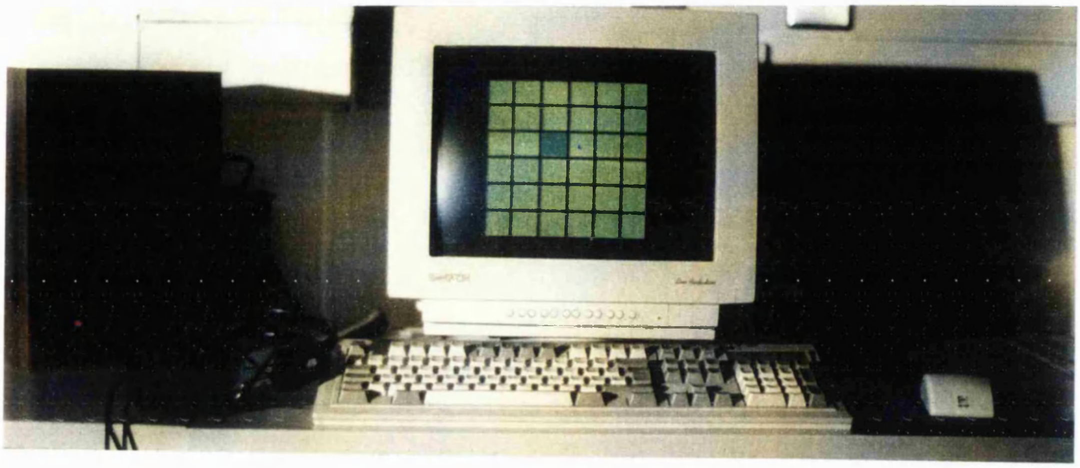
15 native speakers of South East British English participated in this experiment. They were undergraduate students of BA in Linguistics, or BSc in Speech Science/Speech Communication, at University College London. They had no history of hearing problems. They had taken several phonetics/phonology courses and at least one course involving phonetic ear-training sessions, and that fact assured that they brought adequate background knowledge to the task in this experiment. For attendance over the whole experimental period, each subject was paid four pounds. They were not informed of the nature and aim of this experiment before its end.

5.4 Procedure

The experiment was held in the teaching laboratory of Wolfson House, Department of Phonetics, University College London, not in the sound-proof room used in Pilot Experiments 2-3, to accommodate more than one subject at each time slot. This was made possible because the matching software was converted to a DOS version, which enabled simultaneous operation on a number of PCs. The laboratory was kept quiet by removing sources of background noise as much as possible, and subjects listened to the stimuli through covered-ear headphones. None of them reported that their attention had been compromised by background noise. At most three subjects were tested at one time.

Experimental instruction was given to the subjects in the following way: they were given a written explanation of the task required in the experiment, which is shown in Appendix VI. After reading it, they were given an oral instruction presented by the author, using one PC terminal in the laboratory, simulating the task that they were to do in the main experiment. Enquiries about the task of the experiment and the way to operate a PC and its mouse were answered during this oral presentation. The subjects then had one trial matching session and any further questions were answered. In addition, they were instructed to have a break between each session if they became tired.

The task of the subjects was the same as in Pilot Experiment 3. Sitting in front of a PC terminal showing a 6 x 6 grid, they were required to match the vowel quality of the reference /CVC/ and the test /#V#/, as the latter changed its F1/F2 according to the cursor position on the grid, which was moved by clicking with a mouse. The interval between the reference /CVC/ and the test /#V#/ was 300 ms, as in Pilot Experiment 3. When the subjects moved the cursor into a new block or when the space key was pressed, the pair of the reference /CVC/ and the test /#V#/ was replayed. After they had found the block whose test token /#V#/ had the most similar vowel quality to the reference /CVC/, they were instructed to press the "q" key, and the next stimulus pair and grid were presented. Figure 5-1 shows photographs of the actual experimental setup.



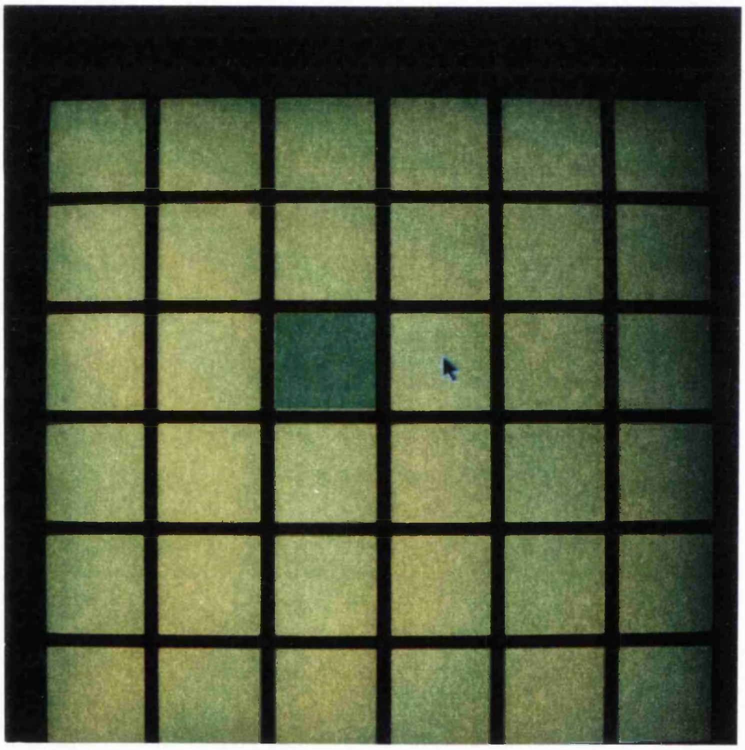


Figure 5-1 : Photographs of the actual setup of the main experiment. Each PC plays the stimuli through the headphones as shown in the upper photograph. The lower photograph shows the grid and the cursor which subjects see on the screen.

The allocation of the direction of F1/F2 on the two axes was randomised as in Pilot Experiment 3. Although the grid in this experiment was apparently identical with that in Pilot Experiment 3, (6 x 6 grid on the screen in both experiments), the steps of the formant frequency change were altered to 0.5 Bark on both F1/F2 axes. This modification was done for consistency of the step along both axes. There was no particular block whose F1/F2 values exactly corresponded to the peak F1/F2 values of /CVC/ as before. To avoid a bias to the response due to location in the grid, the subjects were reminded during the oral instruction that the block whose test token had the most similar vowel quality to the reference token could be in the peripheral 6 x 6 area of the grid. The grid cell with values closest to the F1/F2 peak values of the reference /CVC/ was randomly assigned to one of the central 4 x 4 cells. The F1/F2 values used in the grids are shown in Table 5-1 below.

/bɛb/ F1	300	354	405	461	516	580
/bɛb/ F2	1600	1727	1863	2009	2165	2333
/bæb/ F1	600	664	732	804	880	961
/bæb/ F2	1300	1407	1521	1643	1724	1913
/bɒb/ F1	450	507	567	631	697	767
/bɒb/ F2	800	875	955	1041	1131	1227
/bʊb/ F1	250	300	352	406	461	520
/bʊb/ F2	800	875	955	1041	1131	1227
/dɛd/ F1	300	354	405	461	516	580
/dɛd/F2	1600	1727	1863	2009	2165	2333
/dæd/F1	600	664	732	804	880	961
/dæd/F2	1300	1407	1521	1643	1724	1913
/dɒd/F1	450	507	567	631	697	767
/dɒd/F2	800	875	955	1041	1131	1227
/dʊd/F1	250	300	352	406	461	520
/dʊd/F2	1000	1087	1181	1280	1386	1499

Table 5-1: Formant values used in /#V#/ token to produce a 6 x 6 grid. All values in Hz.

5.5 Results and Discussion

5.5.1 Adequacy of the obtained results

First of all, it was found that each of the 15 subjects finished the 48 sessions within 45 minutes. All subjects claimed that the experimental task was manageable. Some of them took a short break, as recommended. This seems to indicate that thanks to its relatively short time duration and its self-pacing nature, the grid-matching scheme did not excessively test the attention span of subjects.

In order to validate the consistency and reliability of subject performance, the two following criteria for subject-selection were utilised:

a) "perseverance": if subjects choose the centre-4 blocks (the very centre 2 x 2 blocks in the grid) as a response in ALL matching sessions, their results are likely to follow from a non-auditory strategy, and therefore these should be excluded.

b) "consistency": for each subject, all F1/F2 values selected for one token type (one of /bɛb/, /bæb/.... /dʊd/) should usually fall within the range of three F1 or three F2 steps across all six repetitions. The three step range was chosen under the assumption that all consistent responses should target at one particular block, with one step up/down as an error range. Then, over all eight token types, if ONE of the F1/F2 response ranges exceeds three steps in MORE THAN four token types out of eight (i.e. more than a half of all token types), then the results of the subject should be excluded. This is illustrated in Figure 5-2 below.

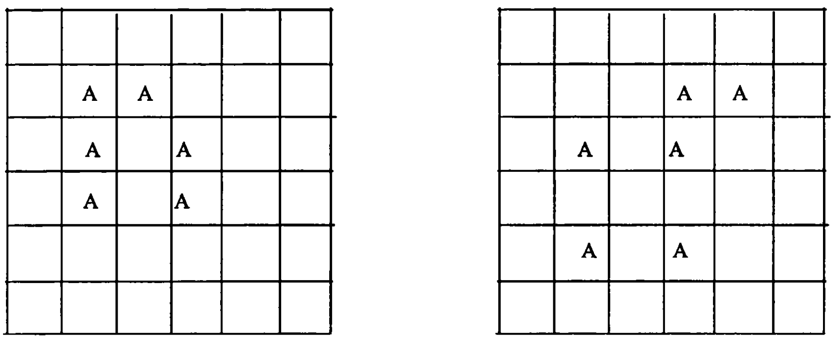


Figure 5-2: Illustration of the "consistency" criterion. The location of the subject's response is displayed as the letter "A". The response pattern of the grid on the left can pass this criterion since the range of the step is 3 x 3 (vertical x horizontal), while the grid on the right cannot pass it because the step is 4 x 4 (vertical x horizontal).

The first criterion a) was used to exclude subjects who could not follow the experimental instructions, and the second criterion b) was used to exclude subjects who lacked consistency.

The results according to the criterion a), the frequency of centre-4 choices, are shown in Table 5-2. Each vertical column stands for token types and each horizontal row corresponds to the individual subject. Within each block the number of occurrences of centre-4 block responses (maximally six, due to six repetitions made for each token type) is indicated.

	bɛb	bæb	bɒb	bʊb	dɛd	dæd	dɒd	dʊd
Subj1	2	4	4	1	3	5	4	4
Subj2	2	2	3	0	1	4	4	3
Subj3	2	4	1	3	5	4	4	3
Subj4	2	1	0	0	2	2	0	1
Subj5	4	2	4	3	4	4	5	3
Subj6	2	5	4	3	3	4	1	2
Subj7	3	5	1	2	5	0	4	3
Subj8	0	2	0	0	0	0	1	1
Subj9	2	2	3	1	2	0	1	0
Subj10	0	6	3	0	3	3	1	4
Subj11	4	6	4	3	3	2	5	4
Subj12	6	6	6	6	6	6	6	6
Subj13	1	6	2	1	3	3	2	0
Subj14	0	2	2	1	0	3	0	1
Subj15	0	0	1	2	0	0	0	0

Table 5-2: Number of centre-4 choices: the number of the responses that fell within the centre 2 x 2 blocks in the 6 x 6 grid. The columns correspond to the token type, while each row represents each subject. "Subj" stands for "Subject"

Table 5-2 reveals that only subject 12 violates the criterion a), since all the responses fell within the centre 4 x 4 blocks, and that implies that subject 12 probably adopted the strategy of selecting the visual centre-4 blocks on the screen without involving auditory judgement. Therefore subject 12 was excluded from the analysis.

The results of the investigation of the F1/F2 ranges are shown on Table 5-3 on the next page. As in Table 5-2, the columns stand for token types and each row is for an individual subject. The numbers in each box of Table 5-3 are F1 range x F2 range. For example, 5 x 4 stands for the result where the F1 range of the matching responses by a subject is within five F1 steps while the F2 range of the matching response is within four F2 steps. In Table 5-3, if the F1 x F2 range is 3 x 3 or smaller, an open circle, not the actual range value, is inserted, to show that the particular trials passed the criterion b).

Table 5-3 reveals that the subjects 8, 14 and 15 did not fulfil the criterion b) since their responses produced a larger F1 x F2 range than 3 x 3 in 6 or 7 tokens types. This is worse than criterion b) and it suggests a quasi-random response by these subjects. Subjects 8, 14, and 15 were also therefore excluded. The analysis of the results of the remaining 11 subjects is discussed in the following sections.

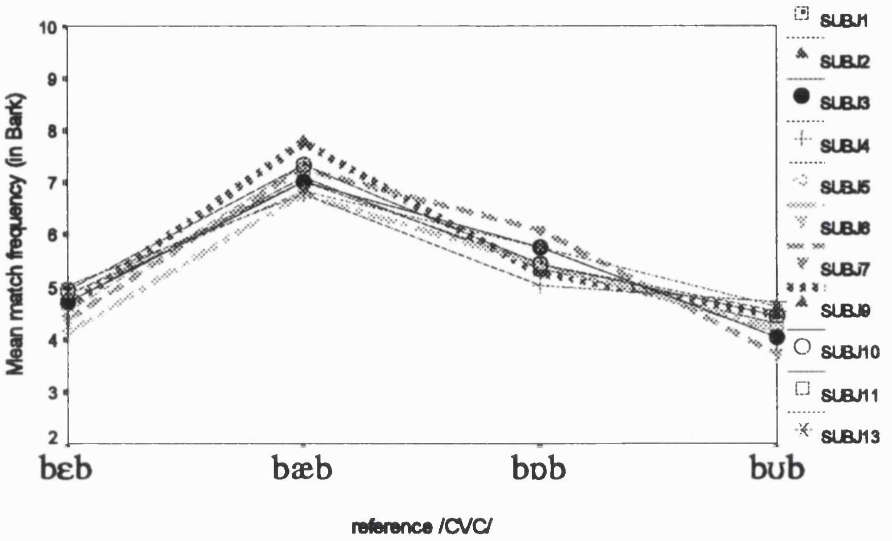
	bɛb	bæb	bɒb	bʊb	dɛd	dæd	dɒd	dʊd
Subj1	O	O	O	O	4 X 3	O	O	4 X 5
Subj2	4 X 3	4 X 5	2 X 4	O	O	O	O	4 X 3
Subj3	O	O	4 X 2	4 X 2	O	O	O	3 X 4
Subj4	3 X 4	O	4 X 4	O	3 X 4	2 X 5	O	O
Subj5	O	O	O	4 X 3	O	O	O	4 X 3
Subj6	5 X 5	O	O	4 X 2	O	O	O	4 X 3
Subj7	3 X 4	O	O	4 X 2	O	O	O	4 X 3
Subj8	4 X 5	O	O	4 X 4	4 X 4	4 X 4	4 X 3	4 X 3
Subj9	4 X 3	4 X 3	O	O	4 X 2	4 X 2	O	O
Subj10	O	O	O	O	O	O	O	O
Subj11	O	O	O	O	3 X 4	4 X 4	O	O
Subj12	O	O	O	O	O	O	O	O
Subj13	O	O	3 X 4	O	O	O	2 X 4	O
Subj14	4 X 4	3 X 4	3 X 4	4 X 4	4 X 4	O	2 X 4	5 X 3
Subj15	4 X 4	4 X 3	4 X 4	3 X 4	O	5 X 6	5 X 5	3 X 4

Table 5-3: Range of matched F1/F2 values of test /#V#/ tokens. The values indicated are (F1 variation steps) X (F2 variation steps) pooled across all six matching sessions per subject. "O" indicates that the all six responses of that subject in that token were within 3 steps both in F1 and F2, which is determined in the criterion b).

5.5.2 Statistical analysis on subject homogeneity

Before obtaining the mean shift index across the 11 remaining subjects, as in the Pilot Experiments 1 and 2, mean matched frequencies were calculated across all six trials for each subject, /CVC/ reference token type and formant. They are plotted in Figures 5-3 and 5-4 to represent the general behaviour of each subject.

F1 result: /bVb/



F1 result: /dVd/

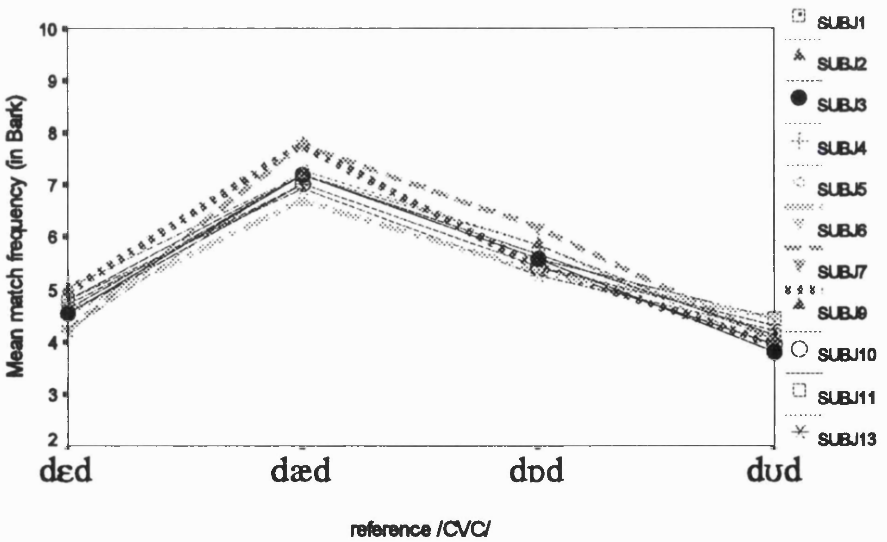
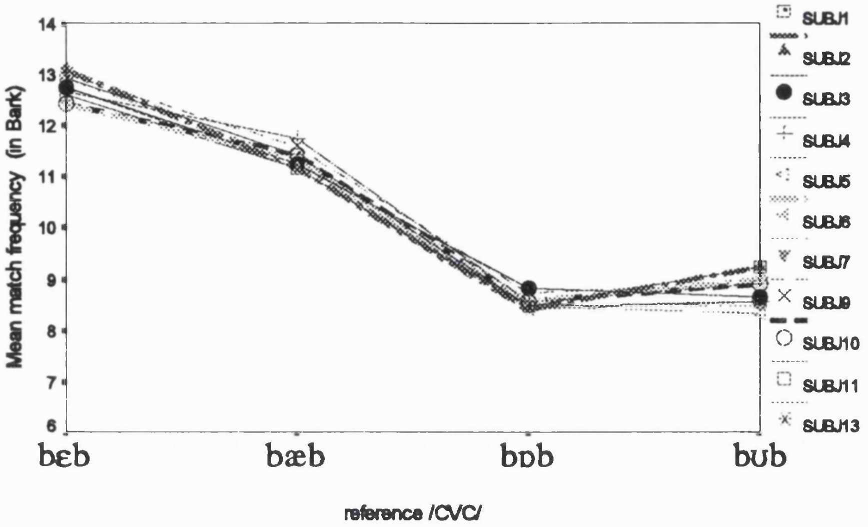


Figure 5-3: Mean matched frequency of each subject: F1

F2 result: /bVb/



F2 result: /dVd/

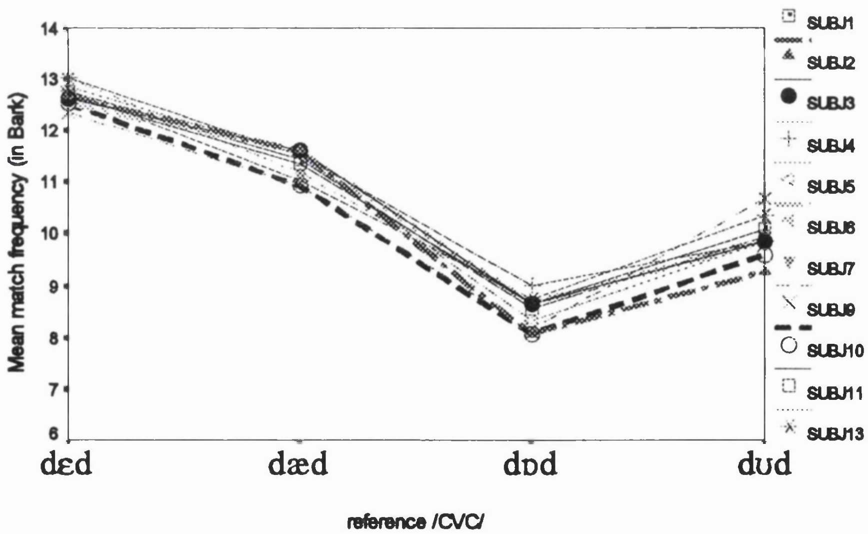


Figure 5-4: Mean matched frequency of each subject: F2

Figures 5-3 and 5-4 show that the matching pattern of 11 subjects appears similar, [illegible] Consequently, Repeated Measures ANOVA was manually carried out to examine whether these subjects were statistically homogenous, with factors of [subject] (11-levels), [consonant] (2-levels), [vowel] (4-levels) and [trial] (6-levels). The highest order interaction, which was employed as a denominator for the F-tests, was [subject]*[consonant]*[vowel]*[trial]. For this analysis, the shift index ([given matched formant value]-[its corresponding /CVC/ trajectory peak]) was obtained in Bark scale for each measurement since in this main experiment the F1/F2 frequency scales in the grid were all represented in Bark scale. The analysis was made separately on each formant. SPSS for Windows (Version 6.1.3) was used as in the previous experiments.

With regard to F1, the analysis showed that [subject] as a main factor was significant ($F(10,150)=3.01$, $p<.01$) and the interaction [subject]*[vowel] was also significant ($F(30,150)=5.32$, $p<.01$), although the interaction [subject]*[consonant] was not significant ($F(10,150)=1.25$, $p>.01$). The analysis of F2 matching confirmed a difference between subjects: subject as a main factor was significant ($F(10,150)=3.06$, $p<.01$), and the interactions [subject]*[vowel] ($F(30,150)=1.99$, $p<.01$) and [subject]*[consonant] ($F(10,150)=3.06$, $p<.01$) were both significant.

Thus the 11 subjects were shown not to form a homogenous group, making it necessary to create subgroups, and therefore the following procedure was undertaken. First, cluster analysis was made on the mean shift index across six trials, for each subject, consonantal context and vowel type. The complete result of this cluster analysis is shown in Appendix VII. Then a subject with a distinctive response pattern was eliminated according to the results of the cluster analysis and the observation from Figures 5-3 and 5-4. This process was repeated until the F-ratio of the Repeated Measures ANOVA was less than that of the 1% significance level, for [subject] as a main factor and its interaction by [consonant] and [vowel]. This elimination process was carried out for each formant type across all vowel types, and eventually created two subject groups for F1 and two groups for F2. The two subject groups for F1 are henceforth called Groups A and B and the two subject groups for F2 are called Groups X and Y. Table 5-4 below displays the subjects belonging to each group and the F-ratios obtained from the Repeated Measures ANOVA. The index of each subject corresponds to that used in Tables 5-2 and 5-3.

F1 Subject Grouping

Group	A	B
subjects	1,2,3,4,5,10,11,13	6,7,9
[subject] as a main factor	F(7,105)=1.93 p>.01	F(2,30)=3.26 p>.01
[subject]*[consonant]	F(1,105)=0.43 p>.01	F(2,30)=0.23 p>.01
[subject]* [vowel]	F(21,105)=1.81 p>.01	F(6,30)=2.36 p>.01

F2 Subject Grouping

Group	X	Y
subjects	1,3,4,5,6,7,10,11,13	2,9
[subject] as a main factor	F(8,120)=2.40 p>.01	F(1,15)=1.60 p>.01
[subject]*[consonant]	F(1,120)=2.26 p>.01	F(1,15)=0.05 p>.01
[subject]*[vowel]	F(24,120)=2.13 p>.01	F(3,15)=1.65 p>.01

Table 5-4: Homogenous subject groups obtained from the Repeated Measures ANOVA. F-ratios of Subject as a main factor, and of interactions between [subject] and other factors are also displayed.

It is observed from Table 5-4 that the subgroups are different for F1 and F2, implying that certain subjects might have used different matching criteria for the different formants. This point is discussed later in this chapter.

From these groups, further repeated measures ANOVAs were made to examine whether [vowel] and [consonant] as a main factor had any effect on subjects' choice, and the resulting F-ratios are displayed in Table 5-5 below. Shaded cells indicate that the F-ratios there are statistically significant at the 1% level.

F1 Subgroups

Group	A	B
[consonant] as a main factor	$F(1, 105)=2.62$ $p>.01$	$F(1,30)=2.10$ $p>.01$
[vowel] as a main factor	$F(3,105)=77.25$ $p<.01$	$F(3,30)=47.15$ $p<.01$

F2 Subgroups

Group	X	Y
[consonant] as a main factor	$F(1, 120)=23.53$ $p<.01$	$F(1,15)=41.20$ $p<.01$
[vowel] as a main factor	$F(3,120)=103.86$ $p<.01$	$F(3,15)=24.70$ $p<.01$

Table 5-5: Results of the repeated measures ANOVA for individual subgroups, examining the effect of factors [consonant] and [vowel] as a main factor

The results show that F1 matching was not affected by the surrounding consonant of the reference /CVC/ token but by its vowel type, since in both groups A and B, [consonant] as a main factor was not significant but [vowel] as a main factor was. This is clearly the result of the F1 trajectory being identical in /b/ and /d/. They also show that F2 matching was affected by both consonant and vowel types of the reference /CVC/ token because both [consonant] and [vowel] as a main factor were significant in groups X and Y. These observations could be ascribed to the fact that, under the assumption that formant trajectory shape influences the matching strategy of subjects, the F1 trajectory shape is convex regardless of the consonantal environment while the F2 trajectory shape is concave in /dVd/ but convex in /bVb/.

Next section 5.5.3 presents three hypotheses from previous studies and investigates whether they can account for the data of this study.

5.5.3 Interpretation of the results: three initial hypotheses

To study the general tendency of the matching process, the mean shift index was calculated across subjects in each subject group for each formant and for each vowel and consonant type, and displayed in Table 5-6 below.

F1

Group A

bɛb	bæb	bɒb	bʊb	dɛd	dæd	dɒd	dʊd
-.48	.11	.09	-.46	-.46	.39	.06	-.31

Group B

bɛb	bæb	bɒb	bʊb	dɛd	dæd	dɒd	dʊd
-.97	.31	.25	-.75	-.61	.64	.19	-.25

F2

Group X

bɛb	bæb	bɒb	bʊb	dɛd	dæd	dɒd	dʊd
.54	-.13	.00	-.45	.36	-.15	-.66	-.31

Group Y

bɛb	bæb	bɒb	bʊb	dɛd	dæd	dɒd	dʊd
.54	-.20	-.07	-.07	.23	-.23	-1.19	-.95

Table 5-6: Mean shift index over subjects in a homogenous group. All values in Bark

Table 5-6 indicates that the strategy of subjects seems to be complex. The mean shift index fluctuates across all F1 and F2; the peak F2 frequencies of /dɒd/ (1135 Hz) and /bʊb/ (1122 Hz) are close but the mean shift index of /dɒd/ is far greater; and some clear differences between the two subject groups are also observed, e.g., in /bʊb/ F2.

In the perceptual processing of dynamic formant trajectories, subjects may have pursued a matching strategy as proposed in previous studies or an entirely novel matching strategy. However the verification of these hypotheses should be accompanied by appropriate statistical analysis since the mean shift that appears substantial in Table 5-6 may be statistically insignificant.

Therefore three possible hypotheses that were proposed in the previous studies or developed through Pilot Experiments were examined statistically in order to investigate how the subjects evaluated the quality of vowels with dynamic trajectories. They are as follows:

(1) peak picking hypothesis: listeners refer to the peak values of the /CVC/ trajectories of the reference token (henceforth called "peak values") to evaluate the vowel quality of the reference /CVC/. This hypothesis is derived from the data of Pilot Experiment 2, where /#V#/ tokens were matched to /#V#/.

(2) compensation hypothesis by Kuwabara function: listeners auditorily "compensate" undershoot formants by the Kuwabara function (henceforth called "K-target values"). The actual function was described in 4.2.5 in detail. This Kuwabara function was proposed as one of the alternative algorithms in the auditory theory of vowel perception by Rosner & Pickering (1994). (Recall that Rosner & Pickering do not present a concrete algorithm of their own.) The K-target values calculated according to the Kuwabara function, as well as the trajectory /CVC/ peak values, are displayed in Table 5-7.

	bɛb	bæb	bɒb	bʊb
F1 peak	542	760	545	491
F1 K-target	673	963	677	605
F2 peak	1806	1621	1009	1122
F2 K-target	2175	1928	1115	1263

	dɛd	dæd	dɒd	dʊd
F1 peak	526	733	565	437
F1 K-target	651	928	703	533
F2 peak	1848	1619	1135	1359
F2 K-target	1797	1492	846	1145

Table 5-7: Trajectory /CVC/ peak values and K-target values calculated by the Kuwabara function. All numbers in Hz.

(3) isolated vowel hypothesis: listeners compensate with formant values taken from the vowels uttered in isolation (henceforth called "isolated vowel formant values"). The isolated formant values were used in Pilot Experiment 1 in 4.2.

Wilcoxon tests were then applied to test whether the difference between the matched and predicted formant values could be explained by one of these three hypotheses.

To obtain the formant shift data for testing, the following steps were taken. First, the differences of the F1 and F2 matches from K-targets and from the isolated vowel were calculated in Bark scale for each subject. Then Repeated Measures ANOVA for each formant number and type of difference was performed across all 11 subjects, followed by the same subgrouping process that was done on the shift index.

The Repeated Measures ANOVA resulted in the same subject grouping as was formed for the shift index: two identical subject groups for F1 and two for F2. Tables 5-8 and 5-9 below present the F-ratios of the Repeated Measures ANOVA for K-targets and isolated vowel formant values.

