Bidirectional perception of lexical prominence in Spanish and Japanese as second languages

José Joaquín Atria

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London

Speech Hearing and Phonetic Sciences

University College London

2016

I, José Joaquín Atria, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

This dissertation was typeset using XƎLaTeX.
The main font in this document is Linux Libertine. Margin notes are set
in Linux Biolinum, and monospaced text is set in Inconsolata.
Japanese text is set in Takao Mincho.
PGFPlots and Inkscape were used to generate plots and figures.
The fleuron ornament was provided by Vectorian.
The digital version of this document contains no raster graphics.
This dissertation was made entirely with Free Software.

Abstract

Perception of novel phonetic contrasts in a second language has been studied extensively, but suprasegmentals have seen relatively little attention even though difficulties at this level can strongly impact comprehension. Available studies suggest that the perception of segmental and suprasegmental categories are subject to similar factors, but evidence is not entirely conclusive. Additionally, studies focusing on the perception of lexical prominence have suggested that perception across languages with different accent types might be particularly problematic, meaning that these effects in particular would be independent of the direction of language learning.

This dissertation explores both of these questions by studying the perception of lexical prominence by second-language learners in Spanish (a stress-accent language) and Japanese (a pitch-accent language). Following a bidirectional approach, it examines whether the perception of phonologically different types of lexical prominence is subject to similar effects as those traditionally identified for cross-linguistic segmental perception, and how these relate to the direction of learning.

A first set of studies provides a comparative acoustic description of prominence in both languages, and presents the results of an identification task with natural words in different positions within a sentence. Using multiple speakers, these tests showed that the difficulties seen by both groups are different and related to features in their L1, and that despite phonological differences, contexts existed in which high performance was possible. A second set of studies explored the sensitivity of non-native listeners to secondary acoustic cues and the development of new accentual categories, and showed effects of learning for both groups and a strong sensitivity to duration for learners of Spanish. Learners of Japanese showed extremely poor category development for unaccented words in particular.

Overall results show that existing research on SLA is applicable to suprasegmental perception, and that the transfer effects affecting both groups have different domains and scope. The implications for language teaching and theories of L2 perception are discussed.

Acknowledgements

Nothing could have prepared me for what doing a PhD would actually entail. The time spent, the uncertainty, the insecurities, the failures and dead ends… and at the end, the elation and relief of a work well done. If I had a data point for each time I stopped and thought that the whole endeavour was hopeless; for each time I changed my mind and thought that it was not only worthwhile but extremely exciting; for each time I looked back and thought that something should have been done differently, my results would have much more statistical power.

Definitely the most humbling part of the entire experience was to realise just how many people are needed to finish a PhD, how many lives are touched in the process, and how thankful you feel for all the ways in which the people around you helped you get to the end.

This is what makes an acknowledgement section so necessary, and why writing one seems so futile: because for every person that is mentioned by name, there are 10 more that are being left behind. Knowing this, this is my attempt to thank all those without whom this would have been a failure. If you think you deserve a place here and cannot find your name, I hope you will forgive my oversight. Next time we meet, the beer is on me.

First I need to thank Valerie Hazan, for her undying support and attention, and her most valuable advice throughout. In the four years (and a bit) that I had the privilege to work under her supervision, she always had her door open, and was always willing to solve my problems and read over my drafts. The comments I received in return were as harsh as they needed to be, never more, and always helpful.

I would also like to thank Paul Iverson, for his expertise and his *take-no-prisoners* approach to giving advice, and to Mark Huckvale, for his advice on intensity normalisation strategies.

Paul is also largely responsible for my spending a year at UMass Amherst, where professors John Kingston and Kristine Yu took me in and offered a completely different perspective to the things I had been doing, and had yet to do. I have particularly fond memories of the last conversation I had with John before heading back, and I only wish I can express how lucky I feel for having had the opportunity to be in his lab. The students in the Linguistics department also made me feel welcome and challenged, and I cherish the friendships I made while with them. We should go climbing again.

In Sophia University I also received the help of professors Arai Takayuki and Kimiyo Nishimura, who not only provided me with a place to work, but assisted in the data collection in ways I did not think would be possible.

Thank you also to the people at Universidad Católica, and in particular to Domingo Román for his constant encouragement, his friendship, and his unwavering faith in me.

UCL and SHAPS (with comma or without) became for me so much more than a place to work, and having to move on is the one part of completing my PhD that makes me sad. Thank you in particular to the members of the lab: Sonia Granlund, Michèle Pettinato, Outi Tuomainen, Wafa Alshangiti, and Lou Stringer; and Yasuaki Shinohara, who only realised he could call himself a member as he was leaving UCL to bigger and better things. The rest of the graduate students are the main reason my stay in UCL has been as memorable as it has. A special thank you to Melanie Pinet, Georgina Roth, Albert Lee, Lucy Carroll, Tim Schoof, Cristiane Hsu, Mauricio Figueroa Candia, Kurt Steinmetzger, Emma Brint, Yue Zhang, Jieun Song, Petra Hödl; my housemates Faith Chiu and Gisela Tomé Lourido; and Mark Wibrow, without whom this dissertation would have been significantly uglier (and less fun to write).

My gratitude goes also to my thesis examiners, professors Peter Howell and Joaquim Llisterri, who helped make the last part of the PhD the most exciting and rewarding of all. They offered me their sincere advice and made significant contributions to the final stages of this little monster.

A special mention is also due to those who participated in my admittedly tedious tests, without whom no science would have been possible; and to those who so efficiently helped me in collecting data for the latter stages in my research, when data collection is the last thing one wants to do. Thank you to Karina Cerda, Rosa Catalán and Imran Hossain in Santiago; Hernán Acuña in Concepción; and Misae Kunitake and Kazuaki Yasuki in Tōkyō. I am in your debt.

I am also grateful to the professors who I met while at TUFS, and in particular to Shigenobu Kawakami, who hosted me while there, and Takuya Kimura, with whom a productive and rewarding academic collaboration lead eventually into this dissertation. That work would not have been the same without the rest of the members of that team: professors Hirotaka Sensui, Atsuko Toyomaru and Miyuki Takasawa. Professor Noriko Nakamura also offered her helpful advice on the teaching of intonation in Japanese and the search for appropriate Japanese words to use. She is also responsible for much of my initial knowledge of the Japanese accent. Many thanks to you all.

Thank you also to my family, and friends, who helped me in all the other ways in which company and support can be helpful. Even in the distance, I never stopped feeling your love.

Finally, thank you to Koto Sadamura, my partner in crime and love, for the support, the motivation, and the visions of a brighter future. Not only did you make this possible, but you also made it so much more enjoyable. You are the reason all of this made sense.

Contents

List of Figures

List of Tables

Glossary

isol Word in isolation.

nint Non-final position in a negative statement.

qfin Final position in an interrogative statement.

qint Non-final position in an interrogative statement.

sfin Final position in a declarative statement.

sint Non-final position in a declarative statement.

1SG 1st person singular.

2SG 2nd person singular.

AFC Alternative forced choice.

AIC Akaike information criterion.

AJH Japanese Society of Hispanists.

AM Autosegmental-Metric.

ANCOVA Analysis of co-variance.

ANOVA Analysis of variance.

CG Category goodness contrast.

CLF1 Chilean older female speaker.

CLM1 Chilean younger male speaker.

FLA Foreign-language acquisition.

HTML HyperText Markup Language.

IPA International Phonetic Association.

J1S2 Japanese students of Spanish.

JPF1 Japanese older female speaker.

JPF2 Japanese younger female speaker.

JPM2 Japanese older male speaker.

LOR Length of residence.

L1 First language.

L2 Second language.

MP3 Mpeg-2 Audio Layer III.

NA Non-speech contrast.

NLM Native Language Magnet Model.

NLM-e Extended Native Language Magnet Model.

N Noun (part of speech).

OJAD Online Japanese Accent Dictionary.

PAM Perceptual Assimilation Model.

PAM-L2 Perceptual Assimilation Model, extended for Second Languages.

PAST Past (grammatic tense).

PENTA Parallel Encoding and Tonal Approximation.

PHP PHP: Hypertext Preprocessor.

PRES Present (grammatic tense).

RMS Root-mean-square.

S1J2 Chilean students of Japanese.

SC Single category contrast.

SLA Second-language acquisition.

SLM Speech Learning Model.

SOV Subject-Object-Verb.

STRAIGHT Speech Transformation and Representation using Adaptive Interpolation of weiGHTed spectrum.

SVO Subject-Verb-Object.

TBU Tone-bearing unit.

TC Two-category contrast.

TOBI Tone and Break Indices.

UC Contrast between categorised and uncategorised sounds.

UU Contrast between two uncategorised sounds.

VUV Voiced-unvoiced.

V Verb (part of speech).

XML Extended Markup Language.

`lame` LAME Ain't an MP3 Encoder.

You can't always get what you want
But if you try sometimes,
You just might find
You get what you need

You Can't Always Get What You Want
The Rolling Stones

Chapter 1

Introduction

The focus of this thesis is the perception of lexical prominence in a second language (L2), and in particular of Spanish and Japanese as L2. An extensive body of work on the acquisition and learning of phonetic contrasts in second languages has managed to provide thorough coverage of the difficulties involved in the process, and in particular of the effects that categorical contrasts in the first language (L1) have on the perception and production of second languages (*e.g.* Broselow, Hurtig, and Ringen, 1987; Best, McRoberts, and Goodell, 2001; Flege, 2003; Kuhl, Conboy, et al., 2008, among others). However, most of the literature covering L2 perception in particular has focused on the perception and acquisition of segments, and the field of suprasegmentals has been relatively neglected: in a recent review (Thomson and Derwing, 2014) only 23% of the 75 papers they reviewed focused on suprasegmentals, with an additional 24% examining them only in combination with segments. And these numbers are at their peak: in that same study the authors state this is the result of "a shift in the tide" of research on the production of L2 prosody; and Derwing and Munro (2005) — an earlier review of L2 studies on pronunciation in general (including segmental and suprasegmental research) — found a dearth of work on pronunciation, and little effect of these studies on the teacher-oriented literature.

Research on suprasegmentals

Part of this tidal shift is the increasing number of studies that support the need to spend resources in the study of L2 suprasegmental perception and production, and in particular on its effects in comprehensibility, intelligibility, and foreign accentedness (Grant, 2014; Wei, Gubbins, and Idemaru, 2015).

Benefits for foreign accentedness

Effects of suprasegmentals on the latter are well established, and so is the fact that they have a stronger effect than segments (Anderson-Hsieh, Johnson, and Koehler, 1992; Koster and Koet, 1993; Munro, 1995; Munro and Derwing, 1999; Wennerstrom, 2000; Chun, 2002; Trofimovich and Baker, 2006, among others). And while it is true that foreign accentedness and comprehensibility are independent from one another (Munro and Derwing, 1999), and that the focus of second language learning should be on the latter, it is also true that suprasegmentals have an effect on other more important aspects of L2 use.

Other benefits

These include problems with categorisation, particularly in the case of poor perception of nuclear stress (Levis, 1999); problems with word segmentation, due to improper use of stress-pattern cues (Sanders, Neville, and Woldorff, 2002); and comprehensibility problems with speech produced by L2 speakers with inaccurate stress positions (Bansal, 1966; Tiffen, 1974; Benrabah, 1987, reviewed in Benrabah, 1997). More generally, Chun (2002) states that prosodic training reduces the effects of L1 transfers in production, and minimises the risk for non-native speakers of pragmatic and sociolinguistic misunderstandings, apart from the beneficial effects on foreign accentedness.

1.1 Language transfer effects

Most of the work on the difficulties learners face when learning an L2 has dealt with the issues related to language experience, including the relationship between the native and non-native language phonemic inventories, the effect of age-related factors, and the amount of L2 use and contact.

A considerable part of the recent research on the interaction between the learner's L1 and L2 has been undertaken in reference to one of three major theoretical frameworks on language learning: the Perceptual Assimilation Model, the Speech Learning Model, and the Native Language Magnet Model.The following section will provide a short overview of each of these models.

1.1.1 Perceptual Assimilation Model

The Perceptual Assimilation Model (PAM; Best, 1995) was first proposed as an explanation of the issues involved in the perception of non-native segments by mature naïve listeners within the direct realist framework (Reed and Jones, 1982; Fowler, 1986; Gibson, 1991). The model supposes that speech perception occurs at a gestural level, and is unmediated by mental associations (*i.e.* that gestures are perceived directly instead of through pre-existing notions of what the intended gesture may be). Since all spoken human languages use the same articulatory organs for their gestures, their categorisation also shares a common space, causing native and non-native gestures to occasionally overlap. The perceptual assimilation that results from these overlaps gives the model its name.

Contrasts

Depending on the magnitude of these overlaps, non-native sounds can be categorised as a better or worse member of an already existing native category; as outside the boundaries of any single native category but still within the native phonological space if no overlap is shown; or as a non-speech sound, outside this phonological space (*i.e.* an uncategorisable sound). Based on this, the model recognises six possible contrasts[1]:

TC A two category contrast, in which each segment is perceived as a member of a different native category.

SC A single category contrast, if both non-native sounds are assimilated as members of the same native category and are both considered to be equally good or bad exemplars of it.

CG A category goodness contrast, if both of them fall within the same native category but one is considered to be a better exemplar than the other.

UC A contrast between a categorised and an uncategorised sound.

UU A contrast between two uncategorisable phones, when they are both perceived as speech sounds but both lie outside the boundaries of existing native categories.

NA A contrast between two sounds perceived to be non-speech sounds.

1 Abbreviations for these contrasts follow those provided originally by Best (1995).

Predictions

The model makes predictions regarding the discriminability of two sounds based on the type of contrast that they generate. The best discrimination is predicted with TC and, to a lesser extent, UC contrasts, since in both cases the sounds will reside on different sides of a category boundary, where discrimination is highest. In other cases discrimination will depend on the specific gestural characteristics of the sounds, and how much they differ from one another and from native categories: with UU contrasts discrimination can be anywhere "from poor to very good"; CG contrasts are predicted to have "moderate to very good" discrimination; while NA contrasts are expected to have "good to very good" discrimination. Single category contrasts are always predicted to have poor discrimination, but not necessarily as low as chance level.