F1 Subject grouping on differences on matched values from K-target values

Group	A	B
subjects	1,2,3,4,5,10,11,13	6,7,9
[subject] as a main factor	F(7,105)=2.31, p>.01	F(2,30)=4.12, p>.01
[subject]*[consonant]	F(1,105)= 2.82, p>.01	F(2,30)=0.37, p>.01
[subject]*[vowel]	F(21,105)=2.00, p>.01	F(6,30)= 1.25, p>.01

F2 Subject grouping on differences on matched values from K-target values

Group	X	Y
subjects	1,3,4,5,6,7,10,11,13	2,9
[subject] as a main factor	F(8,120)=2.37, p>.01	F(1,15)=1.00, p>.01
[subject]*[consonant]	F(1,120)=2.43, p>.01	F(1,15)=0.52, p>.01
[subject]*[vowel]	F(24,120)=1.12, p>.01	F(3,15)=2.00, p>.01

Table 5-8: Subject groups obtained from Repeated Measures ANOVA, calculated on the difference of F1 and F2 from K-target values. Its F-ratios of [subject] as a main factor, and of interactions between [subject] and other factors are displayed.

F1 Subject grouping on differences of matched formant values from isolated vowel formant values

Group	A	B
subjects	1,2,3,4,5,10,11,13	6,7,9
[subject] as a main factor	F(7,105)=2.21, p>.01	F(2,30)=4.35, p>.01
[subject]*[consonant]	F(1,105)=0.50, p>.01	F(2,30)=0.07, p>.01
[subject]*[vowel]	F(21,105)=2.00, p>.01	F(6,30)=1.57, p>.01

F2 Subject grouping on differences of matched formant values from isolated vowel formant values

Group	X	Y
subjects	1,3,4,5,6,7,10,11,13	2,9
[subject] as a main factor	F(8,120)= 2.39, p>.01	F(1,15)=1.90, p>.01
[subject]*[consonant]	F(1,120)=2.27, p>.01	F(1,15)=0.05, p>.01
[subject]*[vowel]	F(24,120)=2.05, p>.01	F(3,15)=1.65, p>.01

Table 5-9: Subject groups obtained from Repeated Measures ANOVA, calculated on the difference of F1 and F2 from isolated vowel formant values. Its F-ratios of [subject] as a main factor, and of interactions between [subject] and other factors are displayed.

Subsequently, these two types of differences (i.e., those from isolated vowel formant values and those from K-target values) were combined over homogenous subgroups of subjects for each reference /CVC/ token type and for each formant, and Wilcoxon tests were carried out on these two types of difference, and also on the shift index (obtained by [matched frequency] - [trajectory peak value]), for each formant number, each reference token type and each subject group, in order to test whether the differences are statistically significant, as in Pilot Experiments 2-3.

The null hypothesis was that the median shift index to the /CVC/ trajectory peak was 0.0 (for the peak picking hypothesis); and that the median difference to the K-target values and the isolated vowel formant values was 0.0 (for the compensation hypothesis by Kuwabara function and the isolated vowel hypothesis). These Wilcoxon tests followed the Bonferroni procedure as in Pilot Experiments, and the total significance level .01 was divided by the number of comparisons, 192 (i.e. 4 vowels x 2 consonants x 2 formants x 4 subject groups x 3 types of difference), producing the individual significance level of .01 / 192 = .0001. The resulting z-values are displayed in the following Tables 5-10 and 5-11. Shaded cells signify that the z-values there are significant at the significance level of .0001. Note that in Groups B (3 subjects) and Y (2 subjects), no values of Wilcoxon's z are significant at the level of .0001, due to the small number of observations.

F1: Group A

	bɛb	bæb	bɒb	bʊb	dɛd	dæd	dɒd	dʊd
peak	-5.415	-3.343	-0.215	-6.020	-5.712	-4.668	-3.005	-5.466
K-target	-6.031	-6.031	-5.969	-6.031	-6.031	-6.031	-6.031	-6.031
iso	-6.031	-5.928	-4.953	-3.671	-6.031	-5.928	-4.410	-2.277

F1: Group B

	bɛb	bæb	bɒb	bʊb	dɛd	dæd	dɒd	dʊd
peak	-3.636	-2.242	-1.981	-3.723	-3.462	-2.981	-2.199	-2.809
K-target	-3.723	-3.723	-3.592	-3.723	-3.723	-3.592	-3.723	-3.723
iso	-3.723	-2.286	-2.199	-0.675	-3.723	-0.849	-1.415	-0.718

Table 5-10: z-values obtained by Wilcoxon tests for the two F1 subject groups, for each token type and difference type. "peak" represents z-values of Wilcoxon test for shift index; "K-target" represents those for the differences of matched formant frequencies from K-target values; and "iso" represents those for the difference of matched formant frequencies from isolated vowel formant values. z-values in shaded cells are significant on the level of .0001, the 1 % level equivalent specified by the Bonferroni procedure.

F2: Group X

	bɛb	bæb	bɒb	bʊb	dɛd	dæd	dɒd	dʊd
peak	-5.397	-2.069	-0.721	-5.226	-6.001	-1.635	-6.373	-4.037
K-target	-6.307	-6.393	-6.307	-6.393	-6.126	-4.765	-6.393	-6.393
iso	-0.607	-5.197	-4.628	-2.027	-1.623	-4.628	-4.059	-6.393

F2: Group Y

	bɛb	bæb	bɒb	bʊb	dɛd	dæd	dɒd	dʊd
peak	-2.554	-0.790	-0.161	-0.494	-2.565	-1.189	-3.115	-3.088
K-target	-3.059	-3.059	-2.981	-3.059	-2.745	-1.882	-3.059	-0.392
iso	-1.882	-2.196	-1.882	-2.588	-1.882	-2.588	-2.196	-2.667

Table 5-11: z-values obtained by Wilcoxon tests for the two F2 subject groups, for each token type and difference type. "peak" represents z-values of Wilcoxon test for shift index; "K-target" represents those for the differences of matched formant frequencies from K-target values; and "iso" represents those for the difference of matched formant frequencies from isolated vowel formant values. z-values in shaded cells are significant on the level of .0001, the 1 % level equivalent specified by the Bonferroni procedure.

First the K-target values derived from the compensation hypothesis by the Kuwabara function are examined. All z-values of Groups A and X in Tables 5-10 and 5-11 show a significant difference under the threshold level of $p < .0001$, the 1% equivalent in each Wilcoxon test, and therefore the null hypothesis was refuted. It shows that subjects did not refer to a formant value compensated by the Kuwabara function applied to /CVC/, denying the use of the K-target compensation strategy in vowel quality perception by the subjects.

Next the peak-picking hypothesis was examined: subjects referred to the trajectory peak of the reference /CVC/ syllables in their vowel quality evaluation. Z-values of Group A in Table 5-10 indicate that 5 out of 8 cases show a significant difference between matched formant values and their /CVC/ trajectory peak value. A similar tendency is observed in Group B, although no significant difference was obtained due to the small number of subjects. Table 5-11 shows that for the shift index, Group X has 4 cases out of 8 where z-values are significant. A similar trend is found in Group Y. This leads to the refutation of the null hypothesis for the trajectory peak matching strategy.

Finally the isolated vowel hypothesis that subjects compensated with the formant values in isolated utterances was investigated. In Table 5-10, significant differences between matched formant values and isolated vowel formant values were found in 6 cases in Group A, and Group B showed a similar tendency. In Table 5-11, 5 cases of Group X show a significant difference, while Group Y has a similar trend. Hence the null hypothesis was refuted.

In summary, the statistical analysis refuted all three hypotheses on vowel quality evaluation, i.e., (1) subjects referred to peak values of the /CVC/ trajectories of the reference token; (2) subjects referred to a compensated formant value using the Kuwabara function; and (3) subjects referred to formant values of utterances without any consonantal context. A new hypothesis or a modified version of one of three above is required to explain the experimental results. This is provided in the next section.

5.5.4 Interpretation of the results: a new hypothesis

Since the results in Pilot Experiment 3 of 4.4 suggest that the F1 trajectory range of the reference /CVC/ may affect the F1 matching of the test /#V#/, the relations between the formant trajectory range of the reference /CVC/ and the matched frequency of the test /#V#/ were investigated for both F1 and F2.

Tables 5-12 and 5-13 shows the relations between trajectory ranges and the mean shift index to the trajectory peak. Table 5-12 is for F1 results, Groups A and B, and Table 5-13 for F2 results, Groups X and Y. In each formant / consonantal environment, the order of the vowels is arranged so that the /CVC/ trajectory range increases from left to right.

Moreover, to verify whether a given mean shift index is substantial enough to be perceived by listeners, the difference limen was introduced for the analysis in Tables 5-12 and 5-13. For the difference limen of formant frequencies, Rosner & Pickering (1994:55) suggest a value between 0.03 and 0.05 for the Weber fraction. Hence the 0.05 Weber fractions for each /CVC/ formant trajectory peak values were calculated in Hz, converted into Bark and utilised as difference limens. Appendix VIII has a table of 0.05 Weber fractions in Bark scale. Shaded cells in Tables 5-12 and 5-13 show that their shifts are further than the difference limen at that frequency.

F1 Group A /bVb/ F1 (F1 range -----> increase)

	ʊ	ɛ	ɒ	æ
range	3.28	3.70	3.73	5.37
mean index	-.46	-.48	.09	.11

/dVd/ F1 (F1 range ------> increase)

	ʊ	ɛ	ɒ	æ
range	2.80	3.57	3.89	5.18
mean index	-.31	-.46	.06	.39

F1 Group B /bVb/ F1 (F1 range -----> increase)

	ʊ	ɛ	ɒ	æ
range	3.28	3.70	3.73	5.37
mean index	-.75	-.97	.25	.31

/dVd/ F1 (F1 range ------> increase)

	ʊ	ɛ	ɒ	æ
range	2.80	3.57	3.89	5.18
mean index	-.25	-.61	.19	.64

Table 5-12: Trajectory range of F1 /CVC/ and its mean shift index, written as "mean index" (= mean of all [matched formant value of /#V#/] - [its reference/CVC/ trajectory peak]) All values in Bark. Shaded cells show that the shift value is more than 0.05 Weber fraction of a /CVC/ formant trajectory peak frequency.

F2 Group X /bVb/ F2 (F2 range -----> increase)

	ɒ	ʊ	æ	ɛ
range	2.02	2.65	4.97	5.67
mean index	.00	-.45	-.13	.54

/dVd/ F2 (F2 range -----> increase)

	ɛ	æ	ʊ	ɒ
range	-0.52	-1.39	-2.51	-3.62
mean index	.36	-.15	-.31	-.66

F2 Group Y /bVb/ F2 (F2 range -----> increase)

	ɒ	ʊ	æ	ɛ
range	2.02	2.65	4.97	5.67
mean index	-.07	-.07	-.20	.54

/dVd/ F2 (F2 range -----> increase)

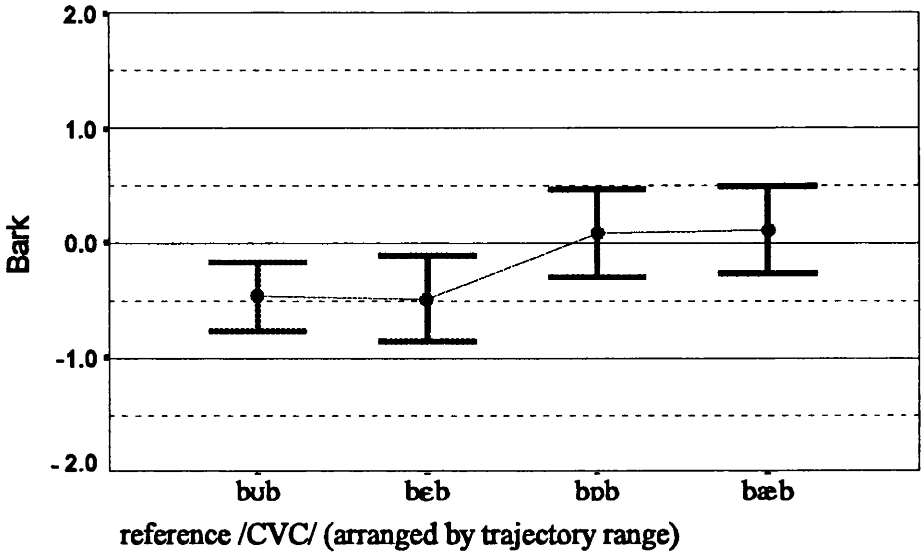
	ɛ	æ	ʊ	ɒ
range	-0.52	-1.39	-2.51	-3.62
mean index	.23	-.23	-.95	-1.19

Table 5-13: Trajectory range of F2 /CVC/ and its mean shift index, written as "mean index" (= mean of all [matched formant value of /#V#/] - [its reference/CVC/ trajectory peak]) All values in Bark. Shaded cells show that the shift value is more than 0.05 Weber fraction of a /CVC/ formant trajectory peak frequency.

This mean differences are also plotted in Figures 5-5 to 5-8, together with error bars of one standard deviation, which were obtained by accumulating all the individual frequency differences between the matched value and the /CVC/ trajectory peak according to each token type.

F1

Group A /bVb/



Group B /bVb/

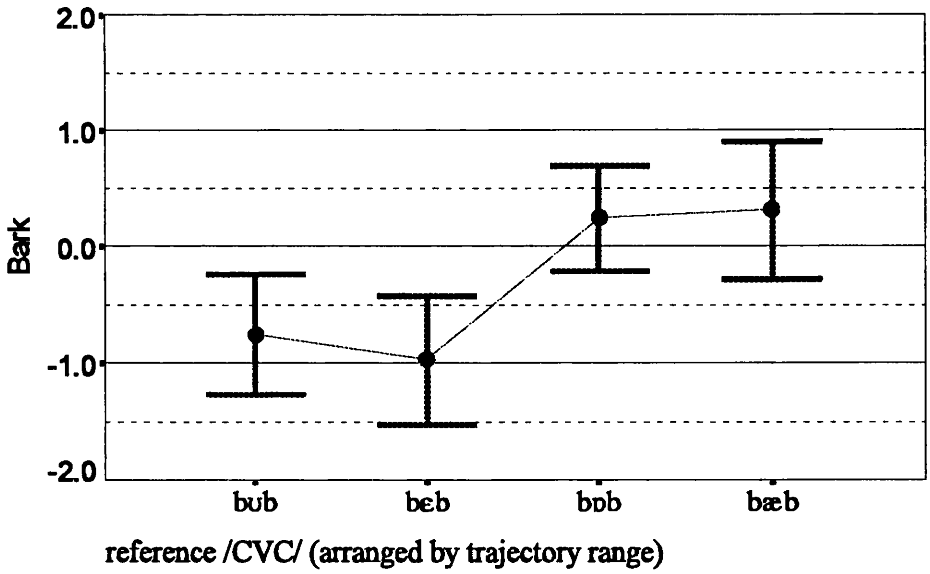
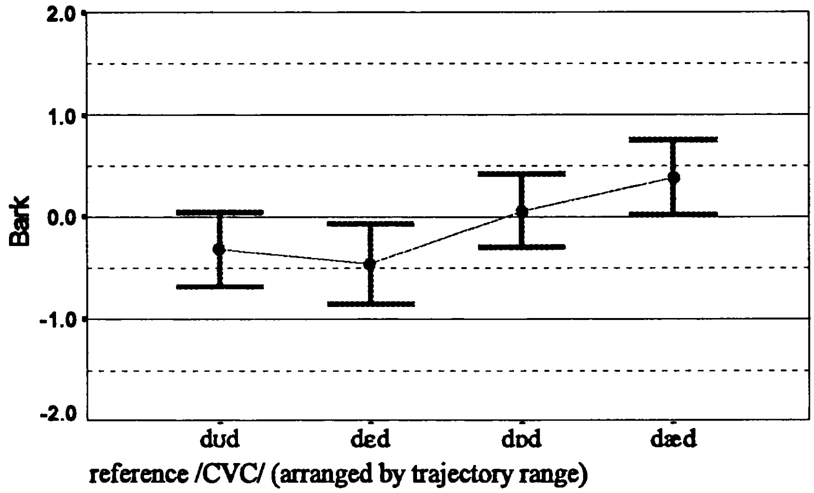


Figure 5-5: Mean shift index (circle in the middle) with an error bar of one standard deviation of /bVb/ F1. The reference /CVC/ tokens are arranged so that their trajectory range increases from left to right in X-axis.

F1

Group A /dVd/



Group B /dVd/

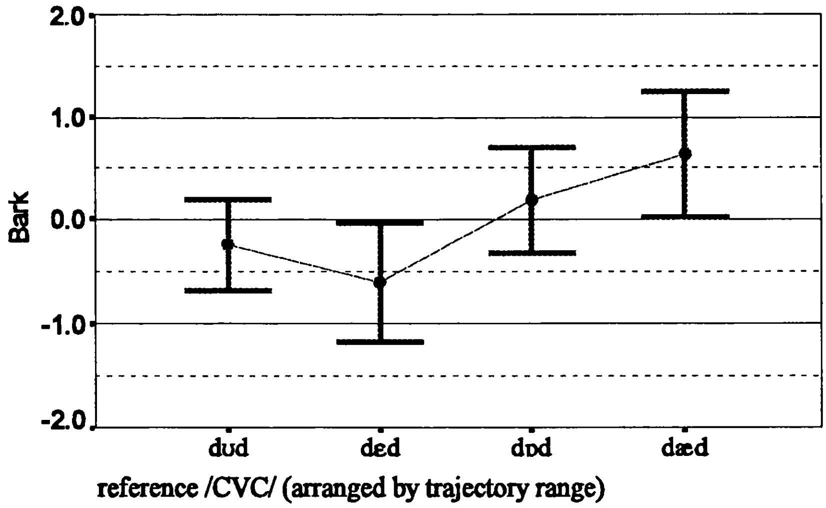
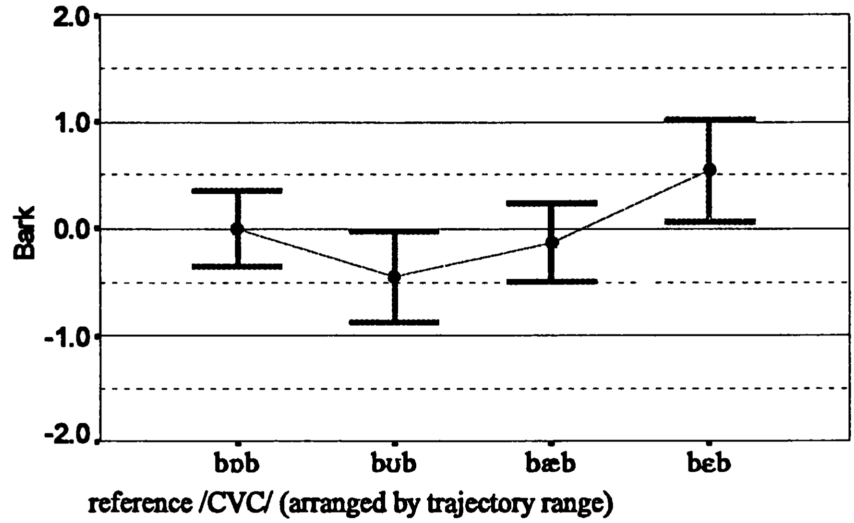


Figure 5-6: Mean shift index (circle in the middle) with an error bar of one standard deviation of /dVd/ F1. The reference /CVC/ tokens are arranged so that their trajectory range increases from left to right in X-axis.

F2

Group X /bVb/



Group Y /bVb/

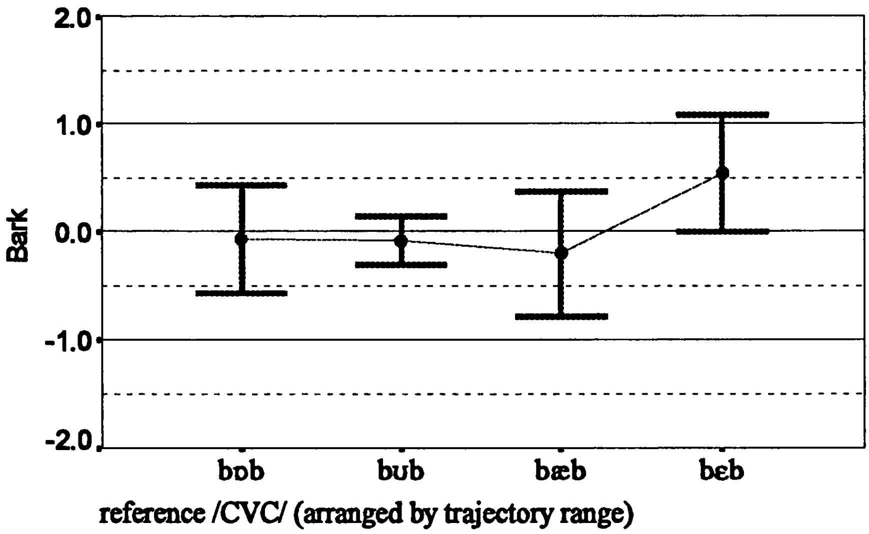
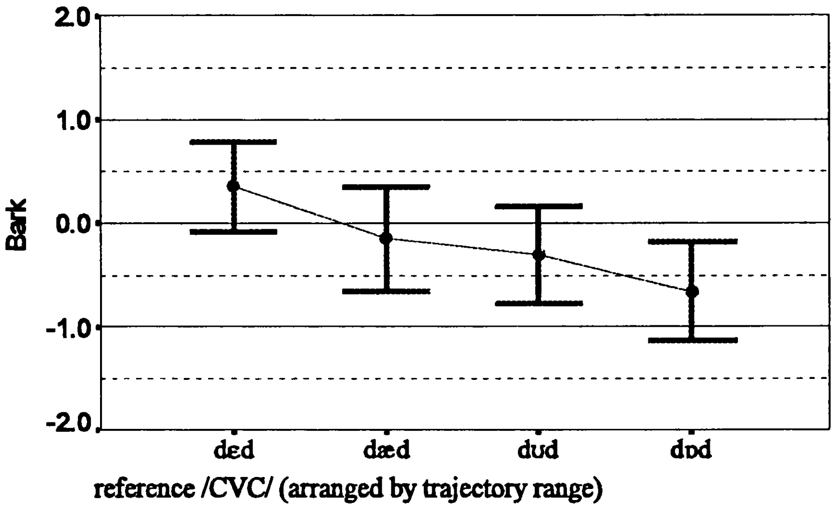


Figure 5-7: Mean shift index (circle in the middle) with an error bar of one standard deviation of /bVb/ F2. The reference /CVC/ tokens are arranged so that their trajectory range increases from left to right in X-axis.

F2

Group X /dVd/



Group Y /dVd/

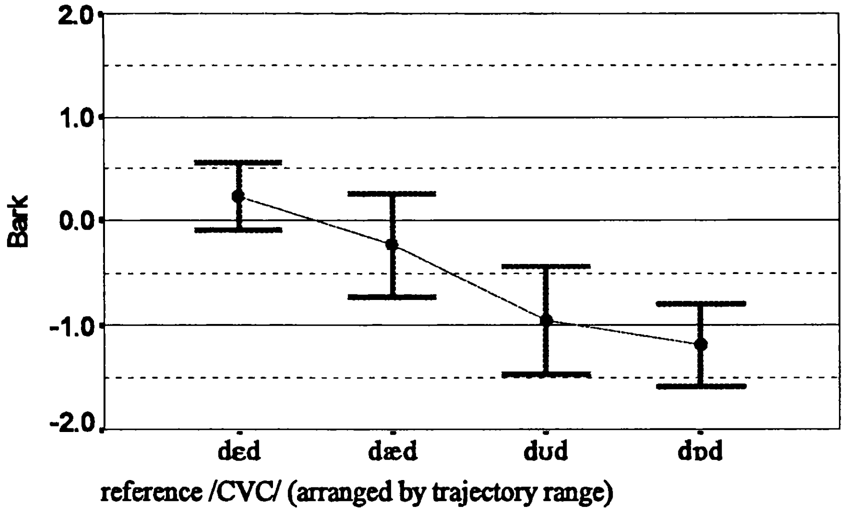


Figure 5-8: Mean shift index (circle in the middle) with an error bar of one standard deviation of /dVd/ F2. The reference /CVC/ tokens are arranged so that their trajectory range increases from left to right in X-axis.

Tables 5-12 and 5-13, together with Figures 5-5 to 5-8, show an interesting relation between the trajectory range and the matched frequency across two groups of subjects: as the trajectory range of a formant in a reference /CVC/ increases, the mean matched formant frequency shifts from within the trajectory range to outside the trajectory. Note that the F2 trajectory in /dVd/ is concave. In other words, when the formant trajectory range is small, subjects select a value somewhere between the /CVC/ edge and peak frequencies to represent its vowel quality, and when the formant trajectory range is large, they select a value beyond the trajectory range (i.e. a value higher than the peak if the trajectory is convex, and a value lower than the peak if it is concave). This is illustrated in Figure 5-9.

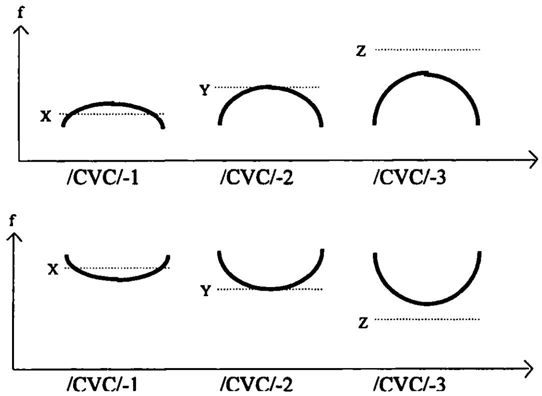


Figure 5-9: Illustration of the influence of the trajectory range on formant matching. As the formant trajectory range of /CVC/ grows larger, as seen from /CVC/-1 to /CVC/-3 above, the matched formant of test /#V#/ moves from within the trajectory range (x Hz in /CVC/-1) to outside the trajectory range (z Hz in /CVC/-3)

This observation agrees with the result of F1 matching in Pilot Experiment 2. Note that, in 20 out of 32 cases in Tables 5-12 and 5-13, their shifts from trajectory peak values are more than a difference limen, implying that they are also auditorily meaningful.

5.5.5 Discussion of the trajectory range effect

This section deals with the potential issues to be addressed concerning the effect of the trajectory range.

Initially, although it is shown that this observation generally holds in all subject groups, there is a slight but discernable difference between the two subject groups for each formant, i.e., between Groups A and B, and between Groups X and Y. We shall investigate whether there are any systematic differences.

The results of Wilcoxon tests in Tables 5-10 and 5-11 show that formant trajectory peak matching to the reference /CVC/ did not account for all the matched results, but we could observe that the peak values corresponded to the matched formant means more than the values of the other hypotheses, since there were more cases where the null hypothesis was not refuted (See Tables 5-10 and 5-11). We might therefore assume that there were two possible components to the matching process: (1) a fundamental strategy is to refer to the trajectory peak of /CVC/, and (2) an influence of the formant trajectory range. Consequently we might account for the difference between the two groups in accordance with the different weighting which these components have for the subjects.

This assumption is justified by observations obtained from Tables 5-12 and 5-13, and Figures 5-5 to 5-8: subjects in Group B showed a more prominent matching shift than those in Group A, suggesting that they were under more influence from the reference trajectory range. In /dVd/ context, subjects in Group Y were under more influence from the reference /CVC/ trajectory range than those in Group X. On the other hand, subjects in Group Y showed the peak value preference in /bVb/ context, since their /bʊb/ displayed a clear peak preference.

Obviously, this interpretation is based upon the supposition that the two formants F1 and F2 are perceptually (psychoacoustically or phonologically) processed differently within a single subject; otherwise one could not explain why two subjects adopted different matching strategies in F1 and they shared the same one in F2, generating subject crossovers in F1 and F2 groupings. For example, Subjects 1 and 7 belong to Group X for F2, but Subject 1 belongs to Group A and Subject 7 to Group B for F1.

The idea that the two formants are subject to independent perceptual processes is slightly odd, but it is shared by some previous studies. For example, DiBenedetto (1989), discussed in 3.3.4, claimed that her data can be best described by assuming that F1 and F2 were processed differently, insisting;

> The hypothesis that F1 and F2 could be perceived differently is not unreasonable, as F1 is within the lower frequency range of the auditory system in which temporal (synchronous firing rate) coding of stimulus frequency occurs, while at higher frequencies (e.g., F2 range), temporal coding breaks down and the coding of frequency is primarily spatial. (DiBenedetto 1989:76)

Moreover, Rosner & Pickering (1994:330) suggested that, in their model, although it is not specified whether the same or somewhat different type of the ASP-function "may operate on E2(t), E2(t) and E3(t)...", "[d]ifferent functions would fit the fact" that the F2 peak value of a vowel is generally somewhat more sensitive than the F1 peak value to coarticulation. These previous studies give a weak support for the hypothesis of separate F1/F2 processing.

There are some other irregularities in the general pattern that might be accounted for: first, the order of the trajectory range reverses with regard to that of the indices from /ʊ/ to /ɛ/ in /bVb/ F1 and /dVd/ F1; and second, in Group X, from /bɒb/ to /bʊb/, shift indices drop while trajectory range increases.

The first irregularity might be explained in terms of natural variability since the shift index differences between /bʊb/ and /bɛb/ and /dʊd/ and /dɛd/ are not large.

The second irregularity could be attributed to the peculiarity of /bɒb/. In /bɒb/ F2, the F2 choices of test /#V#/ are; 800, 875, 955, 1041, 1131 and 1227 Hz. Figure 5-10 shows histograms of responses for F1 results of Groups A and B, for F2 results of Group X.

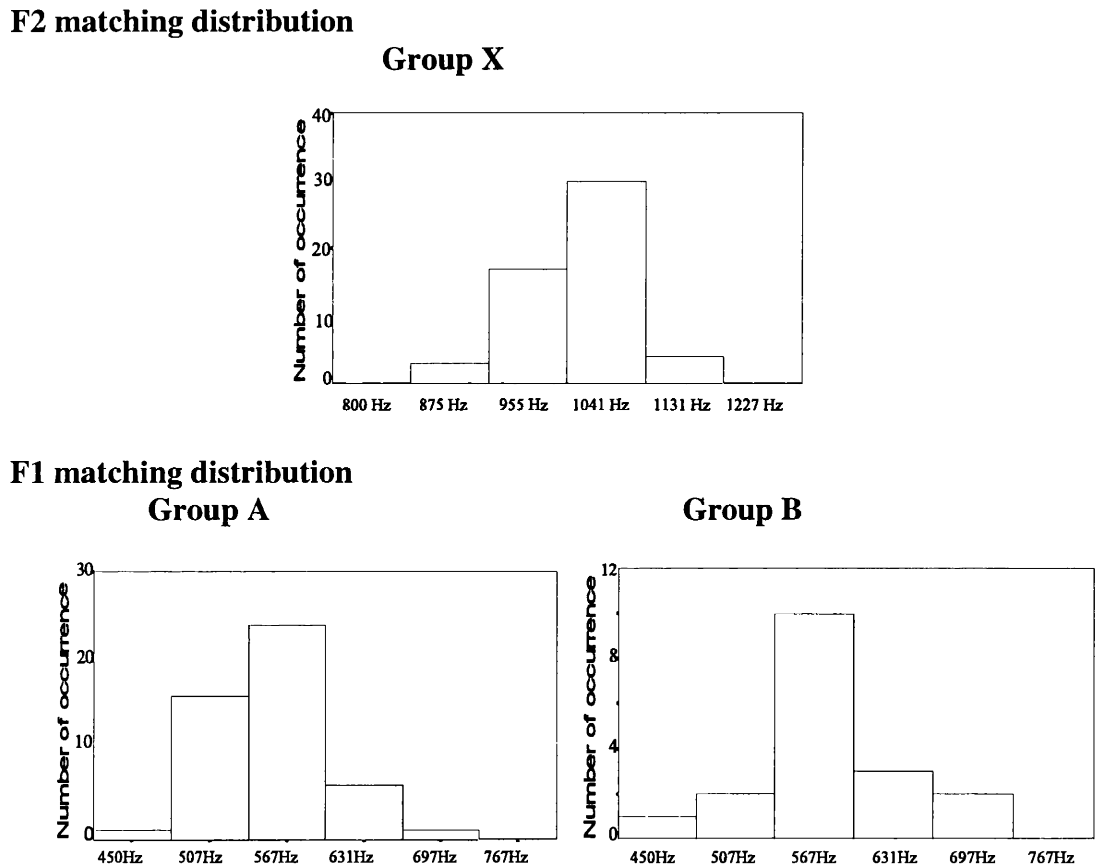


Figure 5-10: Frequency distribution of /bɒb/ F1 (Groups A and B) and F2 (Group X) responses

In the main experiment, subjects in Group X selected F2=1041 Hz most frequently while in Groups A and B, the most frequent F1 choice was 567 Hz. If the interval between two formants is considered, one possible explanation for the peculiarity of

/bɒb/ can be presented: since the Bark interval between 567 Hz and 1041 Hz is only 3.23 Bark, the two formants are very close. Furthermore, the test tokens of (F1,F2) = (567,955), which has 2.73 Bark F1-F2 distance, and (F1,F2) = (567,855), which has 2.23 Bark F1-F2 distance, show a slightly unnatural vowel quality since their two formants are closer than those of any natural vowels. These facts might account for /bɒb/ F2 matching to a higher frequency than is expected by its trajectory range.

Finally it should be mentioned that the results of the main experiment failed to show any positive evidence for the influence of low F1 values on F2 matching, which was proposed in Pilot Experiment 3 in Chapter 4. Although subjects matched the F2 of test /#V#/ tokens with frequencies lower than F2 trajectory peak of /dɒd/ and /dʊd/, this result corresponds to the prediction from the formant trajectory range, which also holds in other cases. A separate experiment is required to focus upon this particular effect of F1 trajectory peak on F2 matching.

5.6 Potential criticisms and suggestions for subsequent experiments

With the results of the main experiment discussed in the previous section, it is necessary to address a number of potential criticisms.

First of all, it was mentioned in 5.5 that there could be a response bias due to the location in the grid of the choice of matched frequency. This led to the "perseverance" criterion that excluded a subject whose responses all occurred in the centre four blocks. However, a similar but less extreme type of influence should be considered: the choice of blocks by subjects may have been drawn towards the centre of the grid. The "perseverance" criterion alone cannot exclude subjects who are influenced by this effect.

In some cases one can reject this influence: Figure 5-11, the histograms of F2 responses of Group X for /bεb/ and /dεd/, show that although the F2 steps shared the same arrangement on the grid, and the block closest to the /CVC/ trajectory peak was in the same position, (3rd from the lowest frequency; shown as "PK"), their frequency distribution patterns and their modal values were different.

/bεb/ F2 **/dεd/ F2**

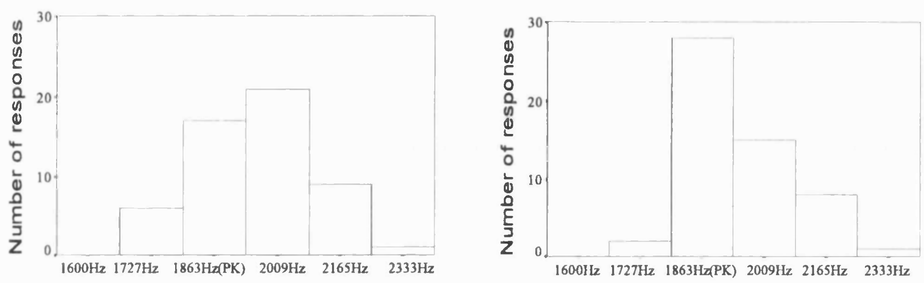


Figure 5-11: Frequency distribution of /bεb/ and /dεd/ F2 responses of Group X. The frequency closest to the /CVC/ trajectory peak is shown as "PK"

However, in other cases it might be suggested that the grid layout influenced the matched value: for example, in the actual grid used for /dʊd/ tokens (shown in Figure 5-12), the blocks with F1 frequencies lower than the /dʊd/ peak value (437 Hz) happened to be allocated in the centre 4 x 4. This could cause a serious problem for the validity of the main experiment.

250 1000	250 1087	250 1181	250 1280	250 1386	250 1499
300 1000	300 1087	300 1181	300 1280	300 1386	300 1499
352 1000	352 1087	352 1181	352 1280	352 1386	352 1499
406 1000	406 1087	406 1181	406 1280	406 1386	406 1499
461 1000	461 1087	461 1181	461 1280	461 1386	461 1499
520 1000	520 1087	520 1181	520 1280	520 1386	520 1499

Figure 5-12: Grid arrangement of /dʊd/. In each block, the number above represents the F1 value of the test /#V#/ token and the number below its F2 value in Hz. The shaded blocks have the closest F1 value to the F1 trajectory peak of /dʊd/.

In connection with the bias due to the grid location, a second criticism is the validity of the grid testing paradigm. Although it has provided some fruitful results with tasks of short duration, the important results demonstrated in grid matching should be replicated in a more conventional matching paradigm, in order to prove that the findings of this experiment are not specific to the grid matching design.

A final potential criticism concerns the trajectory shape of /CVC/, since all /CVC/ trajectories calculated according to Nearey's formula have a long central portion and a rapid transition to the consonantal locus. It was found in a survey before Pilot Experiment 1 that even a slight increase in the slope of the reference /CVC/ trajectory or even a slight decrease of the /CVC/ peak duration had a significant effect on quality, intensifying its "unnaturalness". Hence other trajectory shape types were not utilised in the main experiment. However, these different trajectory shapes might provide results which contradict the observations of the main experiment, even though their frequency range is set to the same values used in the main experiment. Furthermore, the temporal averaging hypothesis of vowel perception, discussed in 3.3.4, could not be investigated in the main experiment since due to the long central portion of Nearey's /CVC/ trajectory, the temporal average frequencies of the dynamic formant trajectory along the central 50% duration were close to the peak formant value. This hypothesis should be examined by the introduction of stimuli with different trajectory shapes.

To address these criticisms, the following three further experiments were designed:

> 1) "hi-lo" F1 Experiment: This experiment further investigates the effect of the trajectory range in the matching frequency. It also examines the influence of any visual effect, and the effect of low F1 on F2 matching. This is a grid-matching experimental scheme in which the task is a match between the reference /CVC/ and the test /#V#/. However, the stimuli comprise six different types of reference /CVC/, whose F1 peak values are different from each other by equal 0.4 Bark steps (and in consequence their F1 trajectory shapes are distinct from each other), but whose F2 peak value and F2 trajectory shape are identical; see Figure 5-13.

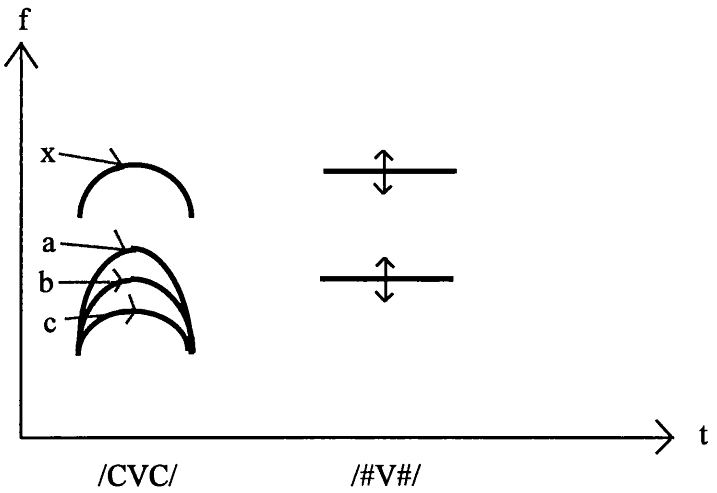


Figure 5-13: Illustration of the "hi-lo" F1 experiment. The reference /CVC/ token cannot alter its formant frequencies, and subjects can change the formant frequencies of the test /#V#/, as in the main experiment. In this experiment, however, all types of reference /CVC/ token have an identical F2 trajectory and one particular peak value (X Hz above), while they have a different F1 trajectory and a different peak value from one type to another (a/b/c Hz above).

2) "Trajectory Shape" Grid Experiment: This experiment is designed to study how the trajectory shape affects the matching strategy and whether the temporal averaging of formant trajectories is involved in vowel quality evaluation. This is a grid-matching experiment whose reference /CVC/ trajectories were obtained from three different formulae, although sharing the same peak F1/F2 values and consonantal loci.

3) "Trajectory Shape" XAXB Experiment: This experiment investigates whether a NON-grid matching experiment can produce as reliable results as a grid matching experiment if some improvement in experimental conditions is made. Here three types of trajectories are used to compare results to those of the "Trajectory Shape" Grid Experiment.

These experiments will be described and their results will be discussed in detail in

Chapter 6.

To conclude, the results of the main experiment are not consistent with the previous models of perception, but show a sensitivity to the formant trajectory range which has not been previously established. These results are however weakened by a possible response bias due to the preference for the centre of grid. Furthermore, the main experiment is subject to other criticisms such as (1) the potential sensitivity to the effect of the trajectory shape of the stimuli; and (2) the fact that such results have not been demonstrated using a similar perceptual testing paradigm. We therefore propose three further experiments to look specifically into these issues.

5.7 Concluding remarks

Overall, this main experiment was designed to investigate the strategy that listeners use to evaluate the quality of a vowel with dynamic formant trajectories, and to investigate whether their strategy corresponds with processes previously proposed in phonology or psychoacoustics. The results show that while subjects referred to the trajectory peak of /CVC/, the formant trajectory range of /CVC/ also affected their matching strategy: when the formant trajectory range was small, subjects selected a value somewhere between the /CVC/ trajectory end frequency and peak frequency to represent its vowel quality, and when the formant trajectory range was large, they selected a value beyond the trajectory range: a value higher than the peak if the trajectory was convex, and a value lower than the peak if it was concave. This is in agreement with what was observed in the F1 results of Pilot Experiment 3.

Chapter 6: Confirmation of the findings in the main experiment

---

6.1 Introduction

The previous chapter described the main experiment, in which the grid-matching test uncovered a sensitivity of subjects to formant trajectory range when they judged vowels in context: as the trajectory range becomes larger, the matched formant frequency of /#V#/ shifts from within the range of peak and /CVC/ loci to the /CVC/ peak, and sometimes beyond it. However, some potential criticisms were raised in section 5.6. In consequence, three further experiments were conducted: "Hi-lo" F1 experiment, "Trajectory Shape" grid experiment, and "Trajectory Shape" XAXB experiment. This chapter presents a description of these three experiments and integrates the results to address the criticisms provided in the previous chapter.

6.2 Experiment 1 ("hi-lo" F1 experiment: the effect of F1 on F2 matching)

6.2.1 Aims

This experiment investigates whether there is any bias due to the grid arrangement. If there is, it would cause a subject to choose one particular region (e.g. around the centre) in a grid, and this would bring about an F1 matching pattern that cannot be explained by the trajectory range hypothesis or a peak matching hypothesis. The effect of the low F1 values on F2 matching, as was observed in Pilot Experiment 3, was also examined. If this effect holds, the matching results of the F2 frequency would not be the same among six /CVC/ tokens with a different F1 value since they would be lowered when F1 values are in a low frequency range, as in /dʊd/.

6.2.2 Materials

This experiment used the same testing scheme as the main experiment, with a reference /CVC/ token and a test /#V#/ token. The reference was /dVd/, 120 ms in duration, and had a F0 declination pattern from 130 Hz to 100 Hz. As was discussed in 5.6, the F1 and F2 values of the nucleus of the reference did not use the results of the pilot acoustic analysis, unlike the other experiments, but were determined as follows. The F1 peak had six positions: 400 Hz, 444 Hz, 490 Hz, 537 Hz, 586 Hz, 637 Hz, chosen to be 0.4 Bark apart within the range of F1 between /dʊd/ (437 Hz) and /dʌd/ (658 Hz). The F2 peak of this experiment was fixed at 1350 Hz, which was intermediate between the F2 frequencies of /dʌd/ (1341 Hz) and /dʊd/ (1359 Hz) in Pilot Experiment 3. These values were chosen to verify the effect of formant trajectory range on the matched formant frequency, taking into consideration the fact that the lower matching of F1 occurred in /dʊd/ (F1=437 Hz, F2=1359 Hz) while it was not prominent in /dʌd/ (F1=658 Hz, F2=1341 Hz). F1/F2 trajectories were calculated according to Nearey's formula, as in the other experiments, with the locus value of /d/ (F1,F2) = (150,2000) Hz. The other parameter settings followed the procedure of the previous experiments. The test /#V#/ token was created in the same way as in the other experiments: 220 ms /#V#/ whose F1/F2 frequencies subjects could change by moving a cursor position in a grid. Each reference /CVC/ token type had eight repetitions, producing 6 (/CVC/ token type) x 8 (repetitions) = 48 matching sessions. The number of repetitions was increased from six per token type in the main experiment to eight in this experiment since the results of the main experiment showed that subjects could manage 48 sessions without losing concentration.

6.2.3 Subjects

Five native speakers of South East British English undertook this experiment. They were either research students or a member of the staff of the Department of Phonetics, University College London. Four of them took part in either the main experiment or Pilot Experiment 3, while one of them participated in this type of experiment for the first time. They were sufficiently motivated without financial reward. None of them had a history of any hearing defect.

6.2.4 Procedure

This Experiment was carried out in the Speech Science Laboratory of Wolfson House, Department of Phonetics, University College London, as in the main experiment, using the same DOS-SFS program and the same task. The allocation of F1/F2 on the two axes of the grid was randomised and the step of the formant change was 0.5 Bark in both F1/F2 axes, as in the main experiment. The F1/F2 values used in the grids are shown in Table 6-1.

s400

F1	250	300	352	406	461	520
F2	1000	1087	1181	1280	1386	1499

s444

F1	389	444	501	561	624	649
F2	1000	1087	1181	1280	1386	1499

s490

F1	326	378	433	490	549	612
F2	1000	1087	1181	1280	1386	1499

s537

F1	421	477	536	597	662	730
F2	1000	1087	1181	1280	1386	1499

s586

F1	465	524	585	649	716	787
F2	1000	1087	1181	1280	1386	1499

s637

F1	455	512	573	636	703	773
F2	1000	1087	1181	1280	1386	1499

Table 6-1: F1/F2 values used in the grid. All values in Hz.

6.2.5 Results and Discussion

As in the main experiment, the two subject-selection criteria were also used in this experiment: "perseverance" to exclude subjects whose matching choices were always in the centre 2 x 2 blocks in the grid; and "consistency" to eliminate subjects whose matching results were outside the range of three steps across all eight repetitions. Specifically: If one of the F1/F2 response ranges across all eight repetitions exceeded three steps in more than three /dVd/ token types out of six (c.f. 'four token types out of eight' in the main experiment) in a given subject, it was decided that the results lacked consistency and they were eliminated.

To investigate the adequacy of the results according to the perseverance criterion, the frequency of the centre-4 choices was investigated for each subject and is shown in Table 6-2. Each vertical column is for token types and each horizontal row stands for an individual subject. Similarly the F1/F2 ranges are shown in the Table 6-3. The numbers in each box of the table are (F1 range x F2 range). As in Table 5-3, if the F1 x F2 range is 3 x 3 or smaller, an open circle, "O", is inserted in the corresponding block to show that it passed the consistency criterion. In both Tables 6-2 and 6-3, s400 token stands for the token whose (peak F1 frequency, peak F2 frequency) is (400,1350). Accordingly, s444 token has (peak F1 frequency, peak F2 frequency) of (444,1350), s490 that of (490,1350), s537 that of (537,1350), s586 that of (586,1350) and s637 that of (637,1350).

	s400	s444	s490	s537	s586	s637
subj 1	2	0	0	2	3	2
subj 2	0	1	1	2	1	2
subj 3	2	2	3	2	3	5
subj 4	5	2	3	3	1	1
subj 5	1	1	2	2	1	3

Table 6-2: Number of "centre-4" choice: the number of the responses that fell within the centre 2 x 2 blocks in the 6 x 6 grid out of eight trials. The columns correspond to the token type while each row represents each subject. 'subj' stands for subject.

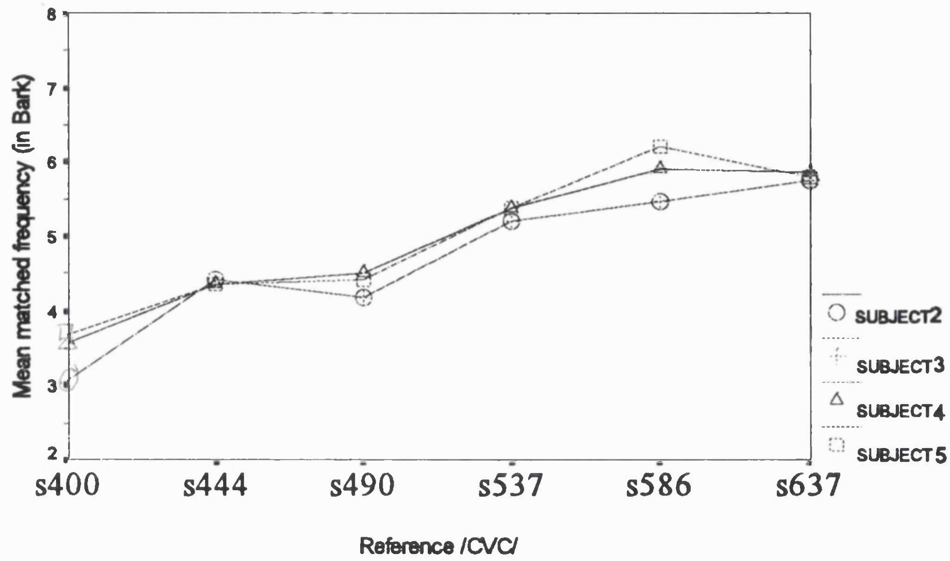
	s400	s444	s490	s537	s586	s637
subj 1	3 x 4	6 x 4	6 x 4	6 x 5	5 x 4	6 x 6
subj 2	O	O	5 x 5	5 x 4	O	4 x 5
subj 3	4 x 3	O	O	O	5 x 4	O
subj 4	O	O	4 x 5	5 x 5	O	O
subj 5	O	3 x 4	O	O	4 x 2	5 x 3

Table 6-3: Range of matched F1/F2 values of test /#V#/ token. The values indicated are (F1 step range) x (F2 step range) pooled across all eight matching sessions per subject. "O" means that all eight responses of that subject in that token were within three steps both in F1 and F2.

The result from this investigation shows that Subject 1 could not pass the second criterion "consistency" although all subjects could meet the first criterion "perseverance". Therefore the data of Subject 1 was ignored.

Then to study the general behaviour of each subject, mean matched frequencies were calculated across all eight trials for each subject, reference /CVC/ type and formant number, following the procedure used in the previous experiments, and they are plotted in Figure 6-1.

F1 mean matched frequency of each subject



F2 mean matched frequency of each subject

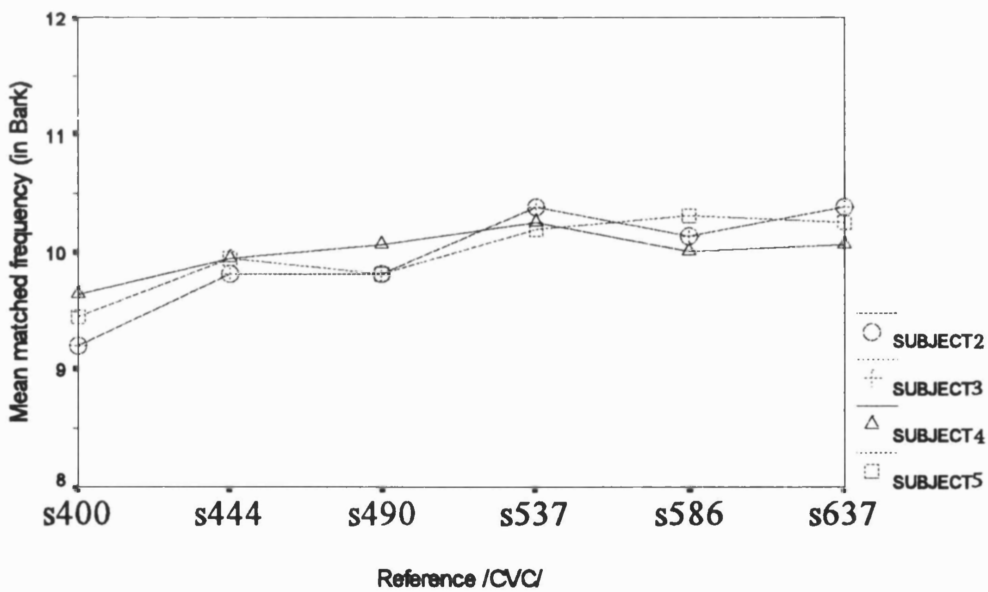


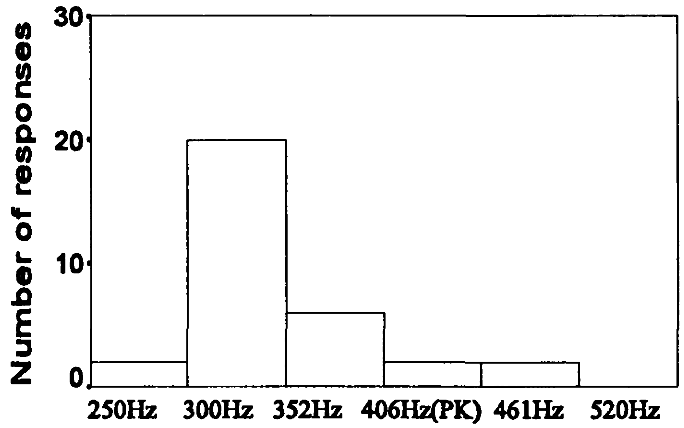
Figure 6-1: Mean matched frequency of each subject for Experiment 1.

Figures 6-1 does not clearly demonstrate whether the matching pattern of four subjects is similar or not. Subsequently, the shift index ([matched formant frequency]-[reference /CVC/ trajectory peak]) was obtained in Bark scale for each reference /CVC/ type, formant number, and trial; and Repeated Measures ANOVA on the shift index, with factors of [fl peak], [subject] and [trial], was performed to investigate the homogeneity of four subjects. The result is as follows: in F1 neither factor [subject] ($F(3,105)=1.37$, $p>.01$) nor the interaction between [subject] and [fl peak] ($F(15,105)=0.92$, $p>.01$) was significant, and also in F2 neither factor [subject] nor the interaction between [subject] and [fl peak] showed significance ($F(3,105)=1.50$, $p>.01$, $F(15,105)=1.75$, $p>.01$, respectively).

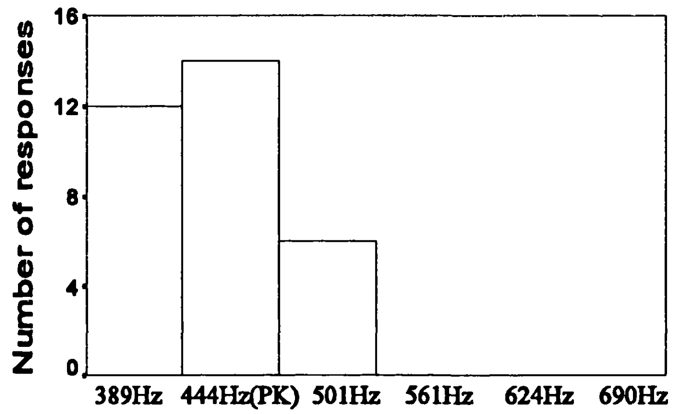
This proved that the four subjects can be treated as a homogenous group. Repeated measures ANOVA also showed that the factor [fl peak] was significant for both F1 and F2 matching ($F(5,105)=3.17$, $p<.01$, $F(5,105)=2.96$, $p<.01$, respectively), demonstrating that subjects were influenced by the F1 peak values in F1 and F2 matching process.

Consequently, across all four homogenous subjects, the mean shift index of F1 and F2 was calculated according to each token type (i.e. s400, s444...), and the results are displayed in Table 6-4 together with the trajectory range in Bark scale. Due to the nature of the stimuli, all the tokens from s400 to s637 had a different F1 trajectory peak value but an identical F2 trajectory peak value. Also across the four subjects, frequency distributions of their responses were combined for each reference /CVC/ token type according to the grid step on the F1 axis and converted to histograms, to inspect the bias of the cursor position in a grid. These histograms are shown in Figures 6-2 and 6-3.

s400



s444



s490

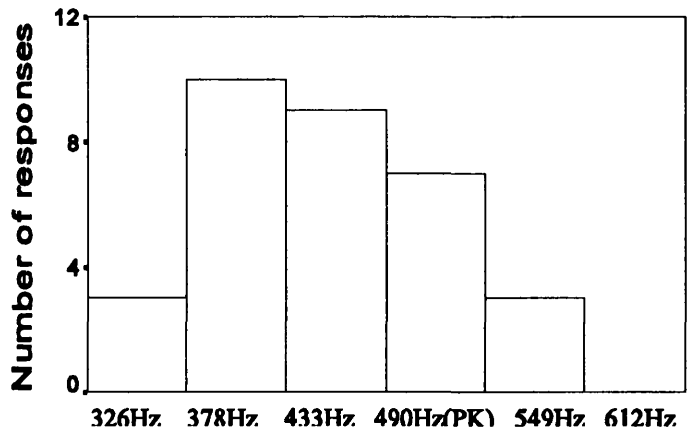
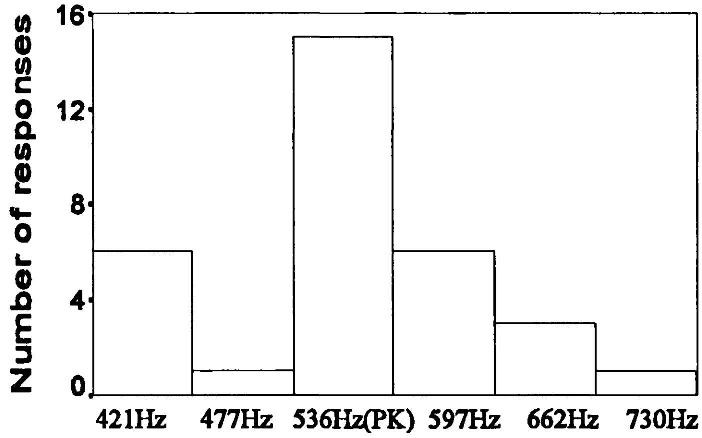
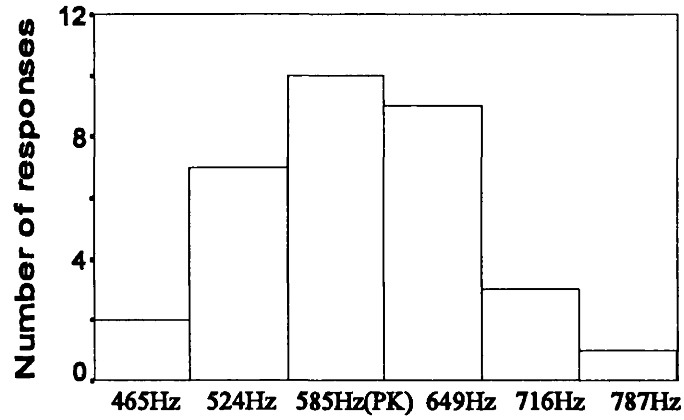


Figure 6-2: Frequency distribution of the F1 results of s400, s444 and s490 tokens. "(PK)" represents the closest frequency of /#V#/ to the F1 /CVC/ trajectory peak.

S537



S586



S637

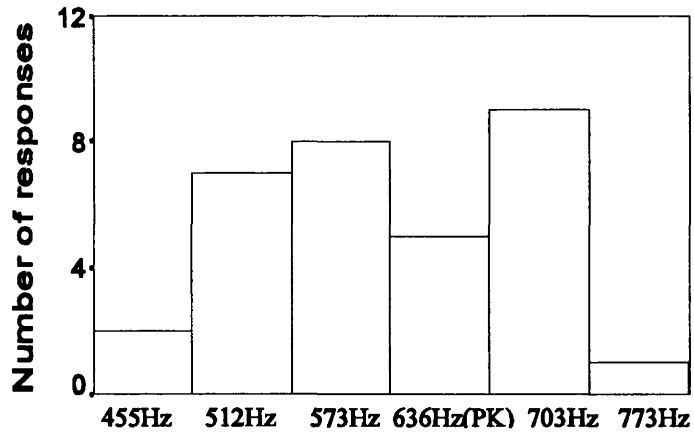


Figure 6-3: Frequency distribution of the F1 results of s537, s586 and s637 tokens. "(PK)" represents the closest frequency of /#V#/ to the F1 /CVC/ trajectory peak.

reference /CVC/ token	F1 trajectory range	Mean Shift index	F2 trajectory range	Mean Shift index
s400	2.472	-.73	2.554	-.83
s444	2.872	-.09	2.554	-.38
s490	3.272	-.55	2.554	-.36
s537	3.672	.02	2.554	-.03
s586	4.072	.09	2.554	-.19
s637	4.472	-.28	2.554	-.06

Table 6-4: Mean shift index (mean across all [matched formant]-[peak of /CVC]) and trajectory ranges of /CVC/ in Experiment 1. All numbers in Bark. The results are obtained from pooled values across 4 homogenous subjects. The label sXXX means that the peak F1/F2 of that token is (XXX,1350) Hz. e.g. s400 token had peaks (F1,F2) of (400,1350) Hz.