Model extensions

The earliest versions of PAM were meant only for the perception by naïve and mature listeners, and was also addressed specifically to the perception of segments. Since then, the model has been extended to include second language learners (PAM-L2; Best and Tyler, 2007). The extension was developed as a way to provide PAM-compliant explanations to the postulates of the Speech Learning Model (SLM; a different perception model, see below). With these extensions, PAM-L2 explicitly recognises that, while perceptual learning processes remain unchanged throughout the lifespan, there is a considerable difference between the way children acquire their L1 and the way adults learn a new language, not only because of changes in the individual, but also and greatly because of changes in the learning environment. It also posits that L2 perception is heavily dependent on the listener's focus of attention, with some sound contrasts requiring attention only at the gestural level (like that between a /p/ and a non-speech sound); and others (and in particular those involving L2 sounds) requiring attention at a phonetic or phonological level, or a combination of the two. For PAM-L2, both L1 and L2 categories exist in a shared space, an *interlanguage*, meaning that L2 input may result on the modification of L1 categories. But since L2 learning involves the interaction of phonetic and phonological levels that are largely independent, it is possible that a new sound is perceived to be part of an interlanguage phonological category while remaining discriminable as a different realisation of that category at a language-specific phonetic level. The shift into and out of cases like this will depend on the shift in the attention focus of the learner, which the authors suggest may move from a phonetic level to higher levels as the learning progresses.

1.1.2 Speech Learning Model

The Speech Learning Model (SLM; Flege, 1995, 2007) had a number of similarities to PAM-L2 (since the latter was developed as a response to the former), but it was always thought of as a model to explain the acquisition of L2 in adult learners with extensive exposure to the target language. A key difference between SLM and both versions of PAM is that SLM posits that L2 perception occurs at an acoustic level instead of a gestural one, and that listeners use this acoustic input to generate mental representations of those sounds, called "phonetic categories". These categories continue to change and adapt throughout the entire lifespan of the person, and the mechanisms involved in this adaptation are the same that act upon the creation of L1 phonetic categories. In fact, both L1 and L2 categories are understood to coexist in a single phonological space, which means that contact with non-native sounds not only can result in the generation of new categories but will also affect already existing ones, which adapt and change as does the interlanguage phonology.

The model stemmed from the research that Flege had done on the way L2 learning was affected by age and the length of residence (LOR) in an L2 environment. What he found was that, in contrast to the idea of the critical period for language acquisition (Lenneberg, 1967), age seemed to have a gradual effect on the ability to learn a new language, and that perceptual plasticity may be diminished but does not disappear (Flege, Mackay, and Meador, 1999; Flege and Mackay, 2011).

Predictions

As more different variations of sounds perceived to be speech are heard, including those coming from non-native systems, they can either fall within the boundaries of an existing category, in which case they will be perceived as a member of that category; or fall outside these boundaries. If the latter happens, some of the more proximal L1 surrounding categories can be expanded to include the new sounds, or new phonetic categories can be generated if a sufficient number of similar variations are encountered. However, this process depends on whether or not they are perceived to be sufficiently distinct from other existing categories. If listeners are unable to hear the difference, they will will be unable to generate a new category to contain them. In this sense, the predictions of SLM are similar to those of PAM.

1.1.3 Native Language Magnet

The Native Language Magnet Model (NLM; Kuhl, Williams, et al., 1992; Kuhl, 2000; later *expanded* into NLM-e; Kuhl, Conboy, et al., 2008) aims at providing an altogether different explanation of the mechanisms involved in language acquisition throughout the lifespan of the individual, from infancy to adulthood. Between 6 and 9 months of age, infants transition from being universal detectors (Eimas et al., 1971) to being language specific perceivers (Werker and Tees, 1984), and they do this by making use of innate abilities for pattern recognition and statistical learning. By being exposed to their ambient native language, children develop a language-specific perceptual filter which distorts their perceptual space in order to highlight the contrasts between the emerging native contrasts. The filter acts by enhancing sensitivity to those features that are relevant and attenuating that of those which are not. The emerging categories are centred around phonetic prototypes, which are mental representations of the native sounds.

In this sense, NLM-e (and its predecessor NLM) is radically opposed to the direct-realist framework that supports PAM, in that NLM-e starts from the idea that speech is represented auditorily in the brain prior to any motor or gestural category being developed. Whatever categories do exist emerge from those mental representations and the effects they have on the perceptual space by means of their perceptual magnet effect. And this "neural commitment" is posited as the likely cause for the effects of the so-called critical period and the general decline in perceptual plasticity. This effect is related to the more general notion of categorical perception, but it is understood to be pre-categorical and distinct from it: while categorical perception has been shown to exist in non-human animals (*e.g.* Kuhl and Miller, 1975, 1978, among others) even for prosodic features (Ramus et al., 2000), only humans have shown signs of perceptual magnets (Kuhl, 1991).

Predictions

Despite this, NLM-e makes similar predictions to the types of contrasts identified by PAM. With one exception: since phonetic categories for NLM-e exist in a perceptual space, the model predicts better discriminability for sounds that map to prototypes that are spatially more distant from each other. However, it is not immediately clear how this distance can be measured across multiple contrasts (Kingston, 2003).

1.1.4 Model comparison

Despite their differences, the three models presented so far have some aspects in which they are in agreement, such as the idea that the mechanisms available in childhood to generate phonological categories remain available even into adulthood, and that L1 and L2 categories coexist in a common *interlanguage* phonological (or perceptual) space which accounts for the bidirectional modifications between the two.

Another aspect which raises few concerns is what is referred to as *category assimilation* in PAM and *equivalence* in SLM. Namely, that when L2 listeners are confronted with novel contrasts, they can be perceived as either good or bad exemplars of existing categories, or as something altogether different which cannot be considered to be a member of any category (assuming they were perceived to be speech sounds at all). This is true for SLM and PAM, since both models deal with perception in a categorical fashion (even if they understand the reality of those categories in radically different ways). And while NLM-e understands perception to be pre-categorical, the effects on the perceptual space that are caused by the phonetic prototypes have a similar effect, in that sounds that fall too close to an existing prototype will be perceived as (better or worse) instances of it, while those that are sufficiently distinct will escape the prototype's attraction and could eventually lead to the generation of new prototypes.

The models differ on their underlying understanding of the mechanisms involved in perception: PAM considers perception from a direct-realist perspective, and understands it to be gesture-based; SLM considers phonetic categories to be mental representations of sounds, not gestures; and NLM-e claims phonetic categories exist in a multidimensional perceptual space, in which each dimension is a relevant phonetic property, and are the result of the distortion that phonetic prototypes have on this space. But they all make similar predictions for the discriminability of non-native contrasts. Because of this, the labels for the types of contrasts initially identified by PAM will be used as shorthand to refer to these mappings, even when not talking about PAM in particular.

1.1.5 From segments to suprasegmentals

There is a general lack of models of perception that explicitly deal with the issues of suprasegmental perception. Recently, a study with learners of tonal languages whose native language were non-tonal (English and French), found that lexical tones were assimilated (in the sense of PAM) to what they refer to categories in the non-native speakers' L1 prosodic system. This was taken as evidence supporting the applicability of at least PAM to the perception of suprasegmentals (So and Best, 2008, 2011).

However, the key question is whether there would be any reason to say that the learning and development of segmental and suprasegmental categories should be different. This is vaguely hinted at in So and Best (2008, 2011), where the authors refer to intonational categories as *i-Categories*, implying that they would be in some way different to other non-intonational categories. But no evidence for this is presented. On the other hand, Dupoux, Pallier, et al. (1997) and Dupoux, Sebastián-Gallés, et al. (2008) report what they refer to as a "stress deafness" effect on French speakers when perceiving the position of the Spanish lexical stress, which they take to be compelling evidence for suprasegmental perception also being subject to the same transfer effects that have been thoroughly researched for the perception of segmental contrasts.

From a purely theoretical perspective, it would be surprising to find a difference between the two: the phonetic categories present in SLM have an acoustic basis, and suprasegmentals evidently function as such, contributing acoustic information; and at least after the modifications introduced to accommodate evidence from L2 learners, PAM-L2, argues that new categories are generated at the phonological level, and it is clear that there are phonological functions encoded in prosody.

On the other hand, evidence from word recognition suggests that judgements based on prosody are slower and more prone to errors (Cutler and Chen, 1997; Cutler, Dahan, and Donselaar, 1997). And there is also evidence that pitch is processed in the brain in ways that are different from segments: Dupoux, Sebastián-Gallés, et al. (2008) states that the particular acoustic properties of suprasegmental cues (which are "slow and large-scale") make them a likely candidate for processing in the right hemisphere. However, according to the "attraction hypothesis" (Shipley-Brown et al., 1988), although pitch processing would rely on the non-linguistic functions of the right hemisphere, its lateralisation would not be total, and would depend on the degree to which it is perceived by the listener to be linguistic in nature, which would "draw" it towards the left hemisphere (Gandour et al., 2000). In this way, even if there is evidence that pitch in particular is processed differently in the brain, segmental and suprasegmental information is "unlikely to be processed fully independently" (p 213).

Trofimovich and Baker (2006) attempted to answer the question of the difference between segmental and suprasegmental categories experimentally from the perspective of language learning, and found that they behaved similarly, with both being subject to the same gradual learning process and depended on the specific aspect that was being studied. And despite their peculiar terminology, evidence from So and Best (2008, 2011) and their suggestions for an extension of the tenets of PAM also ultimately support the idea that both types of categories are similar.

1.2 Prosodic models and annotation

In order to make cross-language comparisons in Spanish and Japanese, a common theoretical framework and tone labelling scheme was necessary. The ToBI (Tone and Break Indices) labelling standard (Silverman et al., 1992) has been extensively used in descriptions of intonation in both Japanese (*e.g.* Pierrehumbert and Beckman, 1988; Venditti, 1995, 2005) and Spanish (*e.g.* Beckman, Díaz-Campos, et al., 2002; Sosa, 2003; Estebas Vilaplana and Prieto Vives, 2008), and therefore presents itself as the obvious choice in this case. ToBI itself is framed on the Autosegmental-Metric (AM) theory (Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986; Pierrehumbert and Beckman, 1988), which assumes that tones reside in a "tier" separate from segments (following the ideas in Goldsmith, 1976), and align with stressed syllables from the metric tier.

ToBI inherits this multi-tiered approach by recognising two main tiers: one for a two-level system of tones, which are marked as high (H) or low (L); and a different tier for "breaks and indices" (Price et al., 1991) marking the rhythmic patterns of speech and the non-tonal grouping of words. A third "miscellaneous" tier is used for annotating additional non-linguistic information such as disfluencies or hesitations. Tones that are aligned with a stressed syllable are marked with a star (H becomes H*, L becomes L*), and accents can be combined via a plus (+) sign to annotate multitonal accents. Additional characters in the ToBI inventory allow for the special marking of other features, although the only two that will become relevant in this dissertation are pitch downtrend (¡) and delayed alignment (>).

There are a number of alternative models of intonation that have been developed and could have been chosen for the description of tone in this case instead of ToBI. Of particular relevance are the Fujisaki model (Fujisaki and Hirose, 1984), which has had a large impact on the discussion of Japanese intonation particularly among Japanese researchers, and the Parallel Encoding and Tonal Approximation (PENTA) model (Xu, 2004), which although considerably more recent has also had considerable success (see for example Lee, 2015). However, as stated above, a key concern for the discussion of the results in this dissertation was to find a framework whose applicability to both languages was well established. And in that regard, despite the possible merits that both the Fujisaki or the PENTA model might have, the position of TOBI is uncontested.

1.3 Lexical prominence

Definition of prominence

This thesis is concerned with one specific aspect of L2 perception: that of lexical prominence. However, the precise definition of lexical prominence deserves some attention, since it has been used as an umbrella term that has meant remarkably different things depending on the different perspectives from where it has been studied. Wagner et al. (2015) presents a recent discussion of this scenario, and identifies three predominant perspectives in the study of prosodic prominence:

1 The *functional* perspective, which "focuses on communicative, often core linguistic functions of prominence" and "lends itself to a categorical classification, with a particular type of functional prominence being either present or not". Studies that consider emphatic or so-called *emotional* speech as prominent are also understood to work from this perspective.

2 The *physical* perspective, which focuses on the aspects of the acoustic signal that are recognised as "physical correlates of prominence" and often "relies on on individual rather than a large set of possible signal correlates, focusing on the interplay and interchangeability of correlates in particular contexts". Prominence will not be understood as categorical but as a continuous phenomenon in studies framed entirely on this perspective.

3 The *cognitive* perspective, which "focuses on perceptual processing" and the interaction between "low-level neural pathways and psycho-acoustic processing mechanisms that contribute to higher-level cognitive processing".

In these terms, this thesis will deal with prominence from a point that is halfway between the functional and the physical perspectives, since it understands lexical prominence as a categorically perceived property of the word, but bases that categorisation primarily on the acoustic aspects of the signal, and not necessarily on communicative or paralinguistic aspects.

Explicit definition

Taken their suggestion for an explicit definition, then, this dissertation will understand a syllable as prosodically prominent when it is perceived to be particularly marked in a language-dependent and contrastive way, by virtue of a specific and also language-dependent behaviour of acoustic cues. In more simple terms, the notion of lexical prominence will be understood to perfectly align with that of *lexical stress* when referring to Spanish, and that of *lexical accent* when referring to Japanese. The following sections will provide an overview of what exactly those "language-dependent behaviours" are.

1.3.1 Spanish

Predictability and distribution of lexical prominence

Spanish has traditionally been considered to be a "stress-accent" language, in which all content words have a lexically determined stress that can occur on any of the last three syllables of the word. According to the positions of the stress, words are said to be *agudas* (*i.e.* accented in the ultima, or oxytone); *graves*[2] (*i.e.* accented in the penult, or paroxytone); and *esdrújulas* (*i.e.* accented on the antepenult, or proparoxytone). Word stress in Spanish serves a number of functions: it is contrastive at both a paradigmatic and a syntagmatic level, distinguishing between stressed and unstressed syllables, and between words in minimal sets (Quilis, 1983). Indeed, there is a large number of minimal trios and pairs that exist in Spanish, a product of the fact that a significant part of the information relative to verbal conjugation is encoded in the position of the stress: as an example, stress position is the only difference between the forms for the first-person present tense and the second-person past tense.