Table 6-4 indicates that the F1 trajectory range influenced the matching strategy of subjects in the same way as was observed in the main experiment: subjects matched an F1 of /CVC/ having a small trajectory range with an F1 frequency of /#V#/ between peak and onset/offset, while a large F1 /CVC/ trajectory range induced a matching of /#V#/ around /CVC/ trajectory peak.

As for the bias of the cursor position, the results generally show that the effect of trajectory range on the vowel quality evaluation process cannot be explained simply in terms of the bias of the visual input, since Figures 6-2 and 6-3 show that the modes of the matched distributions did not concentrate upon any particular group of grid locations. It might be argued that visual influences may have affected the matched F1 frequency of the s444 token, which was close to the F1 trajectory peak, although the grid matching for /dud/, which had an F1 peak of 437 Hz, showed an F1 matched considerably lower than the F1 /CVC/ peak in the main experiment. However, this may be ascribed to the truncated distribution of the F1 choices in the grid matching for the s444 token since there the edge value of 389 Hz was the only F1 choice lower than the F1 peak of s444, and this seems to have truncated the lower edge of the matched sample distribution, shifting the mean matched frequency to a higher value. Thus it is argued that visual bias is not a major criticism of the validity of the matching scheme.

Table 6-4 also shows the results of F2 matching. It demonstrates that the matched F2 in this experiment was lower than the F2 trajectory peak in s400, s444 and s490, while in s537, s586, s637, the matched F2 was close to the trajectory peak. The trajectory range effect predicts that the matched F2 would be lower than the F2 trajectory peak of /dVd/, and this was found in the majority of cases. However, there is a substantial variation in the chosen F2 value. This needs an explanation.

In Table 6-4, the F2 frequency of the test /#V#/ token matched to the reference /CVC/ was lower than the peak when the F1 was in a low frequency region, while it was closer to the F2 peak when the F1 peak of the /CVC/ was high in frequency. This observation confirms the effect of low F1 values on F2 matching, which was proposed in Pilot Experiment 3.

To verify the statistical validity of the observations on F1 and F2 above, Wilcoxon tests were carried out between the shift indices and their hypothetical median, 0.0, for each reference /CVC/ type and formant number, across all 4 subjects and trials. The null hypothesis was that the median shift index was 0.0 The total threshold level of significance was set to $p = .01$, and Bonferroni procedure specified the individual significance level as .01 / 2 (formants) x 6 (/CVC/ types) = .0008. Table 6-5 below shows the results.

	s400	s444	s490	s537	s586	s637
z-value of F1 shift index	-4.525	-1.241	-3.847	-.916	-.243	-2.823
z-value of F2 shift index	-4.880	-3.496	-2.804	-.691	-2.281	-1.383

Table 6-5: Results of Wilcoxon tests on the shift index. The shaded cell shows a significant difference ($p < .0008$).

The results of Wilcoxon tests show that in F1, s400 and s490, tokens with the lower F1 trajectory peak, had the null hypothesis refuted, suggesting a significant difference between the median shift index and the F1 trajectory peak, and they seem to support the effect of F1 trajectory range. They also show that in F2, the shift indices of s400 and s444 tokens were significant, endorsing the effect of low F1 values on F2 matching.

Note that the difference limen in Bark scale for F2 = 1350 Hz, based upon the 0.05 Weber fraction, is 0.31 Bark, and that this means that the F2 mean shift index of s400, s444 and s490 in Table 6-4 is auditorily discernable. Also the difference limen in Bark scale for F1 = 400 Hz is 0.18 Bark, and that for F1= 490 Hz is 0.21 Bark, both of which are smaller than their corresponding mean shift index shown in Table 6-4.

To summarise, the results of this experiment confirm that low F1 values of /CVC/ can affect the F2 matching process by lowering it from the F2 /CVC/ trajectory peak. It must be emphasised that the explanation for the mean matched F1 results of all tokens and the mean matched F2 results of /ɛ/ and /æ/ still requires the hypothesis of the trajectory range effect, and the effect of low F1 values on F2 matching does not reduce its significance, although both effects can contribute to an F2 matched lower than the trajectory peaks of the /CVC/ stimuli in this experiment. If this effect of low F1 values on F2 frequency matching holds, the results of F2 matching also contradict the possibility of a bias from a visual effect as suggested in the main experiment, since the F2 matching results differed among the six token types, while the grid location patterns of F2 were identical.

6.3 Experiment 2 ("Trajectory Shape" grid experiment)

6.3.1 Aims

Experiment 2 has two aims: (1) to enquire into how the trajectory shape affects the matching strategy in vowel quality evaluation; and (2) to study whether time averaging of a dynamic formant trajectory is involved in vowel quality evaluation.

6.3.2 Materials

The material was similar to the other grid-matching experiments, with a reference /CVC/ token and a test /#V#/ token, except that the reference /CVC/ had three types, with identical F1/F2 peak values but different formant trajectory shapes. The consonant used for the reference was /d/ and the F1/F2 values of the vowel in /CVC/ were (565,1135) Hz, adopted from the pilot acoustic analysis of /dɒd/. This particular vowel /ɒ/ was selected since its F2 provided the longest distance from its locus, 2000 Hz, giving the steepest F2 trajectory and the largest difference in a trajectory shape. The duration of the reference /dɒd/ was 120 ms, with F0 declination from 130 Hz to 100 Hz during the voicing.

In addition to the trajectory formula by Nearey, used in the other experiments, two new formulae were introduced. One was a parabola function proposed by van Son (1993), which was discussed in 3.3.6. The formula is reproduced here:

$f(t) = T - Df\{ 4(t/d)^2 - 4t/d + 1\}$ (Hz)

where: $f(t)$ = the value of a formant (F1 or F2) at time t (in Hz)
Df = T - f(on/offset) i.e. the difference between the trajectory peak and the onset/offset of /CVC/ (in Hz)
t = time, $0 < t < d$ (in ms)
d = the total token duration (in ms)
T = the nuclear (peak) formant frequency of /CVC/ (in Hz)

(van Son(1993:72))

In this experiment the trajectory peak T was F1 = 565 Hz and F2 = 1135 Hz. The f(on/offset) was the locus of /d/ in Nearey's formula, (150,2000) Hz, to keep the formant trajectory ranges identical with those created according to Nearey's formula. Duration d was 120 ms in this experiment.

The third formula used to calculate the reference /dVd/ formant trajectories was a sinusoidal function, as used by Pols, Boxelaar & Koopmans-van Beinum (1984), which was discussed in 3.4.1. Since the actual formula of the sinusoidal function which they used was not provided in their research and since they used only the quarter of the sinusoidal function to implement the one-directional formant change, this study used:

To calculate F1;
$F1(t)= i + (T-i) \sin(\theta)$
where
F1(t) = the F1 value (Hz) for a given time t (ms)
T = the nuclear (peak) formant frequency of /CVC/ (in Hz)
i = the onset/offset F1 value of /CVC/ in Hz
$\theta = 180 \times t / 120$ (in degrees)
c.f. 120 = the duration of /CVC/ in ms

To calculate F2;

$F2(t) = i - (i-T)\sin(\theta)$

where

F2(t) = the F2 value (Hz) for a given time t (ms)

T = the nuclear (peak) formant frequency of /CVC/ (in Hz)

i = the onset/offset formant value of /CVC/ (in Hz)

$\theta = 180 \times t / 120$ (in degrees)

c.f. 120 = the duration of /CVC/ in ms

Accordingly, three reference /CVC/ token types were synthesised: interpolating every 10 ms by JSRU synthesiser, with the modification of initial and final /d/ consonants for more realistic quality as before. The spectrograms of these reference /CVC/ tokens are shown in Figures 6-4 to 6-6. Figure 6-4 is the reference /CVC/ whose formant trajectories were calculated according to Nearey's formula. Henceforth this token is called Nearey /CVC/. Figure 6-5 is the spectrogram of the reference /CVC/ calculated by the formula of van Son (1993) and this token is called van Son /CVC/. Figure 6-6 is the spectrogram of /CVC/ using the sinusoidal function argued above, and this particular /CVC/ is named Sine /CVC/.

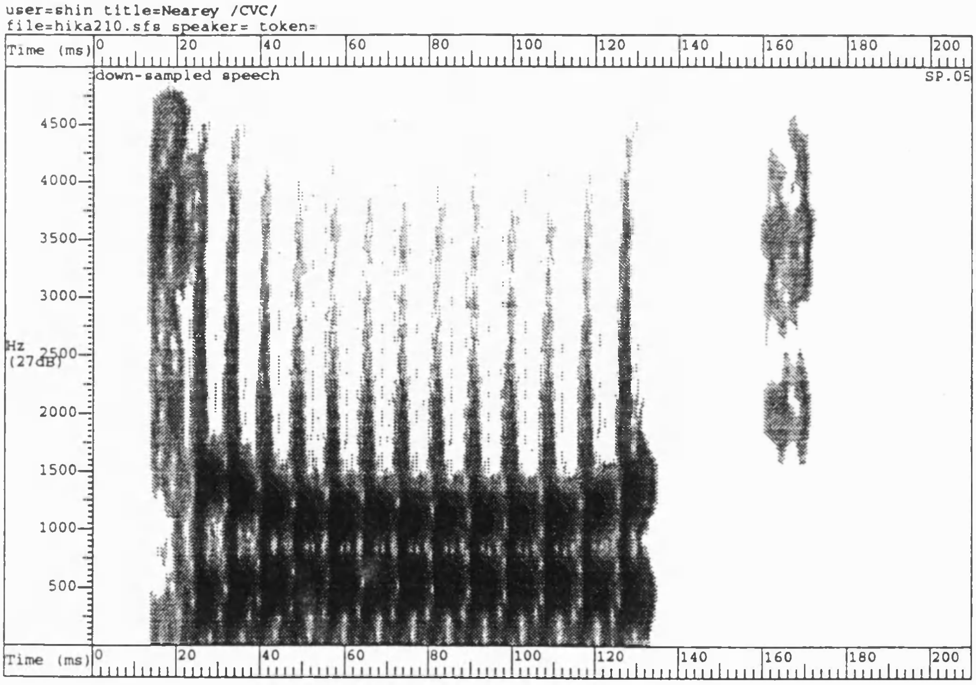


Figure 6-4: Nearey /CVC/ spectrogram

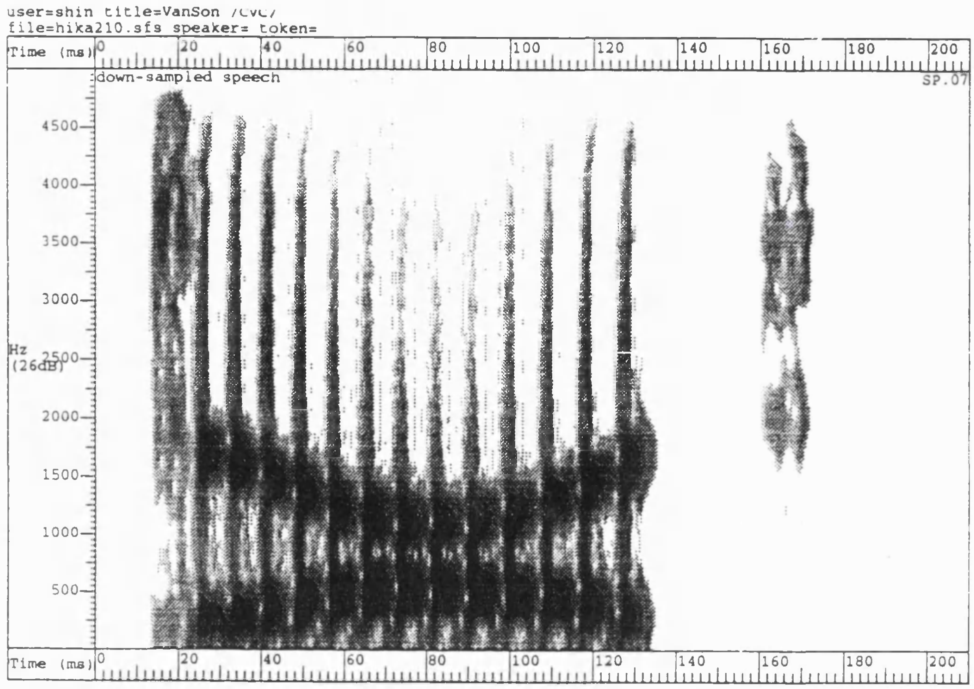


Figure 6-5: Van Son /CVC/ spectrogram

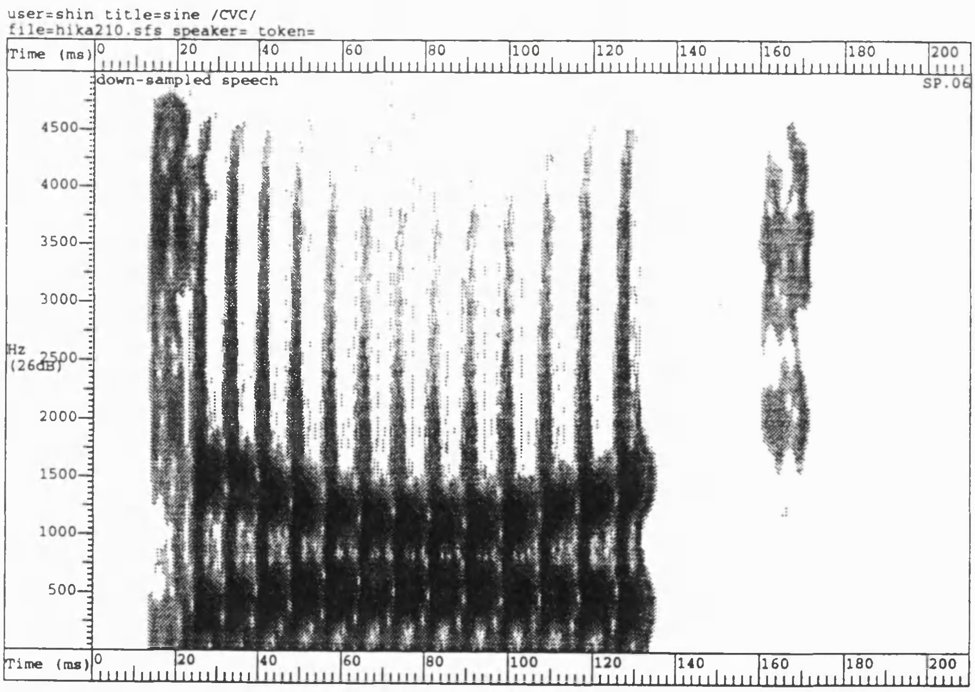


Figure 6-6: Sine /CVC/ spectrogram

Figures 6-4 to 6-6 demonstrate that the types are distinct in trajectory shape. The length of the duration of the /CVC/ nucleus is in the order: van Son /CVC/ < Sine /CVC/ < Nearey /CVC/. This is shown by the temporal averages calculated over 25 % to 75 % of the total duration of /CVC/ listed in Table 6-6.

	Nearey /CVC/	Sine /CVC/	van Son /CVC/
F1 peak	565	565	565
F1 average	564	540	518
F2 peak	1135	1135	1135
F2 average	1138	1185	1231

Table 6-6: Trajectory peak values and temporal average values of three reference /CVC/ tokens. All values in Hz.

The test /#V#/ was created as in the other experiments, with a 220 ms duration and F1/F2 frequencies which subjects could alter by moving a cursor position in the grid. Each of three reference /CVC/ types were presented eight times, creating 3 (token types) x 8 (repetitions) = 24 matching sessions per subject.

6.3.3 Subjects

Five native speakers of South East British English were tested. They were either a research student or one of the staff of the Department of Phonetics, University College London. They performed this experiment voluntarily. Three of them had participated in the main experiment. None of them participated in the previous experiment.

6.3.4 Procedure

This experiment was run as in the other grid-matching experiments. The grid numbers used for /dɒd/ were shown in Table 5-1. Full instructions were given to the two subjects who had not attended the main experiment, while for the other three who had attended the main experiment, only an oral instruction was provided.

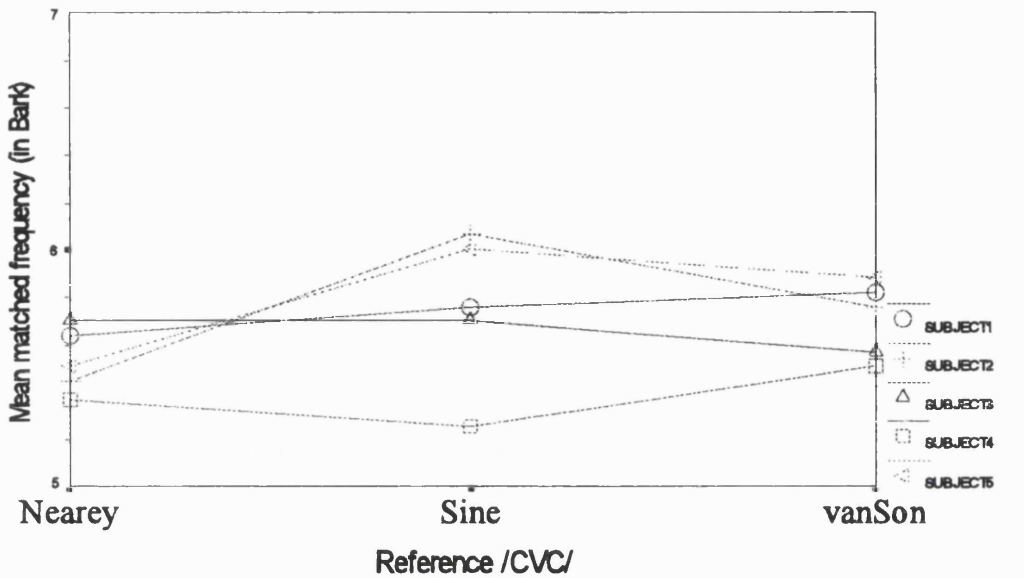
6.3.5 Results and Discussion

The two criteria for subject selection, "consistency" and "perseverance", employed in the main experiment were also used in this experiment, and the results indicated that all five subjects were able to fulfil these criteria.

To study the general behaviour of each subject, mean matched frequencies were calculated across all eight trials for each subject, reference /CVC/ type and formant, following the procedure used in the previous experiments, and they are plotted in Figure 6-7 on the next page.

Figures 6-7 does not clearly demonstrate whether the matching pattern of five subjects is similar or not. Subsequently, as in the previous experiments, the homogeneity of the five subjects was investigated by Repeated Measures ANOVA on the shift index in Bark scale obtained for each formant, reference /CVC/ type and trial, with factors of [subject], [trajectory shape] and [trial]. The significance level was set to .01.

F1 mean matched frequency of each subject



F2 mean matched frequency of each subject

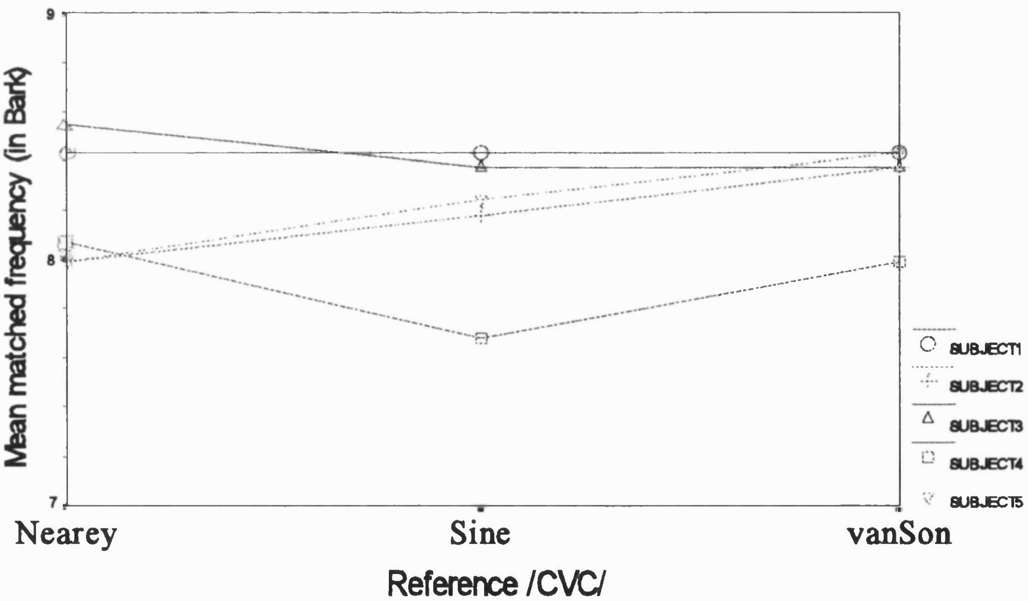


Figure 6-7: Mean matched frequency of each subject for Experiment 2.

With regard to F1, the result showed that the factor [subject] and the interaction of [subject] by [trajectory shape] did not show a significant difference ($F(4,56)=1.08$, $p>.01$, $F(8,56)=1.29$, $p>.01$, respectively). Similarly for F2 results: neither the factor [subject] ($F(4,56)=3.83$, $p>.01$) nor the interaction of [subject] by [trajectory shape] ($F(8,56)=0.71$, $p>.01$) was significant. Therefore, the five subjects were treated as a homogenous group.

The mean shift index for each /CVC/ token type across all five subjects was calculated and is displayed in Table 6-7 below.

	Nearey /CVC/	Sine /CVC/	Van Son /CVC/
F1 trajectory range	3.89	3.89	3.89
F1 mean shift index	.04	.25	.20
F2 trajectory range	-3.62	-3.62	-3.62
F2 mean shift index	-1.09	-1.08	-.95

Table 6-7: Mean shift index across five subjects with the trajectory range.
All values in Bark.

Table 6-7 shows that the difference between the three mean matched values was slight in matched F1 and F2 values, and across the three trajectory types the shift index did not show much difference. This observation was confirmed by the results of Repeated Measures ANOVA, where the factor [trajectory shape] was not significant in F1 ($F(2,56)=0.94$, $p>.01$) nor in F2 ($F(2,56)=1.83$, $p>.01$).

The involvement of the time averaging model was examined. Table 6-7 shows that across the three trajectory types, the mean matched F2 values are below the trajectory peak and the mean matched F1 values are all above the trajectory peak, although in Table 6-6, the temporal average values of three /CVC/ trajectories are below the trajectory peak in F1 and above the trajectory peak in F2. It demonstrates that the results of this experiment were not consistent with the time averaging model.

The results of F1 of all token types were proximate to the F1 peak frequency and all three token types showed lower F2 matching than the F2 trajectory peak. This confirms the results of the main experiment, and Wilcoxon tests on shift indices supported this observation. Table 6-8 demonstrates Wilcoxon's z values for each formant and each trajectory shape. The null hypothesis is that the median shift index was 0.0. Shaded cells represent that z-values there are significant, on the individual significance level of .01 / 2 (formant) x 3 (trajectory shape) = .0017. Table 6-8 displays that in F2 the null hypothesis is refuted while it is not in F1, suggesting that a significant difference exists between the matched formant median and the /CVC/ trajectory peak in F2 but not in F1.

	Nearey /CVC/	Sine /CVC/	Van Son /CVC/
Wilcoxon z for F1 shift index	-1.693	-.605	-.091
Wilcoxon z for F2 shift index	-5.430	-5.349	-5.309

Table 6-8: Wilcoxon's z values calculated on the shift index. The shaded cells represents that the z-values there are significant at the level of .0017.

To summarise, Experiment 2 aimed at establishing whether the difference in the formant trajectory shape of /CVC/ influences the strategy of subjects in vowel quality perception. The results show that the three different formant trajectory shapes did not influence the matching strategy of the subjects in this experiment. They also do not provide evidence that the temporal averaging process was involved in the vowel quality evaluation.

6.4 Experiment 3 ("Trajectory Shape" XAXB experiment)

6.4.1 Aims

In Experiment 3, the two following issues are addressed: (1) whether the important results demonstrated in the grid matching experiment can be reproduced in a more conventional matching experiment; and (2) whether the grid matching paradigm is superior to a conventional matching paradigm as a method to investigate phonetic quality of vowels.

6.4.2 Materials

In this experiment, the /CVC/ tokens were the reference tokens used in Experiment 2: three /dɒd/ token types with a different trajectory shape, Nearey /CVC/, van Son /CVC/ and Sine /CVC/. The peak formant frequencies of F1/F2 were 565 Hz/1135 Hz. There were two types of /#V#/ stimuli: one set had a fixed F1 value and variable F2 values, and the other had variable F1 values and a fixed F2 value. The F1/F2 frequencies of the /#V#/ tokens were determined in the following way, taking into consideration the matching result of the /dɒd/ token in the main experiment:

> 1) F1-fixed /#V#/, whose F1 was 565 Hz (F1 trajectory peak, the result obtained in the main experiment), and whose F2 was one of 976 Hz (1.2 Bark lower than the /CVC/ peak), 1027 Hz (0.8 Bark lower than the peak), 1080 Hz (0.4 Bark lower than the peak), 1135 Hz (the exact peak value), or 1192 Hz (0.4 Bark higher than the peak). Henceforth, they are called 976F2, 1027F2, 1080F2, 1135F2 and 1192F2, respectively. This range was produced to make an approximation to the matching results of the main experiment (1001 Hz), while allowing a choice of the F2 frequency higher than the /CVC/ peak available then (1192 Hz). This stimulus design is illustrated in Figure 6-8.

> 2) F2-fixed /#V#/, whose F2 was 1080 Hz, 0.4 Bark lower than the /CVC/ F2 peak, because of the effect of F2 trajectory on F2 lowering, and whose F1 was one of 493 Hz (0.8 Bark lower than the /CVC/ peak), 528 Hz (0.4 Bark lower than the /CVC/ peak), 585 Hz (the /CVC/ peak value), 602 Hz (0.4 Bark higher than the peak), 641 Hz (0.8 Bark higher than the peak). Henceforth, they are called 493F1, 528F1, 565F1, 602F1 and 641F1. This range was also the result of considering the peak matching tendency of F1 obtained in the main experiment. This stimulus design is shown in Figure 6-9.

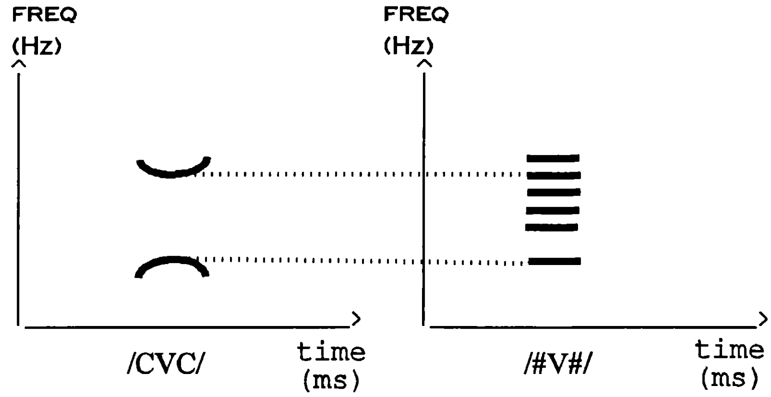


Figure 6-8: Illustration of the "F1-fixed set" used in the shape XAXB experiment. Five types of /#V#/ stimuli, shown in the spectrogram on the right, were created so that their F1 corresponds to the F1 peak of /CVC/, whose spectrogram is shown on the left, while their F2 values varies by 0.4 Bark step. The /#V#/ F2 values are, with reference to the F2 peak of /CVC/; +0.4 Bark, 0 Bark, -0.4 Bark, -0.8 Bark, and -1.2 Bark.

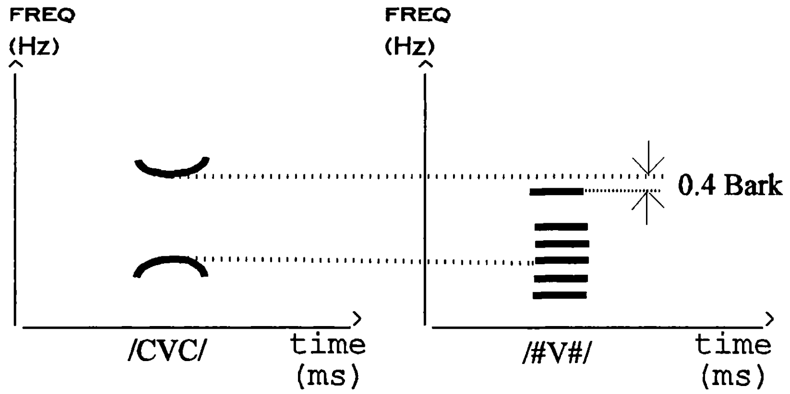


Figure 6-9: Illustration of the "F2-fixed set" used in the shape XAXB experiment. Five types of /#V#/ stimuli, shown in the spectrogram on the right, were created so that their F2, considering the result of the main experiment, is 0.4 Bark lower than the F2 peak of /CVC/, whose spectrogram is shown on the left, while their F1 varies by 0.4 Bark step. The F2 values are, with reference to the F2 peak of /CVC/: +0.8 Bark, 0.4 Bark, 0 Bark, -0.4 Bark and -0.8 Bark.

Therefore this process created, for a given trajectory shape, two sets of five /#V#/ token types, each with distinct formant settings.

6.4.3 Subjects

Six native speakers of South East British English participated in this experiment. They were either postgraduate speech scientists or postgraduate phoneticians of the Department of Phonetics, University College London.