The fact that Spanish stress is so heavily determined both by verb conjugation and more generally by the lexical inventory has made some scholars (*e.g.* Face, 2006) argue that Spanish is a language with a highly predictable stress placement. Particularly in the case of regular verbs, contextual information is often enough for a proficient speaker to completely

2 Words of this type are sometimes called *llanas*, particularly in the Spanish peninsula.

determine the position of the stress based on information on the person, tense, mood and number. This is not the case for irregular verbs, but regular verbs account for a large majority of the lexicon, with Velásquez, Gelbukh, and Sidorov reporting that only 15% of verbs have some degree of irregularity (2002).

However, this predictability is rule-based, since in order to use contextual information to make predictions on the placement of stress in verbs speakers needs to be aware of the rules that govern its distribution. This prior knowledge of the rules is of course a given for native speakers, but it is not clear how accessible these rules are for non-native speakers, or how easy they are to develop. The lack of a demarcative function of Spanish stress (Quilis, 1983), common in languages with relatively fixed stress positions like English (at the beginning of the word) and French (at the end), also supports this (Hidalgo Navarro and Quilis Merín, 2004).

A different kind of predictability is afforded by the fact that some stress positions are much more common than the rest. Quilis reports the results of an early review of Madrid Spanish using a corpus of 20361 words from unstructured and secret interviews (in which the interviewees were not aware they were being recorded). After discarding unstressed words, stressed monosyllabic words, and adverbs ending in -*mente* (which tend to have a type of secondary stress, Ortiz Lira, 2003, see), showed that paroxytone words were four times as common as words with other stress patterns, accounting for just under 80% of a total of 9219.

For nouns, the frequency of different stress positions also depends on the word ending. Table 1.1 (adapted from Face, 2006) shows the distribution of the position of the stressed syllable depending on the final segment of the word. Predictions based on this information is not as accurate as that for regular verbs, but according to Face, it allows for correct predictions on 87% of cases (p 1240). Likewise, Aske (1990) reports a large majority of non-verbs ending in [n], with the exception of those in which the preceding vowel is [e], have final stress.

Unlike the case of verbs, predictability in this case is based on word

Word ending	Oxytone	Paroxytone	Proparoxytone
Vowel	178	2494	178
/s/	20	909	94
Consonant (except /s/)	778	176	2
Total	976	3579	274

Table 1.1: Distribution of stress placement in the 4829 most frequent polysyllabic words of spanish (from Face, 2006).

frequency. But even in this regard, the ability to use this information will depend on the speaker's knowledge of the lexicon, and their level of awareness about stress placement in the first place, both of which are not aspects that can be assumed for non-natives.

Acoustic cues of stress

There are words in Spanish that are not accented, but the kind of words that can be unaccented is limited to function words and clitic pronouns, used commonly in Spanish. Other exceptions exist, such as words that are the first item in some word compounds (*e.g.* tres mil [en. *three thousand*] > /tɾesˈmil/), but these are exceptional cases.

In words that do have prominence, the unit that bears it is unarguably the syllable, and prominent syllables are traditionally described as being longer, louder, and higher in pitch than those that are not. However, the relevance assigned to each has varied widely in different studies.

Early studies on the subject considered the Spanish stress to be an "intensity accent", and primarily correlated with increases in syllable loudness (Cuervo, 1907; Navarro Tomás, 1918). Although this idea has since become widely spread in layman circles, recent experimental work has largely proven it to be false (*cf.* Urrutia Cardenas, 2007).

F_0 as primary cue

Starting with the work of Enríquez, Casado, and Santos (1989), which used synthetic speech to manipulate individual acoustic cues and look at the effects they had on perception, there has a been a growing body of work in support the fundamental frequency as the fundamental acoustic correlate of the Spanish lexical stress. This early work used entirely synthetic stimuli, and manipulated acoustic cues one at a time. This approach made it impossible to look at the interaction between different cues, and it is not entirely clear how natural the resulting stimuli were, nor how it might have impacted their results. But they not only managed to show that F0 had the most constant effect (regardless of testing condition) and was the least subject to listener biases, but they also found that the effectivity of the remaining cues depended, among other things, on the accent pattern in the word. Their results also lead them to suggest that increases in intensity might actually make it more difficult to perceive the position of the stress, but as stated above, this is likely the result of artifacts in their synthesis process.

A similar approach was followed much later by (Llisterri et al., 2003, 2005), but in this case the stimuli were not entirely synthetic. Instead, a perceptual study was carried out using stimuli that were manipulated under a prosodic transplantation paradigm: the stimuli were created by replacing the acoustic features of one word by those of another in which the stress was placed on a different syllable. Importantly, they did thee manipulations also in the combination of different cues, providing the first objective insight into the interaction between the action of these cues.

Like Enríquez, Casado, and Santos (1989), they used real and nonce minimal trios, and showed that modifying syllable duration and intensity, either on their own or in unison, did not result in significant changes in the perception of which syllable was stressed. The manipulation of F0 alone was also not enough to elicit a systematic change in the perception of stress position, but a combination of this cue and one of the other cues, or of all three cues at the same time, was sufficient to alter participants' perception of stress position (Llisterri et al., 2005, p 295).

The idea that F0 served as the primary mark of Spanish lexical stress had already been put forth in 1847 by visionary Bello in his *Gramática*, but these studies (and others since then) have provided empirical evidence that this is the case.

Secondary acoustic cues

These findings, while robust, left questions open regarding the action of the so-called *secondary cues* of the Spanish lexical prominence: duration and intensity. Among the studies that have looked at them specifically, the work by Ortega-Llebaria and colleagues (Ortega-Llebaria, 2006; Ortega-Llebaria and Prieto, 2007, 2009, 2011; Ortega-Llebaria, Prieto, and Vanrell, 2008) deserves special mention.

They considered that the effect of pitch concealed the combination of two separate entities, which they distinguished as sentence accent and of lexical stress. According to them, most of the action of F_0 as a cue encoded information on sentence accent, and the action of duration and intensity on the encoding of lexical stress would only show in contexts in which the action of pitch as an acoustic cue was neutralised. They did this by comparing statements and reported clauses, the latter of which are produced in Spanish with a constant low F_0 with a slow downtrend (Ortega-Llebaria and Prieto, 2011, p 78). The absence of sentence-level intonational movements, they reasoned, would allow them to perform a more detailed examination of the cues involved in lexical stress. To this end, they not only examined F_0 but also duration, intensity, and vowel quality measured directly and through two different methods of calculating spectral tilt: a more traditional approach resulting from the difference of the intensities of the first and second formants of the measured vowel, and a different method described in Fulop, Kari, and Ladefoged (1998) which would allow them to control for the effects of vowel reduction.

Their results confirmed that vowel reduction was not involved in the production of Spanish stress (p 91), and that stressed syllables were produced with larger durations than unstressed ones. When looking at their results for overall intensity, they found that stressed syllables were produced with systematically greater intensity values only in reported statements, which meant that duration was more consistently used for marking stress during production. However, they found no correlate between stress and spectral tilt, and perceptual tests using stimuli with modified duration and spectral tilt showed that, between these, Spanish speakers relied solely on duration for their perception of stress (Ortega-Llebaria, Prieto, and Vanrell, 2008).

Although these results are enlightening and sound, some of the conclusions that have been derived from them, either by the researchers themselves or by other researchers (*e.g.* Kim, 2015), do not necessarily follow. It is undeniable that the shape of the F_0 contour in Spanish is heavily determined by the action of sentence accent, and that neutralising this difference is necessary to examine the effects of duration and intensity on the encoding of lexical stress. It is also undeniable that the Spanish lexical stress is the result of the joint action of a number of different cues, and in this there is agreement in almost any study on the nature of Spanish stress (*e.g.* Enríquez, Casado, and Santos, 1989; Urrutia Cardenas, 2007; Llisterri et al., 2003, 2005; Ruiz Mella and Pereira, 2010). But from this it does not follow to say that what F_0 encodes is exclusively related to sentence accent, or similarly, that the true acoustic correlates of lexical stress are duration and intensity.

Since pitch has been shown systematically to have a primary role in the perception of Spanish stress, some attention must be paid to the behaviour of the F_0 curve and the way in which it is influenced by the context in which the accentual phrase is contained. Peak displacement

Traditionally, it has been stated that the Spanish stress correlates acoustically with a peak in the F_0 curve aligned with the stressed syllable. This is certainly true for *e.g.* words in isolation, but when inserted into longer phrases it is common to find delayed peaks which align with the syllable following that which bears the stress. Indeed, in these cases this phenomenon is extremely common, with Llisterri, Marín, et al. (1995) and Garrido et al. (1993) both independently finding peak displacement in approximately 80% of their corpora. These studies looked also at the contexts in which it was most likely to occur, and they found that it occurred less commonly before pauses, and never across the boundary between a noun phrase and a verbal phrase. This is in agreement with Prieto, Van Santen, and Hirschberg (1995), who found that, in addition to the surrounding prosodic structure, segmental duration had a large effect on peak displacement, with longer vowels and onset durations correlating with longer peak

Tone	Accent type					
	Proparoxytone		Paroxytone		Oxytone	
L*+H	33.55	33.80	49.30	41.90	23.95	33.00
L+H*	48.75	33.25	34.40	26.10	66.55	48.15
H*+L	7.20	14.85	1.95	14.65	6.80	6.90
H+L*	10.50	18.15	14.35	18.40	2.75	12.05

Table 1.2: Distribution of realisations of stress per word accent type and sentence mode. Values on the left are for declarative sentences and those on the right for interrogative sentences. Adapted from results of two different corpora from Fernández Planas and Martínez Celdrán (2003).

delays, except when they were caused by prepausal lengthening.

Fernández Planas and Martínez Celdrán (2003) reports the results of a study on the interactions between duration and pitch as correlates of Spanish stress, which used an automatic signal-based procedure for assigning tone labels derived from the AM model. Their analysis, which considered both declarative and interrogative utterances, identified 12 possible tone configurations which they argued could be understood to be surface representations of 4 underlying tones: L*+H, which shows the effect of peak displacement; L+H*; H*+L; and H+L*.

Table 1.2 presents an overview of their findings, and clearly shows that, regardless of whether the utterance is interrogative or declarative, there is a preference for L*+H and L+H* accents, with the latter being more common for proparoxytone and oxytone words, and the former occurring more frequently in paroxytone words. The table also shows how prevalent the occurrence of structures identified to have peak displacement (L*+H) were, even in those cases when it was not the most frequent.

Non-acoustic cues of stress

There is some additional evidence suggesting that stress perception in Spanish is affected by other non-acoustic cues, such as the uneven distribution of stress positions in the lexicon which was shown in table 1.1. Face (2006) reports the results of a series of tests to examine what he calls the *cognitive factors* (*i.e.* non-acoustic cues) involved in Spanish stress perception. For his studies, he used nonce words to review the effects that four different non-acoustic cues had on stress perception, focusing on the effects of *a*) similar words in the lexicon; *b*) syllable weight; *c*) lexical subregularities (*i.e.* the unequal stress distribution mentioned above); and *d*) the morphological category of the word in question. He concludes that there is a significant effect of lexical similarities and subregularities, which is interpreted as meaning two things: that information from more than just the final segment of the word is taken into consideration; and that words in the lexicon will influence the perception of stress position in other words that are perceived to be similar, in a way that is proportional to the degree of similarity. An effect of morphological category was also found to be significant, so that perception of the stress in the nonce word depended on whether it was contextualised as a verb (in which case it tended to be perceived as paroxytone), or a noun (which were perceived predominantly to be final-accented). There was no effect of syllable weight.

Chilean Spanish

The variety of Spanish used in this dissertation is that spoken in Chile, and more specifically in the central region of the country, surrounding the capital city of Santiago. Most of the studies mentioned so far have dealt with the particulars of Peninsular Spanish in general, or with what could be characterised as "standard Spanish". There are a number of ways in which Chilean Spanish differs from the varieties studied in the research mentioned so far, and this section will cover those that were considered to be relevant to the theme of this dissertation.

Relatively little accentual variation in Spanish

Spanish is the second most widely spoken language in the world, spoken by an estimated 399 million people spread over 31 countries (Lewis, Simons, and Fenning, 2015). Among the many different varieties of Spanish, there are some for which there is considerable variation in the stress systems. Some examples include:

1 Rioplatense Spanish (spoken in the area surrounding Buenos Aires), has a very characteristic and markedly non-traditional verbal *voseo*, in which the position of the accent in some verbal conjugations is shifted.

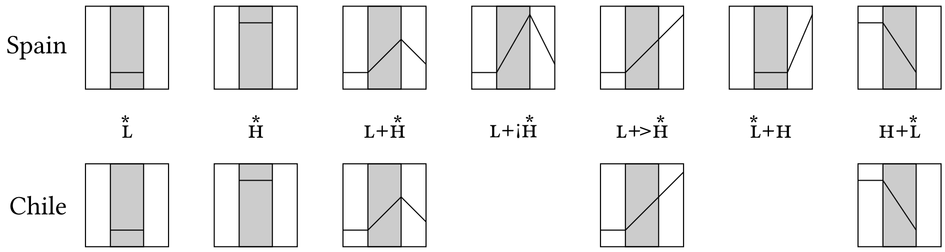


Figure 1.1: Comparison of monotonal and bitonal pitch accent inventories in Castillian Spanish (top) and Chilean Spanish (bottom). Reproduced with data from Estebas-Vilaplana and Prieto (2010) and Ortiz Lira, Fuentes, and Astruc (2010) respectively. The grey area represents the stressed syllable.

2 The Spanish spoken in the Argentine province of Córdoba, known as as *tonada cordobesa*, has a very characteristic pre-tonic lengthening, not found in other varieties.

3 Some varieties spoken in parts of Argentina and the Peruvian Amazon (see Garcia, 2011; García, 2014) have non-standard tri-tonal structures, which although not unheard of, are uncommon in other varieties.

However, particularly considering the number of speakers and the spread of the language, there is comparatively little variation in the stress systems in Spanish, and the differences in most varieties concentrate on the lexicon and the syntax.