6.4.4 Procedure

The stimulus presentation in this matching experiment was XA-XB, where X =/CVC/, and A and B are /#V#/, as in Pilot Experiment 1 discussed in 4.2. The silent interval was 120 ms between X and A or X and B, and it was 240 ms between XA and XB. The /CVC/ was one of Nearey /CVC/, Van Son /CVC/ and Sine /CVC/, while the A/B was one of the F1-fixed set or F2-fixed set /#V#/s above. The A and B in a given XA-XB were from the same set of /#V#/: there was no case where A was adopted from the F1-fixed set /#V#/ and B was adopted from the F2-fixed set /#V#/ or vice versa. One /#V#/ token type could be in positions A or B, creating two possible orders. Each XA-XB combination type was presented to a subject twice. The number of XAXB pairs presented for a given set (i.e. either of F1-fixed set or of F2-fixed set) was:

3 (/CVC/ token types that come in X) x
$_5P_2$ (the order of A/B matters) x
2 (repetitions) = 120

Therefore, the total number of stimulus XAXB pairs presented to the subject was; 120 x 2 (sets) = 240. The order of presentation of these 240 pairs was randomised. They were recorded onto a DAT tape through a 5 kHz low-pass filter.

The task of the subjects was the same as in Pilot Experiment 1. In a sound-proof recording room, they were required to listen to the set of X-A+X-B and to determine whether of the first X-A (/CVC/-/#V#/) or the second X-B (/CVC/-/#V#/) had the closer vowel quality. They were asked to tick a box corresponding the first XA pair or the second XB pair on an answer sheet. The stimuli were played through a loudspeaker at a comfortable level, and as in Pilot Experiment 1, only one subject was tested at one time. The actual session took 40 minutes, including the time of the oral instruction.

6.4.5 Results and Discussion

The results obtained were analysed in the same way as in Pilot Experiment 1 in 4.2. First for each subject, the number of the preference choices was counted with regard to each trajectory type and each token. The maximum should be 16 (2 orders x 2 repetitions x 4 combinations). The complete results are presented as tables in Appendix XI. Then to define homogenous subject groups, Model Selection Log-linear analysis was made on the data of this experiment, with factors of [trajectory shape], [F1-value] and [subject] for F2-fixed /#V#/ set, and with factors of [trajectory shape], [F2-value] and [subject] for F1-fixed /#V#/ set.

The final model created by the analysis for F2-fixed /#V#/ set had the effect of [F1-value] but not the factor [subject] nor its interaction with other factors. The final model with its predicted counts appears in Appendix X. This clearly shows that the results of the F2-fixed /#V#/ can be treated as homogenous across six subjects, and it also suggests that the trajectory shape difference did not affect the way they chose the tokens in the F2-fixed /#V#/. This similarity in response patterns among the three trajectory types is also observed in Figure 6-10, where for each trajectory type the number of responses to each token type (i.e., 493F1, 528F1...) was counted across the subjects and plotted.

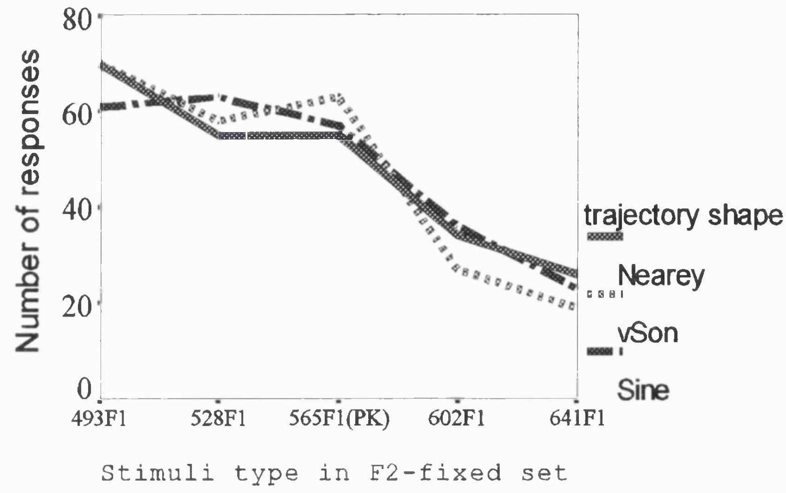


Figure 6-10: Results of F2-fixed sets across all six subjects. The frequency of the /CVC/ F1 trajectory peak is shown as "(PK)".

The Model Selection Log-linear analysis for F1-fixed /#V#/ set resulted in the creation of a final model which showed the interaction of subjects: [F2-value]*[subject]. Appendix XI shows the final model with its predicted counts. The presence of the interaction of subjects displays that responses to the F1-fixed /#V#/ set are not homogenous across six subjects and cannot be treated as a single group. Consequently the predicted counts of the final model were studied and it was discovered that of the six subjects, Subjects 2 and 3 appeared to share the similar response pattern. Then separate Model Selection Log-linear analyses were made on the group of Subjects 2 and 3, and that of Subjects 1, 4, 5 and 6. The final model for responses of Subjects 2 and 3, presented in Appendix XII, did not produce the factor [subject] as a main factor nor its interaction by [F2-value], suggesting that the responses of Subjects 2 and 3 for F1-fixed /#V#/ are homogenous. However, since the final model for responses of Subjects 1, 4, 5 and 6 displayed the interaction of the factor [subject] by [F2-value], the responses of these subjects were re-examined, and it was found that Subject 1 did not exactly show the same response pattern as that of the other three subjects. Therefore another Model Selection Log-linear analysis was carried out on Subjects 4, 5 and 6, without Subject 1, and it produced a final model without the [subject] factor: solely with the factor [F2-value]. This final model is presented in Appendix XIII.

This repeated process of Model Selection Log-linear analysis created three subject groups: (1) Subject 1; (2) Subjects 2 and 3; (3) Subjects 4, 5 and 6. It is noted that the factor [trajectory shape] is not involved in the final models of Groups (2) and (3), and this experiment did not demonstrate any significant difference between the matching patterns of Nearey /CVC/, Since /CVC/ and Van Son /CVC/, supporting the observation in Experiment 2.

For each trajectory type, the number of responses to each token type (i.e., 976F2, 1027F2 ...) was counted separately within these three groups and plotted in Figure 6-11.

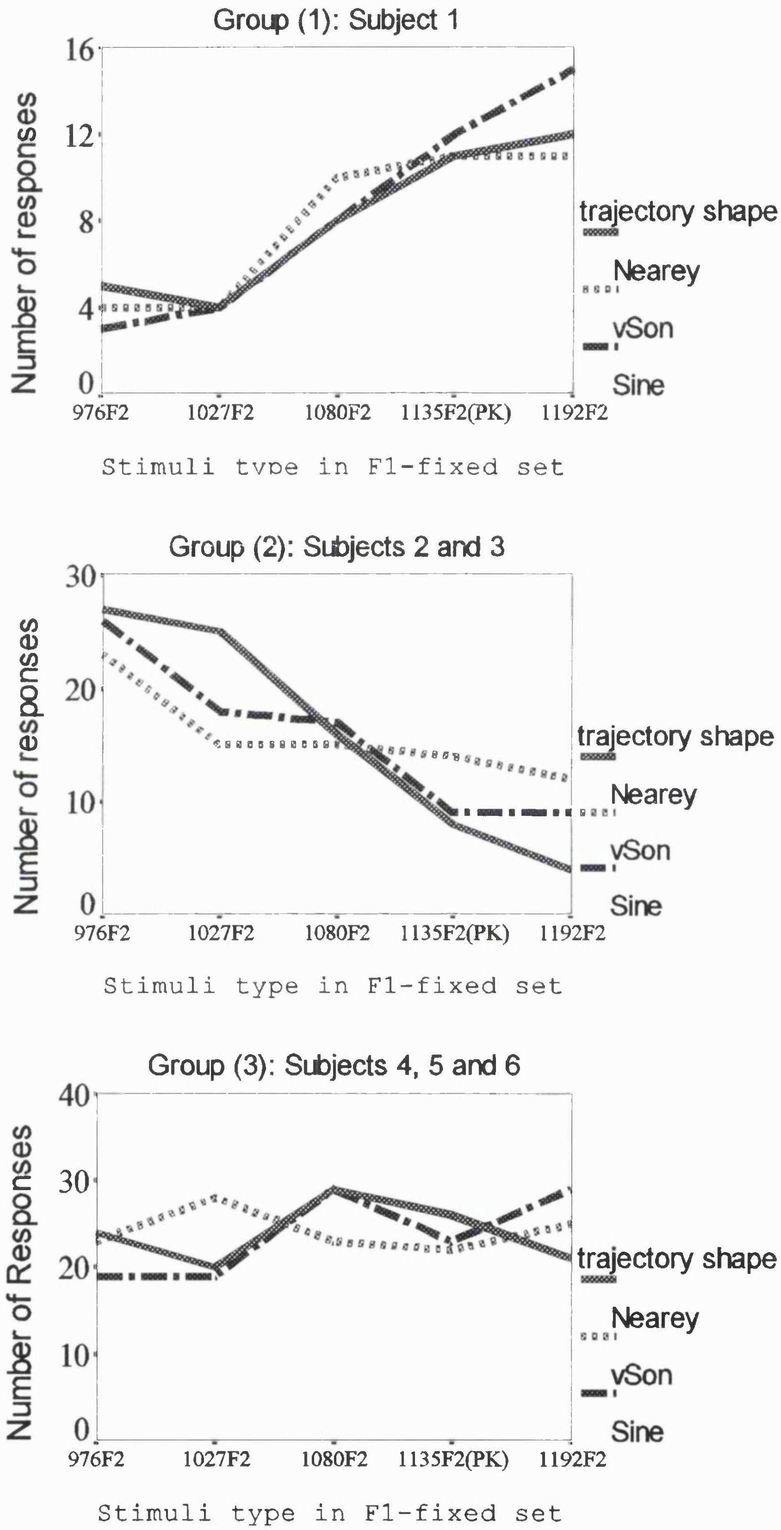


Figure 6-11: Results of F1-fixed sets according to three subject groups. The frequency of the /CVC/ F2 trajectory peak is shown as "(PK)".

It was assumed that the difference between subject groups may have been caused by the different matching strategies adopted by the subjects in each group, and the following interpretations were given to account for the diversity among the three subject groups.

Group (1): peak-matching strategy

The top graph of Figure 6-11 demonstrates that the three trajectory types did not influence the choice of Subject 1. It also shows a tendency towards a preference for the /CVC/ trajectory peak, 1135 Hz, which is similar to the behaviour of a few subjects in the main experiment, who were less sensitive to the /CVC/ trajectory range and preferred values closer to a reference /CVC/ peak.

Group (2): influenced by the trajectory range

In the middle graph of Figure 6-11, Subjects 2 and 3 show a clear preference for a lower F2 frequency. This fact and the results of their F1 matching (see Appendix VIII, the result table) are in accordance with the results of the main experiment: the grid matching for /dɒd/ in the main experiment showed the F1 matched to the trajectory peak, while it showed the F2 matched lower than the trajectory peak. This demonstrates that

- the results of the XAXB experiment do provide some support for the grid matching results.

Group (3): random response

The response pattern provided by Subjects 4, 5 and 6 in the bottom graph of Figure 6-11 is almost flat, suggesting that the three subjects could not judge the vowel quality difference and made a random response. This may be attributed to the difficult perceptual task imposed by the XAXB scheme, and confirms the superiority of the grid-matching paradigm for vowel quality evaluation. The XAXB scheme tests the short-term memory of listeners, since when listening to the second pair XB in XAXB, they have to refer to the vowel quality comparison of the XA pair in their memory to make a judgement.[25] On the other hand, the grid experiment allows them to refer to any stimulus pair whenever they wish, and to concentrate on the more finely matched pairs. Consequently, this task is naturally more manageable.

In summary, this experiment employed the XAXB matching scheme to compare its efficacy with the grid matching experiment scheme. The results show that the XAXB scheme sometimes showed the same tendencies as the grid matching scheme, but overall the subjects gave less consistent results.

6.5 Conclusion of the three experiments

The three experiments described in 6.2 to 6.4 were designed to respond to three criticisms mentioned in 5.6, which were 1) a possible 'response bias' due to the grid location; 2) the validity of the grid-matching scheme; and 3) the possible influence of the trajectory shape of the stimuli.

The first criticism, the effect of the cursor position on the choice of block, was addressed in 6.2, where the F2 grid arrangement as well as the F2 trajectory of the reference /CVC/ were fixed across all matching sessions. The results showed the clear influence of F1 on the F2 matching process, which cannot be attributed to the visual interference of the block location. They also support the effect of the /CVC/ formant trajectory range on matching processes, proposed in the main experiment.

The second criticism, the validity of the grid matching scheme, was addressed in 6.4. The XAXB matching scheme was used with the same material as Experiment 2 above, and there was evidence for the superiority of this grid matching scheme.

[25] One may claim that an introduction of a more conventional AXB paradigm instead of the XAXB paradigm used in this experiment could improve listeners' performance. However, in a pre-experimental enquiry (four subjects participated) using the AXB paradigm with the same /CVC/ and /#V#/ stimuli, two listeners remarked that they could not remember what the first A in the AXB was like when they heard the final B, and that the XAXB paradigm seemed to help their short-term memory.

The final criticism, the trajectory shape of the stimuli, was addressed in 6.3 and 6.4, where three tokens with the same peak / offset / onset formant frequencies but with different trajectory shapes were examined. The results did not present any evidence for the effect of the different trajectory shapes.

The three experiments in this chapter do not contradict the findings of Chapter 5. In vowel quality evaluation with a dynamic /CVC/, listeners follow a /CVC/ peak-picking strategy, but they are influenced by the formant trajectory range: formant frequencies of /#V#/ matched with /CVC/ shift from the peak of the /CVC/ trajectory towards its onset/offset when its trajectory range is small, while they shift towards the values beyond the trajectory range when the trajectory range is large. Some influences of low F1 frequency values on F2 matching were also discovered in these experiments. The results of Experiment 1 revealed a small influence of a visual bias, although this does not falsify the results obtained from the main experiment.

Chapter 7: General Discussion

7.1 Introduction

The experiments undertaken in this study have demonstrated three components to the perceived quality of a vowel showing formant undershoot: (i) an underlying peak-picking perceptual strategy; (ii) an effect of the formant trajectory range; and (iii) an influence of low values of F1 on F2 matching. This chapter discusses the significance of these results and returns to the objectives of this study set out in Chapter 1.

7.2 Significance of the results in the context of previous studies on phonological vowel perception

This section examines to which extent the three components described in 7.1 are supported or contradicted by the previous studies on phonological perception of vowels with undershoot.

(i) An underlying peak-picking strategy

When listeners categorise a vocalic region as a phonological vowel in /CVC/ showing formant undershoot, they cannot simply rely upon the peak frequency of the formant trajectory, since such peaks show a great acoustic variation due to coarticulatory effects with neighbouring segments. Hence the peak-picking strategy that we have demonstrated in these experiments should be attributed to a phonetic / psychoacoustic level of vowel perception.

This assumption does not cause any contradiction to the models proposed by previous works on phonological vowel perception discussed in sections 3.3.1 to 3.3.4: the perceptual overshoot hypothesis by Lindblom & Studdert-Kennedy; the vowel inherent spectral change model by Nearey; the dynamic specification model of vowel perception by Strange and her colleagues; and the temporal averaging hypothesis by DiBenedetto and Huang. One could also hypothesise that listeners phonologically categorise a vowel showing formant undershoot aided by a special psychoacoustic / phonetic process, or by an intervention of a higher perception level. Neither is the existence of an underlying peak-picking strategy a theoretical problem for the model of Rosner & Pickering (1994). Their initial model, discussed in 3.3.7, is based upon the temporal middle point of the vowel, and in their model the problem of the dynamic formant perception is solved by the introduction of a time-integrating ASP function.

Further discussion on this problem of the relationship between the phonological perception of vowels and their perceived quality is offered in Section 7.5.

(ii) A trajectory range effect

There have been a number of studies of phonological vowel perception to which we could relate the trajectory range effect. In the labelling test of Lindblom & Studdert-Kennedy (1967), used in tests of the perceptual target compensation theory described in 3.3.1, the stimuli /jVj/ had a concave F2 trajectory with trajectory ranges of 464 to 724 Hz. These values are similar to those used in the main experiment of this study, where the concave F2 trajectory of /dæd/ had a range of 1.39 Bark (=381 Hz). Our results showed that the mean matched frequency was below the trajectory peak, with its mean shift index being -0.15 Bark in Group X and -0.23 Bark in Group Y. Hence the results of the main experiment predict a perceptual overshoot in the /jVj/ results of Lindblom & Studdert-Kennedy (1967), and indeed this was the case.

The temporal averaging hypothesis of perception discussed in 3.3.4 claims that the phonological perception of a vowel with a dynamic F1 trajectory is determined by the average F1 value from 25 % to 75 % of the total /CVC/ duration. This seems to be in contradiction to the findings of this study when the trajectory range is large. However, DiBenedetto (1989), outlining a modified version of the temporal averaging hypothesis of F1, used F1 stimuli with peak values that fell within the range of American English /i/ - /ε/ continuum, which means that their F1 range was small. This demonstrates that the previous experiments conducted to test the temporal averaging hypothesis do not contradict the trajectory range effect found in this study, because both predict that listeners refer to a formant value somewhere between the /CVC/ edge and peak frequencies for small trajectories.

As discussed in Section 3.3.6, van Son (1993) used a method to evaluate phonological vowel perception by a shift in ranking of the Dutch vowels. He emphasised the importance of the closing transition of /CVC/, because the F2 shift in vowel ranking was more prominent in /#VC/ tokens than in /CVC/ or /CV#/, and claimed that the shift towards /CVC/ onset/offset was observed in F1/F2. This does not agree with the findings of this experiment, in that he did not find shifts beyond the trajectory range which were observed in this study.

The results of the experiments in this study cannot verify the two theories discussed in Chapter 3, the vowel inherent spectral change model by Nearey et al discussed in 3.3.3, and the dynamic specification model of vowel perception by Strange and her colleagues discussed in 3.3.2, since these models do not give concrete acoustic parameters for experimentation. It might be possible that the trajectory range effect is a perceptual component belonging to the psychoacoustic / phonetic level of these models, but this awaits further theoretical development.

In the auditory model of vowel perception described in 3.3.7, Rosner & Pickering (1994) point out that the ASP-function used in their model generates a running average and also "[t]emporally adjacent maxima or minima on the running averages contribute to the final definition of a point in the auditory vowel space" (p.371). This suggests that in their model the auditory process of listeners might be sensitive to the trajectory range by taking the trajectory maximum and minimum into account. In this regard the model of Rosner & Pickering is not incompatible with our results.

(iii) An interaction of low values of F1 on F2 matching

The auditory theory of vowel perception by Rosner & Pickering asserts that an ASP-function applied to one vowel formant path may affect another ASP-function and an F1-F2 interaction does not cause any theoretical setback, as they state, "there may be effects of *Ei(t)* on the output of the ASP-function for *Ej(t)*" (Rosner & Pickering 1994:371). Thus the influence of low values of F1 on F2 matching found in this study is not contradictory to their model.

No other theories than that of Rosner and Pickering, which concern phonological perception of vowels with formant undershoot, deal with this type of influence on F2 matching triggered by low F1. The temporal averaging model proposed by Huang and DiBenedetto does not incorporate the interaction of F1 and F2. Neither the vowel inherent spectral change model by Nearey, nor the dynamic specification model of vowel perception by Strange and her colleagues, is compatible with this type of perceptual interaction of two formants.

7.3 Significance of the results in the context of previous studies on phonetic / psychoacoustic vowel quality perception

In this section, the three perceptual components to the perceived vowel quality in 7.1 are discussed in light of previous phonetic studies reviewed in 3.4.

(i) An underlying peak-picking strategy

With regard to the underlying peak-picking perceptual strategy, we proposed in 7.2 that this perceptual strategy might belong to the phonetic / psychoacoustic level of vowel perception. Although it is a default assumption that listeners might refer to the formant trajectory peaks in vowel quality perception, previous work does not provide adequate discussion or investigation of this strategy, since it has investigated only a single-directional formant change, like Pols, Boxelaar & Koopmans-van Beinum (1984) and the other studies mentioned in their paper.

(ii) A trajectory range effect

The second effect, the effect of the formant trajectory range, is now discussed here. An experiment by Pols, Boxelaar & Koopmans-van Beinum (1984) discussed in 3.4.1 investigated the perceived quality of a single-direction F2 change. Their stimulus structure was considerably different from that of this study, in that their stimuli had a steady F1 trajectory and their F2 change was single-directional.

Hence it is predicted that the results of Pols, Boxelaar & Koopmans-van Beinum (1984) would not show the effect of the formant trajectory range found in our experiments, and indeed their results show that a small formant transition gave a matched frequency beyond the endpoint of the changing formant while in our results the smaller trajectory ranges caused matching between the trajectory peak of /CVC/ and the onset / offset. Thereby, we must assume that either the direction of a trajectory change or the absence of a dynamic F1 trajectory explains why no trajectory range effect was found.

(iii) An interaction of low values of F1 on F2 matching

The experiment by Pols, Boxelaar & Koopmans-van Beinum (1984) used stimuli with a single-directional F2 trajectory and a steady F1, and as in the last section, we could predict that due to the different stimulus structure, their results might not show the interaction of low values of F1 on F2 matching proposed by this study. However, their results do show a possible interaction of low values of F1 on F2 matching.

Figure 7-1 is a reproduction of Figure 4 in their paper. Its horizontal axis, Ft, is the end frequency of an F2 trajectory, which started at 1800 Hz, and its vertical axis, ΔF, is [the matched F2 value] - [the end frequency of F2].

The stimulus vowel was presented in three F1 conditions: F1=400 Hz (shown as a filled square), F1=650 Hz (a filled circle), and no F1 (a filled triangle).

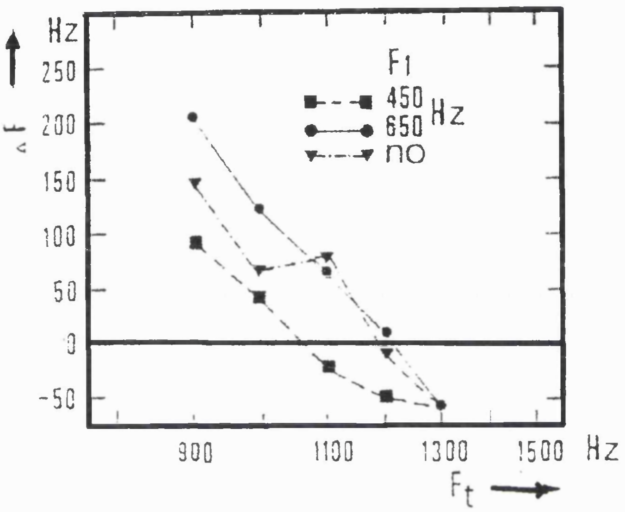


Figure 7-1: Results of F2 matching in Pols, Boxelaar & Koopmans-van Beinum (1984): A reproduction of Figure 4 (p.377)

Figure 7-1 indicates that in the case of an F2 trajectory from 1800 Hz to 1100 Hz, the matched F2 was lower than the F2 endpoint when F1=450 Hz, but it was higher when F1=650 Hz. This figure seems to show the F2 being matched lower than the F2 end / peak frequency, when F1 is low. If true, this suggests that this type of interaction between two formants is a universal psychoacoustic process regardless of the trajectory shape of F1: whether the F1 trajectory is flat or parabola, as long as it is in a low frequency region. This suggestion needs further testing for confirmation.

7.4 Interpretation of results within overall speech perception models

Sections 7.2 and 7.3 enquired into to what extent the findings of this study are supported or contradicted by the previous studies, but did not find any studies that could provide results totally compatible with the three perceptual effects discovered in this study. This section discusses three possible models that might explain how and where these three effects arise in the process of speech perception.

The first model is that all three perceptual effects are all attributed to the level of psychoacoustic perception: these effects are caused by the underlying psychoacoustic process.

As Jamieson (1989) wrote, it seems that speech sounds are subject to processing by the auditory system as any non-speech sounds: they are involved in the same obligatory auditory processes. Subsequently one could hypothesise that there is a psychoacoustic compensation process for formant undershoot that could be the origin of phonological compensation processes: phonological processes are dependent upon the 'bottom-up' influence of obligatory psychoacoustic processing.

With regard to the effect of the formant trajectory range, it is assumed that when the trajectory range is small, the output from the underlying psychoacoustic process of vowel quality evaluation provides a vowel quality judgement of the /CVC/ according to some kind of sampled or averaged value, which is a cause of the observed phonological temporal averaging process; and that when the trajectory range is large, the bottom-up output gives a judgement of vowel quality in /CVC/ by selecting a frequency outside the /CVC/ trajectory range, which is realised as phonological target compensation at the phonological level. However, it is yet to be investigated why the underlying psychoacoustic / phonetic process might show such an odd sensitivity to the dynamic trajectory range.

The underlying peak-picking strategy could also be attributed to the psychoacoustic level, and one could also assume that the interaction of low values of F1 on F2 matching may be caused by the underlying psychoacoustic process.

The second model is that in the psychoacoustic experiments of this study, some proposed phonological processes (such as perceptual target compensation or temporal averaging) influenced the listeners' matching process of vowel quality. These higher-level processes may be affecting listeners' psychoacoustic judgements of vowel quality as a 'top-down' influence from the phonological level to the psychoacoustic level. Then the trajectory range effect could result from the greater coarticulation effect on open vowels. It is also hypothesised that this top-down intervention disturbs the underlying psychoacoustic peak-picking strategy discovered in this study. We could assume that the third effect, the interaction between two formants, arises in the underlying psychoacoustic level since, as was discussed in 7.3, a similar effect was reported on by Pols, Boxelaar & Koopmans-van Beinum (1984). However, this assumption should await further investigation.

The third model is in line with the auditory theory of vowel perception, proposed by Rosner & Pickering (1994) as discussed in 3.3.7.

Rosner & Pickering argue that there are five stages I to V in their model of vowel categorisation, which are: Stage I: ERB transformation; Stage II: Filter bank; Stage III: Suppression; Stage IV: Intensity transformation; and Stage V: Second function and peak/shoulder picking. They claim that out of these five stages, Stages I to IV are open to general audition while Stage V is specific to speech perception. This model implies that the difference between the effects discovered in this study and the results of the previous work on phonological vowel perception could be ascribed to whether any auditory input has gone through Stage V that is specific to speech perception, while the similarities between them could be attributed to the outcome of Stages I to IV and hence shared by all auditory processes.

As was discussed in 7.2, this model could also provide a mechanism of how these effects might arise in this auditory model of vowel perception. A possible interaction between two auditory space functions (ASP-functions) applied to formant paths could explain the effect of low F1 values on F2 matching, and the underlying peak-picking strategy could be explained by the initial model by Rosner & Pickering which is based upon the temporal middle point of the vowel. Furthermore, the ASP-function could be sensitive to the trajectory maximum and minimum of /CVC/, suggesting that this model is not incompatible with the effect of the trajectory range.

One issue regarding this auditory model of vowel perception should be addressed here. To explain the perception of vowels with dynamic formant trajectories, Rosner & Pickering introduced an ASP-function, which integrates over the E1 / E2 values along the dynamic trajectories, and they suggested Kuwabara function as one of its best compromises with a concrete algorithm. In the main experiment, however, the Kuwabara function was not able to predict the experimental results.

However it should be noted that some other weighted averaging algorithms in auditory vowel space could explain the two modes of formant matching: when a formant trajectory in /CVC/ has a small range, listeners do not perceive the dynamic motion of the formant and judge the frequency according to some kind of sampled or average value, and when the trajectory range of /CVC/ is large enough, listeners detect the dynamicity of the /CVC/ trajectory, and judge the frequency according to some frequencies outside the /CVC/ range. This proposal of a different integrating function does not cause a theoretical problem to this model.

To summarise, three models were suggested to account for the three perceptual effects proposed by this experiment in light of the frameworks of previous studies: (1) a single psychoacoustic process causes all the perceptual effects found in this study; (2) the psychoacoustic strategy of peak-picking is under the top-down influence of the phonological perception processes, which are sensitive to a formant trajectory range; and

(3) a single auditory process of vowel perception accommodating both the psychoacoustic perception and the phonological perception, which is compatible with the model of Rosner & Pickering, causes the three perceptual effects.

7.5 Implication of the results of this study and future work

This section re-examines two issues raised in this study, discusses what theoretical implication they could offer, and proposes some possible future work.

First, one issue concerning the subject grouping pattern is discussed. In Chapter 5, we assumed that the difference of subjects' matching behaviour was due to the different matching strategies employed, and hypothesised two possible components to the matching process: (1) a fundamental strategy to refer to the trajectory peak of /CVC/ and (2) an influence of the formant trajectory range. We also proposed that the difference between two subject groups could be ascribed to how strongly subjects in a group were affected by the /CVC/ formant trajectory range. To explain the crossover patterns of subject grouping, we needed to suggest that F1 and F2 were perceived separately.

However, proceeding with the hypothesis that two formants are processed separately raises another problem to be addressed here: why some subjects were strongly influenced by the trajectory range effect although others were not, creating two different matching strategies.

Klatt (1979/1982) proposed that there is a distinction between 'psychophysical' and 'phonetic'[26] judgements of the perceptual distance of vowels. This distinction was

[26] Although we assume that the term 'phonetic' in Klatt (1982) could refer to more subtle difference than phonological, non-linguistic distinction between two speech sounds, the detailed instruction given in Klatt (1979), which was also used in the experiment of Klatt (1982) presumably, is quoted in Carlson & Granstrom (1979), which reads, "Rate only changes that tend to influence vowel identity, disregard changes associated with harshness, speaker identity, or transmission channel. " (Carlson & Granstrom 1979:86). This is vague and could still be interpreted as referring to the phonological distinction between two segments. Hence to avoid ambiguity, the term 'phonetic' is in quotation marks if it is used in Klatt's sense while it is not in a quotation when it is used in the sense of this study.

brought out by the different instructions given to the subjects. He performed an experiment on 'phonetic' judgements of vowels, and borrowed data about 'psychophysical' judgements from Carlson, Granstrom & Klatt (1979).

In the rating of 'phonetic' distance, Klatt's instruction to subjects was to ignore as best as they could "any changes that are associated with a change in speaker or recording conditions." (Klatt 1982:1278) while in that of 'psychophysical' distance, the instruction given was to "respond to the amount of difference, no matter what type of change was heard." (Carlson, Granstrom & Klatt 1979:74) The data of these two studies show that 'phonetic' judgements rest primarily on vowel formant frequencies but the 'psychoacoustic' judgements depend more on phase or spectral information of a vowel.