Chilean Spanish is an example of this. Focusing exclusively on differences in the stress system, a comparison of the pitch accent inventories reported for the Castillian (Estebas-Vilaplana and Prieto, 2010) and Chilean (Ortiz Lira, Fuentes, and Astruc, 2010, only for Santiago) varieties of Spanish shows that most of the inventory is shared, and that the inventory of the Chilean variety (considering both monotonal and bitonal accents) are a subset of the inventory for the more well-studied Castillian variety. Figure 1.1 reproduces the pitch accent inventories reported in these studies for both varieties, and clearly shows the extent of the overlap. This overlap was also evident for boundary tones.

There is some evidence that may suggest a different rate of occurrence of peak displacement in the Spanish spoken in Chile in general and Santiago in particular. In a study of spontaneous recordings of Santiago Spanish, Ortiz Lira (2003) found a predominance of H+L* pre-nuclear pitch accents, mentioning that L*+H accents in Chile "contribute greatly to the perception of speakers as foreigners" (p 4). However, the study by Ortiz Lira examined pre-nuclear accents, and no results are presented for nuclear accents. Likewise, a study on the rising accents of Chilean Spanish reported in Toledo and Astruc (2008) did find some L*+H accents, but only at the end of intermediate intonational phrases, meaning that they are more than likely the result of an interaction with sentence intonation. Although none of these studies provided a review like the ones found in Llisterri, Marín, et al. (1995) and Garrido et al. (1993), their findings do not suggest a rate of occurrence quite as high as the reported 80%. A review focused specifically on this possible difference is lacking, and would be welcome.

1.3.2 Japanese

Japanese is the 9th most widely spoken language in the world, with an estimate 128 million speakers worldwide (Lewis, Simons, and Fenning, 2015), but unlike Spanish, native speakers of Japanese are primarily concentrated in Japan alone. Despite the relatively small geographic scale of the language there are a large number of regional varieties, with a considerable part of the variation concerning the accent system. For the purposes of this dissertation, only the regional variety spoken in the Kantō region surrounding Tōkyō will be relevant, and when references are made to Japanese without otherwise noting a specific variety, the term will refer to the Japanese spoken in this region.

This variety, sometimes called Standard Japanese, has a mora-based "pitch-accent" system (Tsujimura, 2007) in which each mora can have either a H or a L tone (Pierrehumbert, 1980), and accents are marked by a H*+L tone following the accented (*i.e.* prominent) mora.

Syllables and moras

The mora — succinctly defined by McCawley (1978, p 114) as "something of which a long syllable consists of two and a short syllable of one" — is a metric unit which sits in a level between the syllable and individual phones. There is considerable consensus in the literature regarding the particularly relevant function the mora has to play in Japanese phonology and its metric system. There is strong evidence from speech errors (Kubozono, 1989) and language games (Katada, 1990) that suggests the mora is involved in word production, and it is also the unit more commonly used for phrasing in popular song (Vance, 2008, see), with each mora normally being assigned to a different beat[3], and in traditional poetry, which restricts the number of moras that can appear on each verse. The relevance of the mora in Japanese is also ingrained in their writing system: characters in the *hiragana* and *katakana* alphabets represent moras instead of syllables or phones, such that written in *hiragana* the place names for *e.g.* Kansai (かんさい), Ōsaka (おおさか) and Hiroshima (ひろしま) all have four characters because they all have the same number of moras (/ka.N.sa.i/, /o.o.sa.ka/, /hi.ro.ši.ma/), despite the fact that they have 2, 3 and 4 syllables respectively (/kaN.sai/, /oo.sa.ka/, /hi.ro.ši.ma/).

The traditional account of Japanese phonology, which follows the tradition of the seminal work by Pierrehumbert (1980), uses the mora as the tone bearing unit (TBU) in Japanese and models changes in intonation by limiting tone changes to occur across mora boundaries. A distinction is made between regular moras and so-called "deficient" moras, in that a phonological constraint against the placement of the accent on a mora of the latter type (the NADM principle; Labrune, 2012). Regular moras include V, CV, and CyV, in which y is a glide which often causes palatalisation in the consonant; while deficient moras include the nasal coda N (*cf.* Vance, 2008, pp 104-105), the mora preceding a geminate consonant (Q), and the second mora in a long vowel (R). Each syllable has only one non-deficient mora, and therefore only one position in which the accent might take place. This is true both for monomoraic syllables of the form V or C(y)V and for bi-moraic (*i.e.* C(y)V[RNQ]) and tri-moraic (*i.e.* C(y)VR[NQ]) syllables.

One accent position per syllable

3 One of the Japanese words used commonly for mora is, in fact, *haku* (箔) which can also be used in the same sense as English *beat*.

Overview of Japanese accent

Acoustic cues

The main acoustic difference between H and L moras resides in the relative F_0 of the signal. Although no duration differences were found, Cutler and Otake (1994) reports significant acoustic differences in F_0 and intensity range and standard deviation, and found these to interact with word position within the utterance. However, their results do not seem to point to any systematic variation, and it is not clear that speakers use these for perception (Sugitō, 1982; Vance, 1995).

Unaccented words

Japanese words can be accented or unaccented, and this is lexically determined. This means that an n-syllable word has up to $n+1$ possible accent patterns (one for each syllable and one extra pattern for unaccented words), although in practice only some of these will result in actual words.

Default position

Different processes seem to operate on the attribution of accent for words of different historical origins, with a large majority of native Japanese words being unaccented (71%), but only a small number of loanwords having the same pattern (7%; Kubozono, 2006). In the case of Sino-Japanese words, the proportion of accented and unaccented words is much more even, with Sibata (1994) and Kubozono (2006) both independently reporting that unaccented words account for just over half of trimoraic words (52.6%, and 51% respectively).

Most accented words follow the "antepenultimate rule" (McCawley, 1968), and have their accent on the antepenultimate mora or on the next non-deficient mora to the left, since accent placement in Japanese is attributed from the end of the word. There is also ample evidence supporting the fact that accent attribution is done on the basis of counting moras rather than syllables (Labrune, 2012), so even if the mora is considered the TBU (and this is a controversial issue, see *e.g.* Lee, 2015), the default position for the accent is not syllable-bound.

Distribution within the word

Tōkyō Japanese imposes some constraints to the distribution of H and L moras in a word[4]. First, any given word can have only one accent, which

4 Other varieties of Japanese (such as that spoken in Kagoshima) have wildly varying accent systems, some for which there is strong evidence of a syllable-based accent system.

means that there can be only one HL downshift per word. Second, by effect of the initial lowering rule (Haraguchi, 1977), the first and the second mora must have different tones. This means that the first mora in the word will be H only in the case in which that mora is accented.

Final-position neutralisation

One of the consequences of the accent being marked as a pitch fall following the accented unit is that, when this unit is in a final position, what marks the accent occurs after the word has ended. This means that when presented in isolation, there is no discernible difference between final-accented and unaccented words (Sugitō, 1982). Some experiments have suggested that this is not the case, and that unaccented H tones reach lower frequencies than those in accented words (*e.g.* Vance, 1995). It is not clear if this variation is the result of systematic variation, or simply idiolectal differences (see Labrune, 2012, for a review), and previous studies have also revealed a general tendency for H tones to be raised preceding a L (Hyman, 2007).