The instructions given in the main experiment, shown in Appendix VI, requested subjects to find the closest pair (/CVC/-/#V#/) of vowel quality, but the term 'vowel quality' may not have been specific enough to the subjects in the main experiment, although they seemed to have adequate background knowledge in phonetics and they had taken at least one ear-training course. Hence, we might assume that this ambiguity of the term 'vowel quality' could have produced some ambiguity, and consequently, different subjects may have chosen a 'psychophysical' or 'phonetic' perceptual mode.

While this might be used to explain the two subject groups in the main experiment, one important aspect remains to be investigated: whether the peak matching process and related effects are 'phonetic' or 'psychophysical'. We could devise the following possible experiments to address this problem:

(1) an experiment using the same grid-matching scheme but with trained phoneticians as subjects. This experimental design avoids the ambiguity of the term 'vowel quality' in the instruction since the term is clear to the phoneticians, and its results can be compared with those of the main experiment

(2) a set of two experiments using the grid-matching scheme, each of which provides a different instruction, 'phonetic' or 'psychophysical' as in Klatt (1982).

The instructions given in these experiments should be carefully created to prevent producing further ambiguity, and a pre-test training for tests and some subject-selection procedures should be introduced. The difference of the results of these experiments should reflect the diversity of the perceptual strategies adopted by the subjects in this study.

These should be undertaken in future work.

Another issue for future research is investigation into the three models that were suggested to account for the results of the main experiment: (1)three perceptual effects are all attributed to the level of psychoacoustic perception: these effects are caused by a single underlying psychoacoustic process; (2) the psychoacoustic strategy of peak-picking is under the top-down influence of the phonological perception processes, which are sensitive to a formant trajectory range; and (3) a single auditory process of vowel perception accommodating both the psychoacoustic perception and the phonological perception, which is compatible with the model of Rosner & Pickering, causes three perceptual effects. Although research into complete theories of speech perception is beyond the scope of this study, we explore consequences for a model of speech perception implied by each model, and suggest some future experimental designs to help choose between them.

The implication of model (1), the complete 'bottom-up' model, is that all the three perceptual components stated in 7.1 (the peak-picking, the trajectory range effect and the effect of the low F1 on F2 matching) might belong to a level of psychoacoustic sound processing, and therefore the psychoacoustic reaction to a vowel showing formant undershoot is of direct help in specifying the segment linguistically. This consequently implies that this type of acoustic variability is compensated for at a low-level. Whether this hypothesis holds or not should be examined by a controlled experiment: for example, an experiment with both speech and non-speech materials.

In contrast, model (2), the 'top-down' model, implies that although the psychoacoustic process may play some role in the linguistic categorisation of a vowel, the perceptual judgements might be influenced by the intervention from higher phonological processes. This consequently implies that the perceptual robustness of speech sounds (or at least vowels) might be the sole product of linguistic processes, either phonological template matching or more complicated processes, and the low-level psychoacoustic processes feed only primitive perceived features of the acoustic signals. This hypothesis should be tested in future experiments: considering that it suggests that phonological processes dominate the performance of the listeners, we would propose an experiment which examines vowel quality perception using two groups of subjects with a different native language and hence a different phonological vowel system.

The model (3), a single auditory process of vowel perception, implies that the general speech perception model could successfully accommodate both a process of linguistic categorisation and that of general psychoacoustic audition by attributing each process to different stages. This hypothesis could obtain some support by further experiments; for example, assuming that during the language development of children the component dealing with linguistic processing is acquired (or that the innate phonological parameters in that component are specified, as generative grammarians and phonologists claim), we might, if possible, compare the performance of adults and infants when they evaluate the quality of vowels showing formant undershoot. The performance of the infants might be assumed to be under little (or maybe no) influence of a speech mode.

7.6 Discussion of the objectives in Chapter 1

This section, following results and discussion, returns to the objectives of the thesis set out in Chapter 1.

Chapter 2 discussed the terminology of the formant shift in /CVC/ and suggested that the term undershoot is, although not completely satisfactory, the safest choice to describe the phenomenon. This meets objective 1) proposed in Chapter 1.

To meet objective 2) in Chapter 1, to discuss previous studies related to this study, Chapter 3 gave a full discussion and confirmed that the study of phonetic vowel quality evaluation had not been fully investigated and that novel experimentation was required to extend current understanding.

Objective 3) in Chapter 1 was also accomplished: to design an appropriate experimental scheme that can investigate vowel quality perception. The study adopted an interactive grid-matching scheme to minimise the number of the stimuli presented during the experiment, and to maintain the motivation of the subjects in the experimental task. This interactive grid-matching scheme was shown to be superior to the XAXB matching scheme in Chapters 4 to 6.

The experiments showed listeners' dependency upon the trajectory range of /CVC/ showing undershoot in their vowel quality perception with the additional influence of the low F1 values on F2 matching. This addresses the target objective 4) set out in Chapter 1: how listeners evaluate the quality of a vowel in a /CVC/ context with formant undershoot.

Finally the objective 5) of this study, investigation of how and to what extent the psychoacoustic processes of vowel quality evaluation contribute to the identification of phonological segments in coarticulated speech, was discussed in this chapter. Three models which were proposed in this chapter to explain the results of this study led to three hypotheses about the relationship between the phonological perception of vowels whose formants show undershoot and their perceived phonetic quality: (1) there is a low-level compensation process for formant undershoot that could be the origin of phonological compensation; (2) intervention from higher phonological processes influence the low-level processes despite that these low-level processes contribute to the linguistic categorisation of a vowel to some extent; and (3) a general speech perception model could successfully accommodate both a process of linguistic categorisation and that of general psychoacoustic audition by attributing each process to different stages.

Overall, future research on these topics will extend the contributions made by this study into how basic auditory processes are involved in phonological vowel perception.

References

Andruski, J.E. & Nearey, T.M. (1992) "On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables." *Journal of the Acoustical Society of America*, 91, 390-410

Assmann, P.F. & Summerfield, Q. (1989) "Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency." *Journal of the Acoustical Society of America,* 85, 680-697

Bergem, D.R. van. (1988) "The first step to a better understanding of vowel reduction." *Proceedings of the Institute of Phonetic Sciences Amsterdam*, 12, 61-75

Bergem, D.R. van. (1989) "Phonetic and linguistic aspects of vowel reduction." *Proceedings of the Institute of Phonetic Sciences Amsterdam,* 13, 97-105

Bergem, D.R. van. (1991a) "Acoustic and Lexical Vowel Reduction." *Proceedings of the ETRW 'Phonetics and Phonology of Speaking Styles', Barcelona, Spain,* [10-1]-[10-5]

Bergem, D.R. van. (1991b) "The influence of sentence accent, word stress, and word class on the quality of vowels" *Proceedings of Eurospeech 91, Genova, Italy, vol.3*, 1455-1458

Bergem, D.R. van (1993a) "On the perception of acoustic and lexical vowel reduction" *Proceeding of Eurospeech 93, Berlin, Germany, vol 1,* 677-680

Bergem, D.R. van (1993b) "Acoustic vowel reduction as a function of sentence accent, word stress and word class" *Speech Communication*, 12, 1-23

Bergem, D.R. van (1995) *Acoustic and lexical vowel reduction.* PhD Diss., University of Amsterdam

Bladon, R.A.W. & Lindblom, B. (1981) "Modeling the judgment of vowel quality differences." *Journal of the Acoustical Society of America*, 69, 1414-1422

Centmayer, K. (1975) "Interrelations of vowel perception and linguistic context." in G. Fant and M.A.A. Tatham. (eds.) *Auditory analysis and perception of speech.* Academic Press, 143-152

Carlson, R., Granstrom, B., & Klatt, D. (1979) "Vowel perception: the relative perceptual salience of selected acoustic manipulations." *Speech Transmission Lab. Quarterly Progress and Status Report*, 3-4, 73-83

Carlson, R., & Granstrom, B. (1979) "Model predictions of vowel discriminability." *Speech Transmission Lab. Quarterly Progress and Status Report*, 3-4, 84-104

Chistovich, L.A. & Lublinskaya, V.V. (1979a) "The 'center of gravity' effect in vowel spectra and critical distance between the formants: psychoacoustical study of the perception of vowel-like stimuli." *Hearing Research,* 1,185-195

Chistovich, L.A., Sheikin, R.L. & Lublinskaya, V.V. (1979b) " 'Centres of Gravity' and Spectral Peaks as the Determinants of Vowel Quality" in B. Lindblom & S. Ohman. (eds.) *Frontiers of Speech Communication Research.* Academic Press, 143-157

Delattre, P. (1969) "An Acoustic and Articulatory Study of Vowel Reduction in four Languages." *International Review of Applied Linguistics*, 7, 295-325

DiBenedetto, M-G (1987) *An acoustical and perceptual study on vowel height.*

PhD Diss. University of Rome

DiBenedetto, M-G. (1989) "Frequency and time variations of the first formant: Properties relevant to the perception of vowel height." *Journal of the Acoustical Society of America*, 86, 67-77

Fant, G. (1970) *Acoustic Theory of Speech Perception.* Mouton

Fourakis, M. (1991) "Tempo, stress and vowel reduction in American English." *Journal of the Acoustical Society of America*, 90, 1816-1927

Fowler, C. (1980) "Coarticulation and theories of extrinsic timing." *Journal of Phonetics*, 8, 113-133

Fox, R.A. (1989) "Dynamic Information in the Identification and Discrimination of Vowels." *Phonetica.* 46, 97-116

Fox, R.A. (1992) "Perception of Vowel Quality in Phonologically Neutralized Context." in Y. Tohkura, E. Vatikiotis-Bateson & Y. Sagisaka (eds.) *Speech perception, Production and Linguistic Structure.* IOS Press, 21-42

Gottfried, T.L, & Strange, W. (1980) "Identification of coarticulated vowels." *Journal of the Acoustical Society of America*, 68, 1626-1635

Gottfried, T.L., Miller, J.L. & Payton, P.E. (1990) "Effect of speaking rate on the perception of vowels." *Phonetica*, 47, 155-172

Hays, W.L. (1988) *Statistics.* (4th ed.) Holt, Rinehalt and Winston

Holmes, J.N. (1982) "Formant synthesisers: cascade or parallel" *JSRU research* *report.* No 1017

Huang, C.B. (1985) *Perceptual correlates of the tense / lax distinction in General American English.* Master's thesis. MIT, Cambridge, MA.

Huang, C.B. (1991) *An acoustic and perceptual study of vowel formant trajectories in American English.* Report 563, Research Laboratory of Electronics. MIT, Cambridge, MA.

Jamieson, D.G. (1987) "Studies of possible psychoacoustic factors underlying speech perception." in M.E.H. Schouten (ed.) *The Psychophysics of Speech Perception.* Marinus Nijhoff Pub., 220-230

Jenkins, J.J. (1987) "A selective history of issues in vowel perception." *Journal of memory and language*, 26, 542-549

Jenkins, J.J., Strange, W. & Edman, T.R. (1983) "Identification of vowels in "vowelless" syllables." *Perception and Psychophysics*, 34, 441-450

Jenkins, J.J., Strange, W & Miranda, S. (1994) "Vowel identification in mixed-speaker silent-center syllables." *Journal of the Acoustical Society of America,* 95, 1030-1043

Joos, M.A. (1948) "Acoustic Phonetics." *Language,* 24, (Suppl.):1-136

Keating, P.A. & Huffmann, M.K. (1984) "Vowel Variation in Japanese." *Phonetica*, 41, 191-207

Klatt, D.H. (1979) "Perceptual comparisons among a set of vowels similar to /æ/: Some differences between psychophysical distance and phonetic distance." *Journal of the*

Acoustical Society of America 66, S86

Klatt, D.H. (1982) "Prediction of perceived phonetic distance from critical-band spectra: a first step." *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing 1982*, 1278-1281

Koopmans-van Beinum, F.J. (1980) *Vowel Contrast Reduction: An acoustic and perceptual study of Dutch vowels in various speech conditions.* PhD Diss. University of Amsterdam

Kuwabara, H. (1985) "An approach to normalization of coarticulation effects for vowels in connected speech." *Journal of the Acoustical Society of America,* 77, 686-694

Kuwabara, H & Sakai, H. (1972) "Perception of vowels and C-V syllables segmented from connected speech." *Journal of the Acoustical Society of Japan,* 28, 225-234

Lindblom, B. (1963) "Spectrographic study of vowel reduction." *Journal of the Acoustical Society of America*, 35, 1773-1781

Lindblom, B. (1990a) "Explaining phonetic variation: a sketch of the H&H theory." in W.J. Hardcastle & A. Marchal. (eds.) *Speech Production and Speech Modelling.* NATO ASI Series D, vol.55, Kluwer Academic Pub. , 403-439

Lindblom, B. (1990b) "On the notion of 'possible speech sound'." *Journal of Phonetics*, 18, 135-152

Lindblom, B. & Studdert-Kennedy, M. (1967) "On the role of formant transitions in vowel recognition." *Journal of the Acoustical Society of America*, 42, 830-843

MacNeilage, P.F. (1983) *The Production of Speech.* Springer-Verlag

Miller, J.D. (1989) "Auditory-perceptual interpretation of the vowel." *Journal of the Acoustical Society of America*, 35, 2114-2134

Miller, J.L. (1981) "Effects of Speaking Rate on Segmental Distinctions." in P.D. Eimas & J.L. Miller (eds.) *Perspectives on the Studies of Speech.* Lawrence Erlbaum Associates, Pub., 39-74

Moon, S-J. (1991) *An Acoustic and Perceptual Study of Formant Undershoot in Clear and Citation-Form Speech.* PhD Diss. Univ. of Texas at Austin

Nearey, T.M. (1989) "Static, dynamic and relational factors in vowel perception." *Journal of the Acoustical Society of America*, 85, 2088-2113

Nearey, T.M. & Assmann, P.F. (1986) "Modeling the role of inherent spectral change in vowel identification." *Journal of the Acoustical Society of America,* 80, 1293-1308

Nord, L. (1986) "Acoustic Studies of Vowel Reduction in Swedish." *Speech Transmission Lab. Quarterly Progress and Status Report*, 4, 19-36

Norusis, M.J. (1994) *SPSS Advanced Statistics 6.1* (Software manual). SPSS Inc.

Peterson, G.E. & Barney, H.L. (1952) "Control methods used in a study of the vowels." *Journal of the Acoustical Society of America.* 24, 175-184

Pols, L.C.W., Boxelaar, G.W. & Koopmans-van Beinum, F.J. (1984) "Study of the Role of formant transitions in vowel recognition using the matching paradigm." *Proceedings of the Institute of Acoustics,* 6, 371-378

Rosner, B.S. & Pickering, J.B. (1994) *Vowel Perception and Production.* Oxford University Press

Son, R. van (1993) *Spectro-temporal features of vowel segments.* PhD Diss. University of Amsterdam

Stevens, K.N. (1959) "The role of duration in vowel identification." *Quarterly Progress Report, Research Lab. of Electronics, MIT*, No. 52, 136-139

Stevens, K.N. & House, A.S. (1963) "Perturbation of vowel articulations by consonantal context: an acoustical study." *Journal of Speech and Hearing Research*, 6, 111-128

Strange, W. (1987) "Information for Vowels in Formant Transitions." *Journal of Memory and Language*, 26, 550-557

Strange, W. (1989a) "Evolving theories of vowel perception." *Journal of the Acoustical Society of America*, 85, 2081-2087

Strange, W. (1989b) "Dynamic Specification of coarticulated vowels spoken in sentence context." *Journal of the Acoustical Society of America*, 85, 2135-2153

Strange, W., Jenkins, J.J. & Johnson, T.L. (1983) "Dynamic specification of coarticulated vowels. " *Journal of the Acoustical Society of America*, 74, 695-705

Studdert-Kennedy, M. (1981) "Perceiving phonetic segments." in T. Myers, J. Laver & J. Anderson (eds.) *The cognitive representation of speech.* North-Holland Pub. Co., 3-10

Syrdal, A.K. (1985) "Aspects of a model of the auditory representation of

American English vowels." *Speech Communication,* 4, 121-135

Tiffany, W.R. (1959) "Nonrandom sources of variation in vowel quality." *Journal of Speech and Hearing Speech*, 2, 305-317

Verbrugge, R.R. & Rakerd, B. (1986) "Evidence of talker-independent information for vowels" *Language and Speech*, 29, 39-57

Williams, D.R. (1987) "Judgments of coarticulated vowels are based on dynamic information." *Journal of the Acoustical Society of America,* 81, S17

Williams, D.R., Verbrugge, R.R. & Studdert-Kennedy, M (1983) "Judging sine wave stimuli as speech and nonspeech." *Journal of the Acoustical Society of America,* 74, S66

Appendices

Appendix I: Results of Pilot Experiment 1

/bVb/

	/bɛb/pk	/bɛb/tgt	/bæb/pk	/bæb/tgt	/bɒb/pk	/bɒb/tgt	/bʊb/pk	/bʊb/tgt
subject 1	8	2	3	7	7	3	4	6
subject 2	8	2	8	2	5	5	5	5
subject 3	10	0	5	5	4	6	3	7
subject 4	4	6	6	4	9	1	5	5
subject 5	8	2	6	4	6	4	6	4

/dVd/

	/dɛd/pk	/dɛd/tgt	/dæd/pk	/dæd/tgt	/dɒd/pk	/dɒd/tgt	/dʊd/pk	/dʊd/tgt
subject 1	8	2	5	5	7	3	9	1
subject 2	7	3	8	2	2	8	7	3
subject 3	8	2	5	5	4	6	4	6
subject 4	8	2	7	3	6	4	5	5
subject 5	10	0	9	1	9	1	9	1

N.B. 'pk' stands for choices of peak tokens; 'tgt' stands for choices of iso-target tokens

Appendix II: Output of Loglinear analysis on the data of Pilot Experiment 1; Across all subjects

N.B. The symbols in the output stand for following sounds: e = /ɛ/, ae =/æ/, a/o = /ɒ/, u = /ʊ/

* * * * * * * * H I E R A R C H I C A L   L O G   L I N E A R * * * * * * * *

DATA Information

80 unweighted cases accepted.
0 cases rejected because of out-of-range factor values.
0 cases rejected because of missing data.
400 weighted cases will be used in the analysis.

FACTOR Information

Factor	Level	Label
CONS	2	consonant
VOWEL	4	
SUBJECT	5	
PEAK_TGT	2	peak_iso-target_choice

- -

* * * * * * * * H I E R A R C H I C A L   L O G   L I N E A R * * * * * * * *

DESIGN 1 has generating class

CONS*VOWEL*SUBJECT*PEAK_TGT

Note: For saturated models .500 has been added to all observed cells.
This value may be changed by using the CRITERIA = DELTA subcommand.

The Iterative Proportional Fit algorithm converged at iteration 1.
The maximum difference between observed and fitted marginal totals is .000
and the convergence criterion is .250

- -

* * * * * * * * H I E R A R C H I C A L   L O G   L I N E A R * * * * * * * *

Tests that K-way and higher order effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
4	12	11.724	.4681	10.527	.5699	3
3	43	52.260	.1573	47.245	.3033	3

2	70	81.828	.1577	73.043	.3784	2
1	79	114.773	.0053	99.600	.0586	0

- -

Tests that K-way effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
1	9	32.945	.0001	26.557	.0017	0
2	27	29.568	.3339	25.798	.5298	0
3	31	40.536	.1174	36.718	.2208	0
4	12	11.724	.4681	10.527	.5699	0

******** H I E R A R C H I C A L   L O G   L I N E A R ********

Backward Elimination (p = .050) for DESIGN 1 with generating class

CONS*VOWEL*SUBJECT*PEAK_TGT

Likelihood ratio chi square = .00000 DF = 0 P = 1.000

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*VOWEL*SUBJECT*PEAK_TGT	12	11.724	.4681	3

Step 1

The best model has generating class

CONS*VOWEL*SUBJECT
CONS*VOWEL*PEAK_TGT
CONS*SUBJECT*PEAK_TGT
VOWEL*SUBJECT*PEAK_TGT

Likelihood ratio chi square = 11.72426 DF = 12 P = .468

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*VOWEL*SUBJECT	12	1.098	1.0000	3
CONS*VOWEL*PEAK_TGT	3	4.805	.1867	3
CONS*SUBJECT*PEAK_TGT	4	10.072	.0392	3
VOWEL*SUBJECT*PEAK_TGT	12	26.355	.0096	3

Step 2

The best model has generating class

CONS*VOWEL*PEAK_TGT
CONS*SUBJECT*PEAK_TGT
VOWEL*SUBJECT*PEAK_TGT

Likelihood ratio chi square = 12.82222 DF = 24 P = .969

- -

******** HIERARCHICAL LOG LINEAR ********

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*VOWEL*PEAK_TGT	3	4.148	.2459	2
CONS*SUBJECT*PEAK_TGT	4	9.489	.0500	2
VOWEL*SUBJECT*PEAK_TGT	12	25.648	.0120	3

Step 3

The best model has generating class

CONS*SUBJECT*PEAK_TGT
VOWEL*SUBJECT*PEAK_TGT
CONS*VOWEL

Likelihood ratio chi square = 16.97008 DF = 27 P = .932

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*SUBJECT*PEAK_TGT	4	9.592	.0479	3
VOWEL*SUBJECT*PEAK_TGT	12	25.744	.0117	3
CONS*VOWEL	3	.150	.9853	2

Step 4

The best model has generating class

CONS*SUBJECT*PEAK_TGT
VOWEL*SUBJECT*PEAK_TGT

Likelihood ratio chi square = 17.11984 DF = 30 P = .971

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*SUBJECT*PEAK_TGT	4	9.557	.0486	3

VOWEL*SUBJECT*PEAK_TGT 12 25.698 .0118 3

Step 5

The best model has generating class

CONS*SUBJECT*PEAK_TGT
VOWEL*SUBJECT*PEAK_TGT

Likelihood ratio chi square = 17.11984 DF = 30 P = .971

******** HIERARCHICAL LOG LINEAR ********

The final model has generating class

CONS*SUBJECT*PEAK_TGT
VOWEL*SUBJECT*PEAK_TGT

The Iterative Proportional Fit algorithm converged at iteration 0.
The maximum difference between observed and fitted marginal totals is .000
and the convergence criterion is .250

- -

Observed, Expected Frequencies and Residuals.

Factor	Code	OBS count	EXP count	Residual	Std Resid
CONS	b				
VOWEL	e				
SUBJECT	sub1				
PEAK_TGT	peak	8.0	6.9	1.10	.42
PEAK_TGT	target	2.0	2.5	-.48	-.31
SUBJECT	sub2				
PEAK_TGT	peak	8.0	7.8	.20	.07
PEAK_TGT	target	2.0	2.3	-.33	-.22
SUBJECT	sub3				
PEAK_TGT	peak	10.0	9.2	.79	.26
PEAK_TGT	target	.0	1.0	-.97	-.99
SUBJECT	sub4				
PEAK_TGT	peak	4.0	5.8	-1.76	-.73
PEAK_TGT	target	6.0	4.3	1.73	.84
SUBJECT	sub5				
PEAK_TGT	peak	8.0	7.4	.57	.21
PEAK_TGT	target	2.0	1.6	.35	.28
VOWEL	ae				
SUBJECT	sub1				
PEAK_TGT	peak	3.0	3.5	-.45	-.24
PEAK_TGT	target	7.0	7.4	-.45	-.16

SUBJECT	sub2				
PEAK_TGT	peak	8.0	8.3	-.32	-.11
PEAK_TGT	target	2.0	1.9	.13	.10
SUBJECT	sub3				
PEAK_TGT	peak	5.0	5.1	-.12	-.05
PEAK_TGT	target	5.0	4.9	.14	.06
SUBJECT	sub4				
PEAK_TGT	peak	6.0	6.2	-.24	-.10
PEAK_TGT	target	4.0	3.7	.27	.14
SUBJECT	sub5				
PEAK_TGT	peak	6.0	6.2	-.19	-.08
PEAK_TGT	target	4.0	4.1	-.12	-.06
VOWEL	a/o				
SUBJECT	sub1				
PEAK_TGT	peak	7.0	6.0	.96	.39
PEAK_TGT	target	3.0	3.7	-.72	-.38
SUBJECT	sub2				
PEAK_TGT	peak	5.0	3.6	1.36	.71
PEAK_TGT	target	5.0	6.1	-1.07	-.43
SUBJECT	sub3				
PEAK_TGT	peak	4.0	4.1	-.09	-.05
PEAK_TGT	target	6.0	5.8	.16	.07
SUBJECT	sub4				
PEAK_TGT	peak	9.0	7.2	1.80	.67
PEAK_TGT	target	1.0	2.7	-1.67	-1.02
SUBJECT	sub5				
PEAK_TGT	peak	6.0	6.2	-.19	-.08
PEAK_TGT	target	4.0	4.1	-.12	-.06
VOWEL	u				
SUBJECT	sub1				
PEAK_TGT	peak	4.0	5.6	-1.61	-.68
PEAK_TGT	target	6.0	4.3	1.66	.79
SUBJECT	sub2				
PEAK_TGT	peak	5.0	6.2	-1.24	-.50
PEAK_TGT	target	5.0	3.7	1.27	.66
SUBJECT	sub3				
PEAK_TGT	peak	3.0	3.6	-.58	-.31
PEAK_TGT	target	7.0	6.3	.68	.27
SUBJECT	sub4				
PEAK_TGT	peak	5.0	4.8	.20	.09
PEAK_TGT	target	5.0	5.3	-.33	-.14
SUBJECT	sub5				
PEAK_TGT	peak	6.0	6.2	-.19	-.08
PEAK_TGT	target	4.0	4.1	-.12	-.06
CONS	d				
VOWEL	e				
SUBJECT	sub1				
PEAK_TGT	peak	8.0	9.1	-1.10	-.36
PEAK_TGT	target	2.0	1.5	.48	.39
SUBJECT	sub2				
PEAK_TGT	peak	7.0	7.2	-.20	-.07
PEAK_TGT	target	3.0	2.7	.33	.20

SUBJECT	sub3				
PEAK_TGT	peak	8.0	8.8	-.79	-.27
PEAK_TGT	target	2.0	1.0	.97	.96
SUBJECT	sub4				
PEAK_TGT	peak	8.0	6.2	1.76	.70
PEAK_TGT	target	2.0	3.7	-1.73	-.90
SUBJECT	sub5				
PEAK_TGT	peak	10.0	10.6	-.57	-.18
PEAK_TGT	target	.0	.4	-.35	-.59
VOWEL	ae				
SUBJECT	sub1				
PEAK_TGT	peak	5.0	4.5	.45	.21
PEAK_TGT	target	5.0	4.6	.45	.21
SUBJECT	sub2				
PEAK_TGT	peak	8.0	7.7	.32	.12
PEAK_TGT	target	2.0	2.1	-.13	-.09
SUBJECT	sub3				
PEAK_TGT	peak	5.0	4.9	.12	.05
PEAK_TGT	target	5.0	5.1	-.14	-.06
SUBJECT	sub4				
PEAK_TGT	peak	7.0	6.8	.24	.09
PEAK_TGT	target	3.0	3.3	-.27	-.15
SUBJECT	sub5				
PEAK_TGT	peak	9.0	8.8	.19	.06
PEAK_TGT	target	1.0	.9	.12	.13
VOWEL	a/o				
SUBJECT	sub1				
PEAK_TGT	peak	7.0	8.0	-.96	-.34
PEAK_TGT	target	3.0	2.3	.72	.48
SUBJECT	sub2				
PEAK_TGT	peak	2.0	3.4	-1.36	-.74
PEAK_TGT	target	8.0	6.9	1.07	.41
SUBJECT	sub3				
PEAK_TGT	peak	4.0	3.9	.09	.05
PEAK_TGT	target	6.0	6.2	-.16	-.07
SUBJECT	sub4				
PEAK_TGT	peak	6.0	7.8	-1.80	-.64
PEAK_TGT	target	4.0	2.3	1.67	1.09
SUBJECT	sub5				
PEAK_TGT	peak	9.0	8.8	.19	.06
PEAK_TGT	target	1.0	.9	.12	.13
VOWEL	u				
SUBJECT	sub1				
PEAK_TGT	peak	9.0	7.4	1.61	.59
PEAK_TGT	target	1.0	2.7	-1.66	-1.02
SUBJECT	sub2				
PEAK_TGT	peak	7.0	5.8	1.24	.52
PEAK_TGT	target	3.0	4.3	-1.27	-.61
SUBJECT	sub3				
PEAK_TGT	peak	4.0	3.4	.58	.31
PEAK_TGT	target	6.0	6.7	-.68	-.26
SUBJECT	sub4				
PEAK_TGT	peak	5.0	5.2	-.20	-.09
PEAK_TGT	target	5.0	4.7	.33	.15

SUBJECT	sub5				
PEAK_TGT	peak	9.0	8.8	.19	.06
PEAK_TGT	target	1.0	.9	.12	.13

- -

Goodness-of-fit test statistics

Likelihood ratio chi square = 17.11984 DF = 30 P = .971
Pearson chi square = 15.45918 DF = 30 P = .987

- -

Appendix III: Output of Loglinear analysis on the data of Pilot Experiment 1, for a group of Subjects 1, 4 and 5

N.B. The symbols in the output stand for following sounds: e = /ɛ/, ae =/æ/, a/o = /ɒ/, u = /ʊ/

******** HIERARCHICAL LOG LINEAR ********

DATA Information

48 unweighted cases accepted.
32 cases rejected because of out-of-range factor values.
0 cases rejected because of missing data.
240 weighted cases will be used in the analysis.

FACTOR Information

Factor	Level	Label
CONS	2	
VOWEL	4	
PEAK_TGT	2	peak_iso-target_choice
SUBCAT	3	

- -

******** HIERARCHICAL LOG LINEAR ********

DESIGN 1 has generating class

CONS*VOWEL*PEAK_TGT*SUBCAT

Note: For saturated models .500 has been added to all observed cells.
This value may be changed by using the CRITERIA = DELTA subcommand.

The Iterative Proportional Fit algorithm converged at iteration 1.
The maximum difference between observed and fitted marginal totals is .000
and the convergence criterion is .250

- -

Tests that K-way and higher order effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
4	6	6.281	.3924	6.319	.3884	4
3	23	22.588	.4851	20.712	.5986	4
2	40	42.459	.3655	38.511	.5373	2
1	47	75.491	.0052	65.600	.0377	0

- -

Tests that K-way effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
1	7	33.032	.0000	27.089	.0003	0
2	17	19.871	.2809	17.798	.4017	0
3	17	16.306	.5022	14.393	.6391	0
4	6	6.281	.3924	6.319	.3884	0

* * * * * * * * HIERARCHICAL LOG LINEAR * * * * * * * *

Backward Elimination (p = .050) for DESIGN 1 with generating class

CONS*VOWEL*PEAK_TGT*SUBCAT

Likelihood ratio chi square = .00000 DF = 0 P = 1.000

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*VOWEL*PEAK_TGT*SUBCAT	6	6.281	.3924	4

Step 1

The best model has generating class

CONS*VOWEL*PEAK_TGT
CONS*VOWEL*SUBCAT
CONS*PEAK_TGT*SUBCAT
VOWEL*PEAK_TGT*SUBCAT

Likelihood ratio chi square = 6.28143 DF = 6 P = .392

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*VOWEL*PEAK_TGT	3	4.204	.2402	3
CONS*VOWEL*SUBCAT	6	.704	.9944	3
CONS*PEAK_TGT*SUBCAT	2	5.812	.0547	4

VOWEL*PEAK_TGT*SUBCAT	6	8.955	.1761	3

Step 2

The best model has generating class

CONS*VOWEL*PEAK_TGT
CONS*PEAK_TGT*SUBCAT
VOWEL*PEAK_TGT*SUBCAT

Likelihood ratio chi square = 6.98576 DF = 12 P = .859

- -

******** HIERARCHICAL LOG LINEAR ********

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*VOWEL*PEAK_TGT	3	3.771	.2873	2
CONS*PEAK_TGT*SUBCAT	2	5.589	.0612	2
VOWEL*PEAK_TGT*SUBCAT	6	8.357	.2131	2

Step 3

The best model has generating class

CONS*PEAK_TGT*SUBCAT
VOWEL*PEAK_TGT*SUBCAT
CONS*VOWEL

Likelihood ratio chi square = 10.75667 DF = 15 P = .770

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*PEAK_TGT*SUBCAT	2	4.650	.0978	3
VOWEL*PEAK_TGT*SUBCAT	6	7.270	.2966	3
CONS*VOWEL	3	.276	.9645	2

Step 4

The best model has generating class

CONS*PEAK_TGT*SUBCAT
VOWEL*PEAK_TGT*SUBCAT

Likelihood ratio chi square = 11.03278 DF = 18 P = .893

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*PEAK_TGT*SUBCAT	2	4.561	.1022	3
VOWEL*PEAK_TGT*SUBCAT	6	7.181	.3045	2

Step 5

The best model has generating class

CONS*PEAK_TGT*SUBCAT
VOWEL*PEAK_TGT
VOWEL*SUBCAT

Likelihood ratio chi square = 18.21341 DF = 24 P = .793

******** HIERARCHICAL LOG LINEAR ********

Step 5

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*PEAK_TGT*SUBCAT	2	4.561	.1022	4
VOWEL*PEAK_TGT	3	5.414	.1439	2
VOWEL*SUBCAT	6	.138	.9999	2

Step 6

The best model has generating class

CONS*PEAK_TGT*SUBCAT
VOWEL*PEAK_TGT

Likelihood ratio chi square = 18.35152 DF = 30 P = .953

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*PEAK_TGT*SUBCAT	2	4.561	.1022	4
VOWEL*PEAK_TGT	3	5.276	.1527	2

Step 7

The best model has generating class

CONS*PEAK_TGT*SUBCAT
VOWEL

Likelihood ratio chi square = 23.62766 DF = 33 P = .885

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*PEAK_TGT*SUBCAT	2	4.561	.1022	4
VOWEL	3	.000	1.0000	2

Step 8

The best model has generating class

CONS*PEAK_TGT*SUBCAT

Likelihood ratio chi square = 23.62766 DF = 36 P = .944

******** HIERARCHICAL LOG LINEAR ********

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*PEAK_TGT*SUBCAT	2	4.561	.1022	4

Step 9

The best model has generating class

CONS*PEAK_TGT
CONS*SUBCAT
PEAK_TGT*SUBCAT

Likelihood ratio chi square = 28.18908 DF = 38 P = .877

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*PEAK_TGT	1	7.976	.0047	2
CONS*SUBCAT	2	.205	.9025	2
PEAK_TGT*SUBCAT	2	6.499	.0388	2

Step 10

The best model has generating class

CONS*PEAK_TGT
PEAK_TGT*SUBCAT

Likelihood ratio chi square = 28.39418 DF = 40 P = .915

- -

******** HIERARCHICAL LOG LINEAR ********

The final model has generating class

CONS*PEAK_TGT
PEAK_TGT*SUBCAT

The Iterative Proportional Fit algorithm converged at iteration 0.
The maximum difference between observed and fitted marginal totals is .000
and the convergence criterion is .250

Observed, Expected Frequencies and Residuals.

Factor	Code	OBS count	EXP count	Residual	Std Resid
CONS	b				
VOWEL	e				
PEAK_TGT	peak				
SUBCAT	group 2	8.0	6.0	2.00	.82
SUBCAT	group 3	4.0	6.0	-2.00	-.82
SUBCAT	group 4	8.0	6.0	2.00	.82
PEAK_TGT	target				
SUBCAT	group 2	2.0	4.0	-2.00	-1.00
SUBCAT	group 3	6.0	4.0	2.00	1.00
SUBCAT	group 4	2.0	4.0	-2.00	-1.00
VOWEL	ae				
PEAK_TGT	peak				
SUBCAT	group 2	3.0	6.0	-3.00	-1.22
SUBCAT	group 3	6.0	6.0	.00	.00
SUBCAT	group 4	6.0	6.0	.00	.00
PEAK_TGT	target				
SUBCAT	group 2	7.0	4.0	3.00	1.50
SUBCAT	group 3	4.0	4.0	.00	.00
SUBCAT	group 4	4.0	4.0	.00	.00
VOWEL	a/o				
PEAK_TGT	peak				
SUBCAT	group 2	7.0	6.0	1.00	.41
SUBCAT	group 3	9.0	6.0	3.00	1.22
SUBCAT	group 4	6.0	6.0	.00	.00
PEAK_TGT	target				
SUBCAT	group 2	3.0	4.0	-1.00	-.50
SUBCAT	group 3	1.0	4.0	-3.00	-1.50
SUBCAT	group 4	4.0	4.0	.00	.00
VOWEL	u				
PEAK_TGT	peak				
SUBCAT	group 2	4.0	6.0	-2.00	-.82
SUBCAT	group 3	5.0	6.0	-1.00	-.41
SUBCAT	group 4	6.0	6.0	.00	.00

PEAK_TGT	target				
SUBCAT	group 2	6.0	4.0	2.00	1.00
SUBCAT	group 3	5.0	4.0	1.00	.50
SUBCAT	group 4	4.0	4.0	.00	.00
CONS	d				
VOWEL	e				
PEAK_TGT	peak				
SUBCAT	group 2	8.0	7.7	.33	.12
SUBCAT	group 3	8.0	7.7	.33	.12
SUBCAT	group 4	10.0	7.7	2.33	.84
PEAK_TGT	target				
SUBCAT	group 2	2.0	2.3	-.33	-.22
SUBCAT	group 3	2.0	2.3	-.33	-.22
SUBCAT	group 4	.0	2.3	-2.33	-1.53
VOWEL	ae				
PEAK_TGT	peak				
SUBCAT	group 2	5.0	7.7	-2.67	-.96
SUBCAT	group 3	7.0	7.7	-.67	-.24
SUBCAT	group 4	9.0	7.7	1.33	.48
PEAK_TGT	target				
SUBCAT	group 2	5.0	2.3	2.67	1.75
SUBCAT	group 3	3.0	2.3	.67	.44
SUBCAT	group 4	1.0	2.3	-1.33	-.87
VOWEL	a/o				
PEAK_TGT	peak				
SUBCAT	group 2	7.0	7.7	-.67	-.24
SUBCAT	group 3	6.0	7.7	-1.67	-.60
SUBCAT	group 4	9.0	7.7	1.33	.48
PEAK_TGT	target				
SUBCAT	group 2	3.0	2.3	.67	.44
SUBCAT	group 3	4.0	2.3	1.67	1.09
SUBCAT	group 4	1.0	2.3	-1.33	-.87
VOWEL	u				
PEAK_TGT	peak				
SUBCAT	group 2	9.0	7.7	1.33	.48
SUBCAT	group 3	5.0	7.7	-2.67	-.96
SUBCAT	group 4	9.0	7.7	1.33	.48
PEAK_TGT	target				
SUBCAT	group 2	1.0	2.3	-1.33	-.87
SUBCAT	group 3	5.0	2.3	2.67	1.75
SUBCAT	group 4	1.0	2.3	-1.33	-.87

- -

Goodness-of-fit test statistics

Likelihood ratio chi square = 28.39418 DF = 40 P = .915
Pearson chi square = 25.87490 DF = 40 P = .959

- -

Appendix IV: Output of Loglinear analysis on the data of Pilot Experiment 1, for a group of Subjects 2 and 3

N.B. The symbols in the output stand for following sounds: e = /ɛ/, ae =/æ/, a/o = /ɒ/, u = /ʊ/

******** HIERARCHICAL LOG LINEAR ********

DATA Information

32 unweighted cases accepted.
48 cases rejected because of out-of-range factor values.
0 cases rejected because of missing data.
160 weighted cases will be used in the analysis.

FACTOR Information

Factor	Level	Label
CONS	2	
VOWEL	4	
PEAK_TGT	2	peak_iso-target_choice
SUBJECT	2	

******** HIERARCHICAL LOG LINEAR ********

DESIGN 1 has generating class

CONS*VOWEL*PEAK_TGT*SUBJECT

Note: For saturated models .500 has been added to all observed cells.
This value may be changed by using the CRITERIA = DELTA subcommand.

The Iterative Proportional Fit algorithm converged at iteration 1.
The maximum difference between observed and fitted marginal totals is .000
and the convergence criterion is .250

******** HIERARCHICAL LOG LINEAR ********

Tests that K-way and higher order effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
4	3	2.749	.4319	2.240	.5241	3
3	13	12.946	.4520	11.090	.6032	3
2	25	35.039	.0875	30.583	.2032	2
1	31	39.283	.1460	34.000	.3251	0

- -

Tests that K-way effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
1	6	4.244	.6437	3.417	.7549	0
2	12	22.093	.0365	19.492	.0773	0
3	10	10.197	.4234	8.850	.5464	0
4	3	2.749	.4319	2.240	.5241	0

* * * * * * * * H I E R A R C H I C A L L O G L I N E A R * * * * * * * *

Backward Elimination (p = .050) for DESIGN 1 with generating class

CONS*VOWEL*PEAK_TGT*SUBJECT

Likelihood ratio chi square = .00000 DF = 0 P = 1.000

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*VOWEL*PEAK_TGT*SUBJECT	3	2.749	.4319	3

Step 1

The best model has generating class

CONS*VOWEL*PEAK_TGT
CONS*VOWEL*SUBJECT
CONS*PEAK_TGT*SUBJECT
VOWEL*PEAK_TGT*SUBJECT

Likelihood ratio chi square = 2.74912 DF = 3 P = .432

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*VOWEL*PEAK_TGT	3	3.294	.3485	3
CONS*VOWEL*SUBJECT	3	.051	.9970	3
CONS*PEAK_TGT*SUBJECT	1	.000	1.0000	3

Step 2

The best model has generating class

CONS*VOWEL*PEAK_TGT
CONS*VOWEL*SUBJECT
VOWEL*PEAK_TGT*SUBJECT

Likelihood ratio chi square = 2.74877 DF = 4 P = .601

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*VOWEL*PEAK_TGT	3	3.325	.3442	3
CONS*VOWEL*SUBJECT	3	.053	.9968	2
VOWEL*PEAK_TGT*SUBJECT	3	6.961	.0731	3

Step 3

The best model has generating class

CONS*VOWEL*PEAK_TGT
VOWEL*PEAK_TGT*SUBJECT
CONS*SUBJECT

Likelihood ratio chi square = 2.80183 DF = 7 P = .903

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*VOWEL*PEAK_TGT	3	3.284	.3499	3
VOWEL*PEAK_TGT*SUBJECT	3	6.935	.0740	3
CONS*SUBJECT	1	.077	.7815	2

Step 4

The best model has generating class

CONS*VOWEL*PEAK_TGT
VOWEL*PEAK_TGT*SUBJECT

Likelihood ratio chi square = 2.87876 DF = 8 P = .942

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
CONS*VOWEL*PEAK_TGT	3	3.209	.3605	3
VOWEL*PEAK_TGT*SUBJECT	3	6.860	.0765	3

Step 5

The best model has generating class

VOWEL*PEAK_TGT*SUBJECT
CONS*VOWEL
CONS*PEAK_TGT

Likelihood ratio chi square = 6.08775 DF = 11 P = .867

******** HIERARCHICAL LOG LINEAR ********

Step 5

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
VOWEL*PEAK_TGT*SUBJECT	3	6.860	.0765	3
CONS*VOWEL	3	.032	.9985	2
CONS*PEAK_TGT	1	.263	.6080	2

Step 6

The best model has generating class

VOWEL*PEAK_TGT*SUBJECT
CONS*PEAK_TGT

Likelihood ratio chi square = 6.11964 DF = 14 P = .963

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
VOWEL*PEAK_TGT*SUBJECT	3	6.860	.0765	3
CONS*PEAK_TGT	1	.231	.6307	2

Step 7

The best model has generating class

VOWEL*PEAK_TGT*SUBJECT
CONS

Likelihood ratio chi square = 6.35080 DF = 15 P = .973

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter

VOWEL*PEAK_TGT*SUBJECT	3	6.860	.0765	3
CONS	1	.000	1.0000	2

Step 8

The best model has generating class

VOWEL*PEAK_TGT*SUBJECT

Likelihood ratio chi square = 6.35080 DF = 16 P = .984

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
VOWEL*PEAK_TGT*SUBJECT	3	6.860	.0765	3

Step 9

The best model has generating class

VOWEL*PEAK_TGT
VOWEL*SUBJECT
PEAK_TGT*SUBJECT

Likelihood ratio chi square = 13.21108 DF = 19 P = .828

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
VOWEL*PEAK_TGT	3	20.568	.0001	2
VOWEL*SUBJECT	3	.175	.9815	2
PEAK_TGT*SUBJECT	1	1.435	.2310	2

Step 10

The best model has generating class

VOWEL*PEAK_TGT
PEAK_TGT*SUBJECT

Likelihood ratio chi square = 13.38600 DF = 22 P = .922

- -

* * * * * * * * H I E R A R C H I C A L L O G L I N E A R * * * * * * * *

The final model has generating class

VOWEL*PEAK_TGT

PEAK_TGT*SUBJECT

The Iterative Proportional Fit algorithm converged at iteration 0.
The maximum difference between observed and fitted marginal totals is .000
and the convergence criterion is .250

- -

Observed, Expected Frequencies and Residuals.

Factor	Code	OBS count	EXP count	Residual	Std Resid
CONS	b				
VOWEL	e				
PEAK_TGT	peak				
SUBJECT	sub2	8.0	8.3	-.25	-.09
SUBJECT	sub3	10.0	8.3	1.75	.61
PEAK_TGT	target				
SUBJECT	sub2	2.0	1.8	.25	.19
SUBJECT	sub3	.0	1.8	-1.75	-1.32
VOWEL	ae				
PEAK_TGT	peak				
SUBJECT	sub2	8.0	6.5	1.50	.59
SUBJECT	sub3	5.0	6.5	-1.50	-.59
PEAK_TGT	target				
SUBJECT	sub2	2.0	3.5	-1.50	-.80
SUBJECT	sub3	5.0	3.5	1.50	.80
VOWEL	a/o				
PEAK_TGT	peak				
SUBJECT	sub2	5.0	3.8	1.25	.65
SUBJECT	sub3	4.0	3.8	.25	.13
PEAK_TGT	target				
SUBJECT	sub2	5.0	6.3	-1.25	-.50
SUBJECT	sub3	6.0	6.3	-.25	-.10
VOWEL	u				
PEAK_TGT	peak				
SUBJECT	sub2	5.0	4.8	.25	.11
SUBJECT	sub3	3.0	4.8	-1.75	-.80
PEAK_TGT	target				
SUBJECT	sub2	5.0	5.3	-.25	-.11
SUBJECT	sub3	7.0	5.3	1.75	.76
CONS	d				
VOWEL	e				
PEAK_TGT	peak				
SUBJECT	sub2	7.0	8.3	-1.25	-.44
SUBJECT	sub3	8.0	8.3	-.25	-.09
PEAK_TGT	target				
SUBJECT	sub2	3.0	1.8	1.25	.94
SUBJECT	sub3	2.0	1.8	.25	.19

VOWEL	ae				
PEAK_TGT	peak				
SUBJECT	sub2	8.0	6.5	1.50	.59
SUBJECT	sub3	5.0	6.5	-1.50	-.59
PEAK_TGT	target				
SUBJECT	sub2	2.0	3.5	-1.50	-.80
SUBJECT	sub3	5.0	3.5	1.50	.80
VOWEL	a/o				
PEAK_TGT	peak				
SUBJECT	sub2	2.0	3.8	-1.75	-.90
SUBJECT	sub3	4.0	3.8	.25	.13
PEAK_TGT	target				
SUBJECT	sub2	8.0	6.3	1.75	.70
SUBJECT	sub3	6.0	6.3	-.25	-.10
VOWEL	u				
PEAK_TGT	peak				
SUBJECT	sub2	7.0	4.8	2.25	1.03
SUBJECT	sub3	4.0	4.8	-.75	-.34
PEAK_TGT	target				
SUBJECT	sub2	3.0	5.3	-2.25	-.98
SUBJECT	sub3	6.0	5.3	.75	.33

- -

Goodness-of-fit test statistics

Likelihood ratio chi square = 13.38600 DF = 22 P = .922
Pearson chi square = 11.54094 DF = 22 P = .966

- -

Appendix V: Results of Wilcoxon tests on Pilot Experiment 3

120 ms /bVb/

	bɪb	bɛb	bæb	bɒb	bʌb	bʊb
F1	-2.94	-3.36	-2.53	-1.65	-2.89	-3.46
Prob.	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002
F2	-1.70	-1.65	-0.41	-0.25	-2.06	-1.80
Prob.	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002

120 ms /dVd/

	dɪd	dɛd	dæd	dɒd	dʌd	dʊd
F1	-2.58	-3.51	-3.36	-2.06	-3.15	-2.84
Prob.	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002
F2	-1.44	-2.79	-1.70	-3.20	-0.67	-3.51
Prob.	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002

220 ms /bVb/

	bɪb	bɛb	bæb	bɒb	bʌb	bʊb
F1	-3.51	-3.51	-2.74	-1.49	-1.96	-3.51
Prob.	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002
F2	-1.96	-3.20	-0.98	-1.08	-2.06	-1.18
Prob.	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002

220 ms /dVd/

	dɪd	dɛd	dæd	dɒd	dʌd	dʊd
F1	-3.51	-3.51	-1.80	-1.24	-3.10	-3.51
Prob.	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002
F2	-0.72	-2.74	-0.62	-1.24	-0.25	-3.52
Prob.	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002	p>.0002

N.B. Each cell displays the z-value. Prob stands for probability. The significance level of each test, .0002, is determined by Bonferroni procedure.

Appendix VI: Written instruction given to subjects in the main experiment

---

Instruction

Thank you very much for attending this experiment.

What we would ask you to do is the following thing:

< Find the closest pair of vowel quality!>

You can see a 6 x 6 grid on the PC screen.
Clicking each block by a mouse, you can listen to a pair of **/CVC/ and /#V#/** on that block
The first token of each pair is the same in all block, but the second /#V#/ changes its quality block by block.
Your task is to find out a block,
where the vowel quality of **the first token and second token** is closest by playing around the grid.

When you think you find a block of a closest sound pair, hit the [space] key, and the next trial grid appears.

This experiment has one session,
with 48 stimuli.

Appendix VII: Results of the cluster analysis made on the mean shift index of 11 subjects

N.B. "sb" stands for subject

<F1 results>

Squared Euclidean Dissimilarity Coefficient Matrix

Case	sb1	sb2	sb3	sb4	sb5
sb2	.4871				
sb3	.6346	.7577			
sb4	.7085	.9810	1.5063		
sb5	.7196	1.4455	.4424	1.7479	
sb6	1.6248	1.7723	1.2948	1.4443	1.5576
sb7	2.2835	2.5210	1.2585	4.3028	1.0385
sb9	.8076	1.6795	1.7500	1.9753	1.4796
sb10	.2598	.7051	.5688	.3893	.7594
sb11	.3079	.7920	.6957	.5996	.7035
sb13	.3670	.6345	.7882	.3465	1.2882

Case	sb6	sb7	sb9	sb10	sb11
sb7	3.1621				
sb9	3.0740	3.0353			
sb10	1.2818	2.8637	1.3092		
sb11	.7787	2.0740	1.3195	.4359	
sb13	.9278	3.3493	1.4742	.1812	.3687

Cluster Membership of Cases using Average Linkage (Between Groups)

		Number of Clusters				
Label	Case	6	5	4	3	2
sb1	1	1	1	1	1	1
sb2	2	2	1	1	1	1
sb3	3	3	2	1	1	1
sb4	4	1	1	1	1	1
sb5	5	3	2	1	1	1
sb6	6	4	3	2	1	1
sb7	7	5	4	3	2	2
sb9	8	6	5	4	3	1
sb10	9	1	1	1	1	1
sb11	10	1	1	1	1	1
sb13	11	1	1	1	1	1

<F2 results>

Squared Euclidean Dissimilarity Coefficient Matrix

Case	sb1	sb2	sb3	sb4	sb5
sb2	1.1973				
sb3	.6312	1.3191			
sb4	.6418	2.4707	.7840		
sb5	1.1368	1.4107	.4090	1.2334	
sb6	.4608	1.3273	.2964	.7042	.8098
sb7	.7323	1.7090	.5257	.8635	.3291
sb9	1.0901	2.9846	1.7073	2.1341	1.9113
sb10	1.1223	1.1862	1.1735	2.2219	1.1785
sb11	.9629	2.0506	.8325	1.5879	.6933
sb13	.7309	2.5226	.8027	1.1087	.7615

Case	sb6	sb7	sb9	sb10	sb11
sb7	.9353				
sb9	1.7987	1.6184			
sb10	.8875	1.2890	1.6224		
sb11	.9537	.3592	1.4382	.7192	
sb13	.9567	.3332	1.0916	1.4554	.2774

- -

Cluster Membership of Cases using Average Linkage (Between Groups)

		Number of Clusters				
Label	Case	6	5	4	3	2
sb1	1	1	1	1	1	1
sb2	2	2	2	2	2	1
sb3	3	1	1	1	1	1
sb4	4	3	1	1	1	1
sb5	5	4	3	1	1	1
sb6	6	1	1	1	1	1
sb7	7	4	3	1	1	1
sb9	8	5	4	3	3	2
sb10	9	6	5	4	2	1
sb11	10	4	3	1	1	1
sb13	11	4	3	1	1	1

Appendix VIII: .05 Weber fraction of each formant trajectory peak value in Bark scale

/bVb/

	bɛb	bæb	bɒb	bʊb
F1	0.22	0.26	0.22	0.21
F2	0.32	0.31	0.28	0.29

/dVd/

	dɛd	dæd	dɒd	dʊd
F1	0.21	0.25	0.22	0.19
F2	0.32	0.31	0.29	0.31

Appendix IX: Result tables of the supplementary XAXB Experiment

<Results of F2-fixed sets: all numbers are a count of the choices>

Nearey /CVC/

	493F1	528F1	565F1	602F1	641F1
subject1	12	10	7	8	3
subject2	6	9	13	7	5
subject3	10	8	10	5	7
subject4	14	12	9	3	2
subject5	14	8	9	6	3
subject6	14	8	7	5	6

Sine /CVC/

	493F1	528F1	565F1	602F1	641F1
subject1	11	10	11	4	4
subject2	7	12	8	9	4
subject3	6	10	10	6	8
subject4	15	9	9	6	1
subject5	12	10	9	5	4
subject6	10	12	10	6	2

VanSon /CVC/

	493F1	528F1	565F1	602F1	641F1
subject1	8	9	12	6	3
subject2	6	12	11	6	4
subject3	13	9	12	2	4
subject4	16	10	9	3	2
subject5	15	10	9	4	2
subject6	12	8	10	6	4

<Results of F1-fixed sets: all numbers are a count of the choices>

Nearey /CVC/

	976F2	1027F2	1080F2	1135F2	1192F2
subject1	5	4	8	11	12
subject2	14	12	8	4	2
subject3	13	13	8	4	2
subject4	9	8	7	10	6
subject5	8	7	12	7	6
subject6	7	5	10	9	9

Sine /CVC/

	976F2	1027F2	1080F2	1135F2	1192F2
subject1	3	4	8	12	15
subject2	14	9	9	4	4
subject3	12	9	8	5	5
subject4	8	5	7	9	11
subject5	3	4	12	11	9
subject6	8	10	10	3	9

Van Son /CVC/

	976F2	1027F2	1080F2	1135F2	1192F2
subject1	4	4	10	11	11
subject2	10	7	7	9	7
subject3	13	8	8	5	5
subject4	11	6	7	10	6
subject5	8	9	10	8	6
subject6	4	13	6	4	13

Appendix X: Output of Loglinear analysis on the F2-fixed data set of the supplementary XAXB experiment, across all six subjects

* * * * * * * HIERARCHICAL LOG LINEAR * * * * * * * *

DATA Information

90 unweighted cases accepted.
0 cases rejected because of out-of-range factor values.
0 cases rejected because of missing data.
717 weighted cases will be used in the analysis.

FACTOR Information

Factor	Level	Label
F1_VALUE	5	F2-fixed token type
SHAPE	3	trajectory shape
SUBJECT	6	

- -

* * * * * * * * HIERARCHICAL LOG LINEAR * * * * * * * *

DESIGN 1 has generating class

F1_VALUE*SHAPE*SUBJECT

Note: For saturated models .500 has been added to all observed cells.
This value may be changed by using the CRITERIA = DELTA subcommand.

The Iterative Proportional Fit algorithm converged at iteration 1.
The maximum difference between observed and fitted marginal totals is .000
and the convergence criterion is .250

- -

Tests that K-way and higher order effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
3	40	17.740	.9991	17.185	.9994	2
2	78	51.529	.9910	50.234	.9939	2
1	89	151.820	.0000	140.197	.0004	0

- -

Tests that K-way effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
1	11	100.291	.0000	89.962	.0000	0
2	38	33.790	.6645	33.049	.6975	0
3	40	17.740	.9991	17.185	.9994	0

******** HIERARCHICAL LOG LINEAR ********

Backward Elimination (p = .050) for DESIGN 1 with generating class

F1_VALUE*SHAPE*SUBJECT

Likelihood ratio chi square = .00000 DF = 0 P = 1.000

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F1_VALUE*SHAPE*SUBJECT	40	17.740	.9991	2

Step 1

The best model has generating class

F1_VALUE*SHAPE
F1_VALUE*SUBJECT
SHAPE*SUBJECT

Likelihood ratio chi square = 17.73966 DF = 40 P = .999

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F1_VALUE*SHAPE	8	4.471	.8123	2
F1_VALUE*SUBJECT	20	29.275	.0825	2
SHAPE*SUBJECT	10	.076	1.0000	2

Step 2

The best model has generating class

F1_VALUE*SHAPE
F1_VALUE*SUBJECT

Likelihood ratio chi square = 17.81551 DF = 50 P = 1.000

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F1_VALUE*SHAPE	8	4.455	.8139	2
F1_VALUE*SUBJECT	20	29.259	.0828	2

******** H I E R A R C H I C A L   L O G   L I N E A R ********

Step 3

The best model has generating class

F1_VALUE*SUBJECT
SHAPE

Likelihood ratio chi square = 22.27086 DF = 58 P = 1.000

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F1_VALUE*SUBJECT	20	29.259	.0828	2
SHAPE	2	.025	.9876	2

Step 4

The best model has generating class

F1_VALUE*SUBJECT

Likelihood ratio chi square = 22.29579 DF = 60 P = 1.000

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F1_VALUE*SUBJECT	20	29.259	.0828	2

Step 5

The best model has generating class

F1_VALUE
SUBJECT

Likelihood ratio chi square = 51.55481 DF = 80 P = .994

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F1_VALUE	4	100.236	.0000	2
SUBJECT	5	.029	1.0000	2

* * * * * * * * H I E R A R C H I C A L   L O G   L I N E A R * * * * * * * *

Step 6

The best model has generating class

F1_VALUE

Likelihood ratio chi square = 51.58410 DF = 85 P = .998

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F1_VALUE	4	100.236	.0000	0

Step 7

The best model has generating class

F1_VALUE

Likelihood ratio chi square = 51.58410 DF = 85 P = .998

- -

* * * * * * * * H I E R A R C H I C A L   L O G   L I N E A R * * * * * * * *

The final model has generating class

F1_VALUE

The Iterative Proportional Fit algorithm converged at iteration 0.
The maximum difference between observed and fitted marginal totals is .000
and the convergence criterion is .250

- -

Observed, Expected Frequencies and Residuals.

Factor	Code	OBS count	EXP count	Residual	Std Resid
F1_VALUE	493F1				
SHAPE	Nearey				
SUBJECT	sub1	12.0	11.2	.83	.25
SUBJECT	sub2	6.0	11.2	-5.17	-1.55
SUBJECT	sub3	10.0	11.2	-1.17	-.35
SUBJECT	sub4	14.0	11.2	2.83	.85
SUBJECT	sub5	14.0	11.2	2.83	.85
SUBJECT	sub6	14.0	11.2	2.83	.85
SHAPE	vSon				
SUBJECT	sub1	8.0	11.2	-3.17	-.95
SUBJECT	sub2	6.0	11.2	-5.17	-1.55
SUBJECT	sub3	13.0	11.2	1.83	.55
SUBJECT	sub4	16.0	11.2	4.83	1.45
SUBJECT	sub5	15.0	11.2	3.83	1.15
SUBJECT	sub6	12.0	11.2	.83	.25
SHAPE	Sine				
SUBJECT	sub1	11.0	11.2	-.17	-.05
SUBJECT	sub2	7.0	11.2	-4.17	-1.25
SUBJECT	sub3	6.0	11.2	-5.17	-1.55
SUBJECT	sub4	15.0	11.2	3.83	1.15
SUBJECT	sub5	12.0	11.2	.83	.25
SUBJECT	sub6	10.0	11.2	-1.17	-.35
F1_VALUE	528F1				
SHAPE	Nearey				
SUBJECT	sub1	10.0	9.8	.22	.07
SUBJECT	sub2	9.0	9.8	-.78	-.25
SUBJECT	sub3	8.0	9.8	-1.78	-.57
SUBJECT	sub4	12.0	9.8	2.22	.71
SUBJECT	sub5	8.0	9.8	-1.78	-.57
SUBJECT	sub6	8.0	9.8	-1.78	-.57
SHAPE	vSon				
SUBJECT	sub1	9.0	9.8	-.78	-.25
SUBJECT	sub2	12.0	9.8	2.22	.71
SUBJECT	sub3	9.0	9.8	-.78	-.25
SUBJECT	sub4	10.0	9.8	.22	.07
SUBJECT	sub5	10.0	9.8	.22	.07
SUBJECT	sub6	8.0	9.8	-1.78	-.57
SHAPE	Sine				
SUBJECT	sub1	10.0	9.8	.22	.07
SUBJECT	sub2	12.0	9.8	2.22	.71
SUBJECT	sub3	10.0	9.8	.22	.07
SUBJECT	sub4	9.0	9.8	-.78	-.25
SUBJECT	sub5	10.0	9.8	.22	.07
SUBJECT	sub6	12.0	9.8	2.22	.71
F1_VALUE	565F1				
SHAPE	Nearey				
SUBJECT	sub1	7.0	9.7	-2.72	-.87
SUBJECT	sub2	13.0	9.7	3.28	1.05
SUBJECT	sub3	10.0	9.7	.28	.09
SUBJECT	sub4	9.0	9.7	-.72	-.23

SUBJECT	sub5	9.0	9.7	-.72	-.23
SUBJECT	sub6	7.0	9.7	-2.72	-.87
SHAPE	vSon				
SUBJECT	sub1	12.0	9.7	2.28	.73
SUBJECT	sub2	11.0	9.7	1.28	.41
SUBJECT	sub3	12.0	9.7	2.28	.73
SUBJECT	sub4	9.0	9.7	-.72	-.23
SUBJECT	sub5	9.0	9.7	-.72	-.23
SUBJECT	sub6	10.0	9.7	.28	.09
SHAPE	Sine				
SUBJECT	sub1	11.0	9.7	1.28	.41
SUBJECT	sub2	8.0	9.7	-1.72	-.55
SUBJECT	sub3	10.0	9.7	.28	.09
SUBJECT	sub4	9.0	9.7	-.72	-.23
SUBJECT	sub5	9.0	9.7	-.72	-.23
SUBJECT	sub6	10.0	9.7	.28	.09
F1_VALUE	602F1				
SHAPE	Nearey				
SUBJECT	sub1	8.0	5.4	2.61	1.12
SUBJECT	sub2	7.0	5.4	1.61	.69
SUBJECT	sub3	5.0	5.4	-.39	-.17
SUBJECT	sub4	3.0	5.4	-2.39	-1.03
SUBJECT	sub5	6.0	5.4	.61	.26
SUBJECT	sub6	5.0	5.4	-.39	-.17
SHAPE	vSon				
SUBJECT	sub1	6.0	5.4	.61	.26
SUBJECT	sub2	6.0	5.4	.61	.26
SUBJECT	sub3	2.0	5.4	-3.39	-1.46
SUBJECT	sub4	3.0	5.4	-2.39	-1.03
SUBJECT	sub5	4.0	5.4	-1.39	-.60
SUBJECT	sub6	6.0	5.4	.61	.26
SHAPE	Sine				
SUBJECT	sub1	4.0	5.4	-1.39	-.60
SUBJECT	sub2	9.0	5.4	3.61	1.56
SUBJECT	sub3	6.0	5.4	.61	.26
SUBJECT	sub4	6.0	5.4	.61	.26
SUBJECT	sub5	5.0	5.4	-.39	-.17
SUBJECT	sub6	6.0	5.4	.61	.26
F1_VALUE	641F1				
SHAPE	Nearey				
SUBJECT	sub1	3.0	3.8	-.78	-.40
SUBJECT	sub2	5.0	3.8	1.22	.63
SUBJECT	sub3	7.0	3.8	3.22	1.66
SUBJECT	sub4	2.0	3.8	-1.78	-.91
SUBJECT	sub5	3.0	3.8	-.78	-.40
SUBJECT	sub6	6.0	3.8	2.22	1.14
SHAPE	vSon				
SUBJECT	sub1	3.0	3.8	-.78	-.40
SUBJECT	sub2	4.0	3.8	.22	.11
SUBJECT	sub3	4.0	3.8	.22	.11
SUBJECT	sub4	2.0	3.8	-1.78	-.91
SUBJECT	sub5	2.0	3.8	-1.78	-.91
SUBJECT	sub6	4.0	3.8	.22	.11

SHAPE	Sine				
SUBJECT	sub1	4.0	3.8	.22	.11
SUBJECT	sub2	4.0	3.8	.22	.11
SUBJECT	sub3	8.0	3.8	4.22	2.17
SUBJECT	sub4	1.0	3.8	-2.78	-1.43
SUBJECT	sub5	4.0	3.8	.22	.11
SUBJECT	sub6	2.0	3.8	-1.78	-.91

- -

* * * * * * * * H I E R A R C H I C A L   L O G   L I N E A R * * * * * * * *

Goodness-of-fit test statistics

Likelihood ratio chi square = 51.58410 DF = 85 P = .998
Pearson chi square = 50.33982 DF = 85 P = .999

- -

Appendix XI: Output of Loglinear analysis on the F1-fixed data set of the supplementary XAXB experiment, across all six subjects

******** HIERARCHICAL LOG LINEAR ********

DATA Information

90 unweighted cases accepted.
0 cases rejected because of out-of-range factor values.
0 cases rejected because of missing data.
720 weighted cases will be used in the analysis.

FACTOR Information

Factor	Level	Label
F2_VALUE	5	F1-fixed token type
SHAPE	3	trajectory shape
SUBJECT	6	

- -

******** HIERARCHICAL LOG LINEAR ********

DESIGN 1 has generating class

F2_VALUE*SHAPE*SUBJECT

Note: For saturated models .500 has been added to all observed cells.
This value may be changed by using the CRITERIA = DELTA subcommand.

The Iterative Proportional Fit algorithm converged at iteration 1.
The maximum difference between observed and fitted marginal totals is .000
and the convergence criterion is .250

- -
- -

Tests that K-way and higher order effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
3	40	29.398	.8913	29.002	.9012	3
2	78	109.962	.0100	105.517	.0207	2
1	89	112.579	.0464	107.250	.0912	0

- -

Tests that K-way effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
1	11	2.618	.9949	1.733	.9992	0
2	38	80.563	.0001	76.514	.0002	0
3	40	29.398	.8913	29.002	.9012	0

******** HIERARCHICAL LOG LINEAR ********

Backward Elimination (p = .050) for DESIGN 1 with generating class

F2_VALUE*SHAPE*SUBJECT

Likelihood ratio chi square = .00000 DF = 0 P = 1.000

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE*SHAPE*SUBJECT	40	29.398	.8913	3

Step 1

The best model has generating class

F2_VALUE*SHAPE
F2_VALUE*SUBJECT
SHAPE*SUBJECT

Likelihood ratio chi square = 29.39826 DF = 40 P = .891

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE*SHAPE	8	5.027	.7547	2
F2_VALUE*SUBJECT	20	75.496	.0000	2
SHAPE*SUBJECT	10	.225	1.0000	2

Step 2

The best model has generating class

F2_VALUE*SHAPE
F2_VALUE*SUBJECT

Likelihood ratio chi square = 29.62288 DF = 50 P = .990

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE*SHAPE	8	4.935	.7645	2
F2_VALUE*SUBJECT	20	75.404	.0000	2

******** HIERARCHICAL LOG LINEAR ********

Step 3

The best model has generating class

F2_VALUE*SUBJECT
SHAPE

Likelihood ratio chi square = 34.55752 DF = 58 P = .994

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE*SUBJECT	20	75.404	.0000	2
SHAPE	2	.000	1.0000	2

Step 4

The best model has generating class

F2_VALUE*SUBJECT

Likelihood ratio chi square = 34.55752 DF = 60 P = .997

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE*SUBJECT	20	75.404	.0000	2

Step 5

The best model has generating class

F2_VALUE*SUBJECT

Likelihood ratio chi square = 34.55752 DF = 60 P = .997

******** HIERARCHICAL LOG LINEAR ********

The final model has generating class

F2_VALUE*SUBJECT

The Iterative Proportional Fit algorithm converged at iteration 0.
The maximum difference between observed and fitted marginal totals is .000
and the convergence criterion is .250

- -

Observed, Expected Frequencies and Residuals.

Factor	Code	OBS count	EXP count	Residual	Std Resid
F2_VALUE	976F2				
SHAPE	Nearey				
SUBJECT	sub1	5.0	4.0	1.00	.50
SUBJECT	sub2	14.0	12.7	1.33	.37
SUBJECT	sub3	13.0	12.7	.33	.09
SUBJECT	sub4	9.0	9.3	-.33	-.11
SUBJECT	sub5	8.0	6.3	1.67	.66
SUBJECT	sub6	7.0	6.3	.67	.26
SHAPE	vSon				
SUBJECT	sub1	4.0	4.0	.00	.00
SUBJECT	sub2	10.0	12.7	-2.67	-.75
SUBJECT	sub3	13.0	12.7	.33	.09
SUBJECT	sub4	11.0	9.3	1.67	.55
SUBJECT	sub5	8.0	6.3	1.67	.66
SUBJECT	sub6	4.0	6.3	-2.33	-.93
SHAPE	Sine				
SUBJECT	sub1	3.0	4.0	-1.00	-.50
SUBJECT	sub2	14.0	12.7	1.33	.37
SUBJECT	sub3	12.0	12.7	-.67	-.19
SUBJECT	sub4	8.0	9.3	-1.33	-.44
SUBJECT	sub5	3.0	6.3	-3.33	-1.32
SUBJECT	sub6	8.0	6.3	1.67	.66
F2_VALUE	1027F2				
SHAPE	Nearey				
SUBJECT	sub1	4.0	4.0	.00	.00
SUBJECT	sub2	12.0	9.3	2.67	.87
SUBJECT	sub3	13.0	10.0	3.00	.95
SUBJECT	sub4	8.0	6.3	1.67	.66
SUBJECT	sub5	7.0	6.7	.33	.13
SUBJECT	sub6	5.0	9.3	-4.33	-1.42
SHAPE	vSon				
SUBJECT	sub1	4.0	4.0	.00	.00
SUBJECT	sub2	7.0	9.3	-2.33	-.76
SUBJECT	sub3	8.0	10.0	-2.00	-.63
SUBJECT	sub4	6.0	6.3	-.33	-.13
SUBJECT	sub5	9.0	6.7	2.33	.90
SUBJECT	sub6	13.0	9.3	3.67	1.20
SHAPE	Sine				

SUBJECT	sub1	4.0	4.0	.00	.00
SUBJECT	sub2	9.0	9.3	-.33	-.11
SUBJECT	sub3	9.0	10.0	-1.00	-.32
SUBJECT	sub4	5.0	6.3	-1.33	-.53
SUBJECT	sub5	4.0	6.7	-2.67	-1.03
SUBJECT	sub6	10.0	9.3	.67	.22
F2_VALUE	1080F2				
SHAPE	Nearey				
SUBJECT	sub1	8.0	8.7	-.67	-.23
SUBJECT	sub2	8.0	8.0	.00	.00
SUBJECT	sub3	8.0	8.0	.00	.00
SUBJECT	sub4	7.0	7.0	.00	.00
SUBJECT	sub5	12.0	11.3	.67	.20
SUBJECT	sub6	10.0	8.7	1.33	.45
SHAPE	vSon				
SUBJECT	sub1	10.0	8.7	1.33	.45
SUBJECT	sub2	7.0	8.0	-1.00	-.35
SUBJECT	sub3	8.0	8.0	.00	.00
SUBJECT	sub4	7.0	7.0	.00	.00
SUBJECT	sub5	10.0	11.3	-1.33	-.40
SUBJECT	sub6	6.0	8.7	-2.67	-.91
SHAPE	Sine				
SUBJECT	sub1	8.0	8.7	-.67	-.23
SUBJECT	sub2	9.0	8.0	1.00	.35
SUBJECT	sub3	8.0	8.0	.00	.00
SUBJECT	sub4	7.0	7.0	.00	.00
SUBJECT	sub5	12.0	11.3	.67	.20
SUBJECT	sub6	10.0	8.7	1.33	.45
F2_VALUE	1135F2				
SHAPE	Nearey				
SUBJECT	sub1	11.0	11.3	-.33	-.10
SUBJECT	sub2	4.0	5.7	-1.67	-.70
SUBJECT	sub3	4.0	4.7	-.67	-.31
SUBJECT	sub4	10.0	9.7	.33	.11
SUBJECT	sub5	7.0	8.7	-1.67	-.57
SUBJECT	sub6	9.0	5.3	3.67	1.59
SHAPE	vSon				
SUBJECT	sub1	11.0	11.3	-.33	-.10
SUBJECT	sub2	9.0	5.7	3.33	1.40
SUBJECT	sub3	5.0	4.7	.33	.15
SUBJECT	sub4	10.0	9.7	.33	.11
SUBJECT	sub5	8.0	8.7	-.67	-.23
SUBJECT	sub6	4.0	5.3	-1.33	-.58
SHAPE	Sine				
SUBJECT	sub1	12.0	11.3	.67	.20
SUBJECT	sub2	4.0	5.7	-1.67	-.70
SUBJECT	sub3	5.0	4.7	.33	.15
SUBJECT	sub4	9.0	9.7	-.67	-.21
SUBJECT	sub5	11.0	8.7	2.33	.79
SUBJECT	sub6	3.0	5.3	-2.33	-1.01

F2_VALUE	1192F2				
SHAPE	Nearey				
SUBJECT	sub1	12.0	12.7	-.67	-.19
SUBJECT	sub2	2.0	4.3	-2.33	-1.12
SUBJECT	sub3	2.0	4.0	-2.00	-1.00
SUBJECT	sub4	6.0	7.7	-1.67	-.60
SUBJECT	sub5	6.0	7.0	-1.00	-.38
SUBJECT	sub6	9.0	10.3	-1.33	-.41
SHAPE	vSon				
SUBJECT	sub1	11.0	12.7	-1.67	-.47
SUBJECT	sub2	7.0	4.3	2.67	1.28
SUBJECT	sub3	5.0	4.0	1.00	.50
SUBJECT	sub4	6.0	7.7	-1.67	-.60
SUBJECT	sub5	6.0	7.0	-1.00	-.38
SUBJECT	sub6	13.0	10.3	2.67	.83
SHAPE	Sine				
SUBJECT	sub1	15.0	12.7	2.33	.66
SUBJECT	sub2	4.0	4.3	-.33	-.16
SUBJECT	sub3	5.0	4.0	1.00	.50
SUBJECT	sub4	11.0	7.7	3.33	1.20
SUBJECT	sub5	9.0	7.0	2.00	.76
SUBJECT	sub6	9.0	10.3	-1.33	-.41

- -

******** H I E R A R C H I C A L   L O G   L I N E A R ********

Goodness-of-fit test statistics

Likelihood ratio chi square = 34.55752 DF = 60 P = .997
Pearson chi square = 34.06990 DF = 60 P = .997

- -

Appendix XII: Output of Loglinear analysis on the F1-fixed data set of the supplementary XAXB experiment, for a group of Subjects 2 and 3

* * * * * * * * H I E R A R C H I C A L L O G L I N E A R * * * * * * * *

DATA Information

30 unweighted cases accepted.
60 cases rejected because of out-of-range factor values.
0 cases rejected because of missing data.
238 weighted cases will be used in the analysis.

FACTOR Information

Factor	Level	Label
F2_VALUE	5	F1-fixed token type
SHAPE	3	trajectory shape
SUBGROUP	2	subject sub grouping; subjects 2 and 3

- -

* * * * * * * * H I E R A R C H I C A L L O G L I N E A R * * * * * * * *

DESIGN 1 has generating class

F2_VALUE*SHAPE*SUBGROUP

Note: For saturated models .500 has been added to all observed cells.
This value may be changed by using the CRITERIA = DELTA subcommand.

The Iterative Proportional Fit algorithm converged at iteration 1.
The maximum difference between observed and fitted marginal totals is .000
and the convergence criterion is .250

- -

Tests that K-way effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
1	7	36.086	.0000	33.555	.0000	0
2	14	9.702	.7836	9.939	.7667	0
3	8	2.129	.9768	2.120	.9771	0

* * * * * * * * H I E R A R C H I C A L L O G L I N E A R * * * * * * * *

Backward Elimination (p = .050) for DESIGN 1 with generating class

F2_VALUE*SHAPE*SUBGROUP

Likelihood ratio chi square = .00000 DF = 0 P = 1.000

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE*SHAPE*SUBGROUP	8	2.129	.9768	2

Step 1

The best model has generating class

F2_VALUE*SHAPE
F2_VALUE*SUBGROUP
SHAPE*SUBGROUP

Likelihood ratio chi square = 2.12922 DF = 8 P = .977

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE*SHAPE	8	9.311	.3168	2
F2_VALUE*SUBGROUP	4	.378	.9843	2
SHAPE*SUBGROUP	2	.003	.9984	2

Step 2

The best model has generating class

F2_VALUE*SHAPE
F2_VALUE*SUBGROUP

Likelihood ratio chi square = 2.13251 DF = 10 P = .995

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE*SHAPE	8	9.316	.3163	2
F2_VALUE*SUBGROUP	4	.383	.9838	2

******** H I E R A R C H I C A L   L O G   L I N E A R ********

Step 3

The best model has generating class

F2_VALUE*SHAPE
SUBGROUP

Likelihood ratio chi square = 2.51547 DF = 14 P = 1.000

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE*SHAPE	8	9.316	.3163	2
SUBGROUP	1	.017	.8968	2

Step 4

The best model has generating class

F2_VALUE*SHAPE

Likelihood ratio chi square = 2.53228 DF = 15 P = 1.000

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE*SHAPE	8	9.316	.3163	2

Step 5

The best model has generating class

F2_VALUE
SHAPE

Likelihood ratio chi square = 11.84830 DF = 23 P = .973

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE	4	36.061	.0000	2
SHAPE	2	.008	.9958	2

******** H I E R A R C H I C A L   L O G   L I N E A R ********

Step 6

The best model has generating class

F2_VALUE

Likelihood ratio chi square = 11.85665 DF = 25 P = .988

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE	4	36.061	.0000	0

Step 7

The best model has generating class

F2_VALUE

Likelihood ratio chi square = 11.85665 DF = 25 P = .988

- -

******** HIERARCHICAL LOG LINEAR ********

The final model has generating class

F2_VALUE

The Iterative Proportional Fit algorithm converged at iteration 0.
The maximum difference between observed and fitted marginal totals is .000
and the convergence criterion is .250

- -

Observed, Expected Frequencies and Residuals.
gp1= subject 2; gp2 = subject 3

Factor	Code	OBS count	EXP count	Residual	Std Resid
F2_VALUE	976F2				
SHAPE	Nearey				
SUBGROUP	gp1	14.0	12.7	1.33	.37
SUBGROUP	gp2	13.0	12.7	.33	.09
SHAPE	vSon				
SUBGROUP	gp1	10.0	12.7	-2.67	-.75
SUBGROUP	gp2	13.0	12.7	.33	.09
SHAPE	Sine				
SUBGROUP	gp1	14.0	12.7	1.33	.37
SUBGROUP	gp2	12.0	12.7	-.67	-.19
F2_VALUE	1027F2				
SHAPE	Nearey				
SUBGROUP	gp1	12.0	9.7	2.33	.75
SUBGROUP	gp2	13.0	9.7	3.33	1.07
SHAPE	vSon				
SUBGROUP	gp1	7.0	9.7	-2.67	-.86
SUBGROUP	gp2	8.0	9.7	-1.67	-.54

SHAPE Sine					
SUBGROUP	gp1	9.0	9.7	-.67	-.21
SUBGROUP	gp2	9.0	9.7	-.67	-.21
F2_VALUE 1080F2					
SHAPE Nearey					
SUBGROUP	gp1	8.0	8.0	.00	.00
SUBGROUP	gp2	8.0	8.0	.00	.00
SHAPE vSon					
SUBGROUP	gp1	7.0	8.0	-1.00	-.35
SUBGROUP	gp2	8.0	8.0	.00	.00
SHAPE Sine					
SUBGROUP	gp1	9.0	8.0	1.00	.35
SUBGROUP	gp2	8.0	8.0	.00	.00
F2_VALUE 1135F2					
SHAPE Nearey					
SUBGROUP	gp1	4.0	5.2	-1.17	-.51
SUBGROUP	gp2	4.0	5.2	-1.17	-.51
SHAPE vSon					
SUBGROUP	gp1	9.0	5.2	3.83	1.69
SUBGROUP	gp2	5.0	5.2	-.17	-.07
SHAPE Sine					
SUBGROUP	gp1	4.0	5.2	-1.17	-.51
SUBGROUP	gp2	5.0	5.2	-.17	-.07
F2_VALUE 1192F2					
SHAPE Nearey					
SUBGROUP	gp1	2.0	4.2	-2.17	-1.06
SUBGROUP	gp2	2.0	4.2	-2.17	-1.06
SHAPE vSon					
SUBGROUP	gp1	7.0	4.2	2.83	1.39
SUBGROUP	gp2	5.0	4.2	.83	.41
SHAPE Sine					
SUBGROUP	gp1	4.0	4.2	-.17	-.08
SUBGROUP	gp2	5.0	4.2	.83	.41

- -

Goodness-of-fit test statistics

Likelihood ratio chi square = 11.85665 DF = 25 P = .988
Pearson chi square = 12.13748 DF = 25 P = .985

- -

Appendix XIII: Output of Loglinear analysis on the F1-fixed data set of the supplementary XAXB experiment, for a group of Subjects 4, 5 and 6

* * * * * * * * HIERARCHICAL LOG LINEAR * * * * * * * *

DATA Information

45 unweighted cases accepted.
45 cases rejected because of out-of-range factor values.
0 cases rejected because of missing data.
360 weighted cases will be used in the analysis.

FACTOR Information

Factor	Level	Label
SHAPE	3	trajectory shape
F2_VALUE	5	F1-fixed token type
SUBJECT	3	

- -

* * * * * * * * HIERARCHICAL LOG LINEAR * * * * * * * *

DESIGN 1 has generating class

SHAPE*F2_VALUE*SUBJECT

Note: For saturated models .500 has been added to all observed cells.
This value may be changed by using the CRITERIA = DELTA subcommand.

The Iterative Proportional Fit algorithm converged at iteration 1.
The maximum difference between observed and fitted marginal totals is .000
and the convergence criterion is .250

- -

Tests that K-way and higher order effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
3	16	16.112	.4452	15.819	.4657	3
2	36	35.504	.4920	34.024	.5629	2
1	44	37.609	.7406	35.750	.8076	0

- -

Tests that K-way effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
1	8	2.105	.9776	1.726	.9883	0
2	20	19.393	.4964	18.206	.5739	0
3	16	16.112	.4452	15.819	.4657	0

* * * * * * * * HIERARCHICAL LOG LINEAR * * * * * * * *

Backward Elimination (p = .050) for DESIGN 1 with generating class

SHAPE*F2_VALUE*SUBJECT

Likelihood ratio chi square = .00000 DF = 0 P = 1.000

\- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
SHAPE*F2_VALUE*SUBJECT	16	16.112	.4452	3

Step 1

The best model has generating class

SHAPE*F2_VALUE
SHAPE*SUBJECT
F2_VALUE*SUBJECT

Likelihood ratio chi square = 16.11168 DF = 16 P = .445

\- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
SHAPE*F2_VALUE	8	5.404	.7137	2
SHAPE*SUBJECT	4	.136	.9978	2
F2_VALUE*SUBJECT	8	14.058	.0803	2

Step 2

The best model has generating class

SHAPE*F2_VALUE
F2_VALUE*SUBJECT

Likelihood ratio chi square = 16.24733 DF = 20 P = .701

\- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
SHAPE*F2_VALUE	8	5.301	.7249	2
F2_VALUE*SUBJECT	8	13.956	.0829	2

* * * * * * * * HIERARCHICAL LOG LINEAR * * * * * * * *

Step 3

The best model has generating class

F2_VALUE*SUBJECT
SHAPE

Likelihood ratio chi square = 21.54869 DF = 28 P = .802

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE*SUBJECT	8	13.956	.0829	2
SHAPE	2	.017	.9917	2

Step 4

The best model has generating class

F2_VALUE*SUBJECT

Likelihood ratio chi square = 21.56540 DF = 30 P = .869

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE*SUBJECT	8	13.956	.0829	2

Step 5

The best model has generating class

F2_VALUE
SUBJECT

Likelihood ratio chi square = 35.52109 DF = 38 P = .585

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE	4	2.088	.7195	1
SUBJECT	2	.000	1.0000	2

* * * * * * * * H I E R A R C H I C A L   L O G   L I N E A R * * * * * * * *

Step 6

The best model has generating class

F2_VALUE

Likelihood ratio chi square = 35.52109 DF = 40 P = .672

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
F2_VALUE	4	2.088	.7195	0

Step 7

The best model has no factors (constant only model)

Likelihood ratio chi square = 37.60926 DF = 44 P = .741

- -

Step 8

The best model has no factors (constant only model)

Likelihood ratio chi square = 37.60926 DF = 44 P = .741

- -

* * * * * * * * H I E R A R C H I C A L   L O G   L I N E A R * * * * * * * *

The final model has no factors (constant only model)

* * * * * * * * H I E R A R C H I C A L   L O G   L I N E A R * * * * * * * *

The Iterative Proportional Fit algorithm converged at iteration 0.
The maximum difference between observed and fitted marginal totals is 8.000
and the convergence criterion is .250

Observed, Expected Frequencies and Residuals.

Factor	Code	OBS count	EXP count	Residual	Std Resid
SHAPE	Nearey				
F2_VALUE	976F2				
SUBJECT	sub4	9.0	8.0	1.00	.35
SUBJECT	sub5	8.0	8.0	.00	.00
SUBJECT	sub6	7.0	8.0	-1.00	-.35
F2_VALUE	1027F2				
SUBJECT	sub4	8.0	8.0	.00	.00
SUBJECT	sub5	7.0	8.0	-1.00	-.35
SUBJECT	sub6	5.0	8.0	-3.00	-1.06
F2_VALUE	1080F2				
SUBJECT	sub4	7.0	8.0	-1.00	-.35
SUBJECT	sub5	12.0	8.0	4.00	1.41
SUBJECT	sub6	10.0	8.0	2.00	.71
F2_VALUE	1135F2				
SUBJECT	sub4	10.0	8.0	2.00	.71
SUBJECT	sub5	7.0	8.0	-1.00	-.35
SUBJECT	sub6	9.0	8.0	1.00	.35
F2_VALUE	1192F2				
SUBJECT	sub4	6.0	8.0	-2.00	-.71
SUBJECT	sub5	6.0	8.0	-2.00	-.71
SUBJECT	sub6	9.0	8.0	1.00	.35
SHAPE	vSon				
F2_VALUE	976F2				
SUBJECT	sub4	11.0	8.0	3.00	1.06
SUBJECT	sub5	8.0	8.0	.00	.00
SUBJECT	sub6	4.0	8.0	-4.00	-1.41
F2_VALUE	1027F2				
SUBJECT	sub4	6.0	8.0	-2.00	-.71
SUBJECT	sub5	9.0	8.0	1.00	.35
SUBJECT	sub6	13.0	8.0	5.00	1.77
F2_VALUE	1080F2				
SUBJECT	sub4	7.0	8.0	-1.00	-.35
SUBJECT	sub5	10.0	8.0	2.00	.71
SUBJECT	sub6	6.0	8.0	-2.00	-.71
F2_VALUE	1135F2				
SUBJECT	sub4	10.0	8.0	2.00	.71
SUBJECT	sub5	8.0	8.0	.00	.00
SUBJECT	sub6	4.0	8.0	-4.00	-1.41
F2_VALUE	1192F2				
SUBJECT	sub4	6.0	8.0	-2.00	-.71
SUBJECT	sub5	6.0	8.0	-2.00	-.71
SUBJECT	sub6	13.0	8.0	5.00	1.77
SHAPE	Sine				
F2_VALUE	976F2				
SUBJECT	sub4	8.0	8.0	.00	.00
SUBJECT	sub5	3.0	8.0	-5.00	-1.77
SUBJECT	sub6	8.0	8.0	.00	.00
F2_VALUE	1027F2				

SUBJECT	sub4	5.0	8.0	-3.00	-1.06
SUBJECT	sub5	4.0	8.0	-4.00	-1.41
SUBJECT	sub6	10.0	8.0	2.00	.71
F2_VALUE	1080F2				
SUBJECT	sub4	7.0	8.0	-1.00	-.35
SUBJECT	sub5	12.0	8.0	4.00	1.41
SUBJECT	sub6	10.0	8.0	2.00	.71
F2_VALUE	1135F2				
SUBJECT	sub4	9.0	8.0	1.00	.35
SUBJECT	sub5	11.0	8.0	3.00	1.06
SUBJECT	sub6	3.0	8.0	-5.00	-1.77
F2_VALUE	1192F2				
SUBJECT	sub4	11.0	8.0	3.00	1.06
SUBJECT	sub5	9.0	8.0	1.00	.35
SUBJECT	sub6	9.0	8.0	1.00	.35

- -

Goodness-of-fit test statistics

Likelihood ratio chi square = 37.60926 DF = 44 P = .741
Pearson chi square = 35.75000 DF = 44 P = .808

- -