Regardless of what is the source of this variation, the fact that a phonetic contrast exists does not mean that it is used by speakers as a tool for discrimination. And there is strong evidence that this phonetic difference is not perceptually relevant even when it is found (Sugitō, 1982; Vance, 1995). Systematic discrimination between final-accented and unaccented words is only achieved in the presence of a disambiguating additional unit which can show that the fall took place: either a deficient mora (since these cannot bear the accent) or a grammatical particle. Figure 1.2, taken from Vance (2008), shows all possible accent patterns in a set of 4-mora words both in isolation and followed by the topicaliser particle /wa/ は.

Prosodic interactions of the Japanese accent

Since pitch is the only systematic cue used for the recognition of accent placement in Japanese, the realisation of the accent has become particularly robust against interactions with *e.g.* utterance-final sentence tonemes. Figure 1.3 shows the F_0 contours of two japanese utterances as reported by Vance (2008). They both have the same sequence of sounds (/sorewa

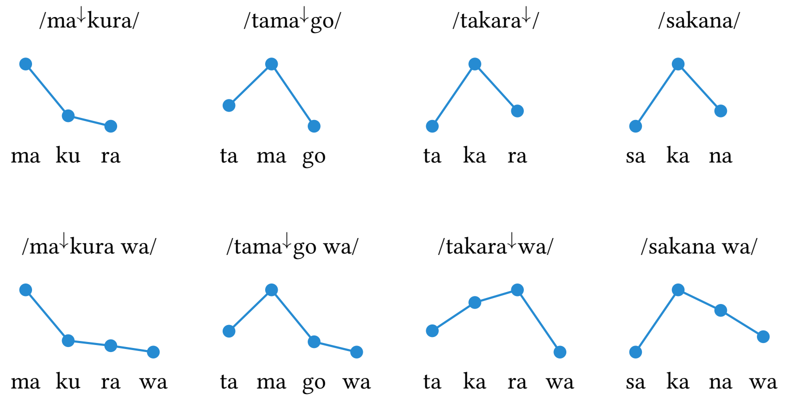


Figure 1.2: Intonation patterns on isolated words (top) and topic phrases (bottom). The contrast between final-accented words (*e.g.* /takara↓/) and unaccented words (*e.g.* /sakana/) is lost in isolation (top row), but adding an additional syllable makes the difference clear. Based on figures from Vance (2008).

banana/, accented in "ba") but one of them is an interrogation (*is that a banana?*) and the other a declaration (*that is a banana*). On both sentences there is an F_0 rise that leads to a peak at the end of the accented mora (with a small intervening dip resulting from the /b/ in /ba/) and a sharp fall during the following mora, which is indeed what marks the accent. In this example, the only noticeable difference between the two is the sharp rise at the end of the utterance marking it as an interrogation, but this rise is entirely contained during the second half of this final syllable, leaving the contour of the word up to the beginning of the interrogative toneme largely unchanged. In order to accommodate for this, the final syllable is roughly twice as long as that of the final mora in the statement.

Interaction with particles

Most particles behave like the topicaliser /wa/ in the sense that they do not change the accent of the preceding word. This list includes /ga/ が, /o/ を, /ni/ に, /de/ で, /e/ へ, /to/ と, /mo/ も, /ya/ や, /kara/ から, and the plain copula /da/ だ in phrases of the form N+/da/. (Vance, 2008).

However, the possessive particle /no/ の behaves differently when modifying a final-accented polysyllabic word, in which case the compound often (not always) becomes unaccented (Vance, 2008, p 156). This accent change seems to be more infrequent when the modified word ends in a long syllable.

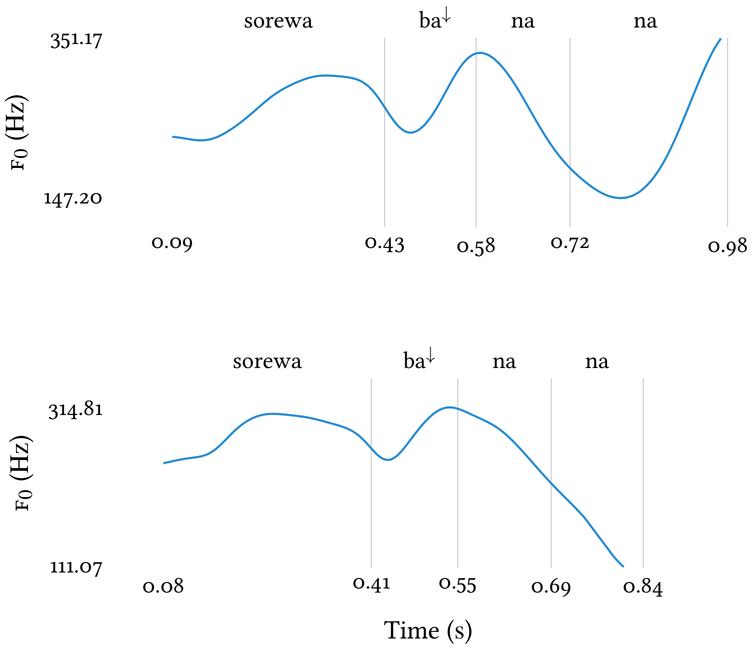


Figure 1.3: Stylised F0 contours from a Japanese interrogative (top) and declarative (bottom) sentence. values on both axes are comparable. curves based on Vance (2008).

There is a different pattern that appears with most polysyllabic particles. Those in this category seem to have an accent of their own on the first mora, but which only manifests when the particle modifies an unaccented word. Otherwise, the accent of the preceding word remains unaffected. Particles that show this behaviour are /desu/ です, /dešita/ でした, /bakari/ ばかり, /demo/ でも, /made/ まで, /nomi/ のみ, /sae/ さえ, /yori/ より and /šika/ しか (Vance, 2008).

Linguistic relevance

Accent in Japanese is not necessarily as relevant for discrimination as it is in *e.g.* Spanish. In a comparative study of the contrastiveness of Chinese, English, and Japanese lexical prominence, Sibata and Shibata (1990) found that 13.57% of observed minimal pairs were distinguished by differences in accent position, and that word-accent accounts for 19.72% of the discrimination in the group of 2-mora combinations with 3 differing accent patterns (p 323). This has led some researchers to question the relevance of the Japanese accent in discrimination: Labrune (2012) claims that "the importance of the distinctive function should not be exaggerated" (p 181), and Vance (2008) says that the accent opposition is basically one of accented *v.* unaccented, with a large majority of accented words having their accents in the default position, and only a few exceptions which are of frequent use and therefore easy to memorise (p 155).

However, Sibata and Shibata state that

> arguing whether or not the [Japanese] word-accent has a distinctive function would be fruitless. It would be best, instead, to think about it in terms of how much, or to what degree, does the accent have — or lack — this distinctive function. In other words, to approach the subject from a quantitative perspective. (p 318, personal translation)

Furthermore, there is evidence supporting the idea that prosodic features play a considerable role in speech perception by natives. A series of experiments reported in Minematsu and Hirose (1995) found significantly different recognition times for isolated words presented with correct and incorrect accent placement. And when presented with progressively larger fragments of 4-mora words, native speakers were able to recognise initial-accented words after listening to significantly smaller fragments than those needed to recognise words with the accent on the second mora or words that had no accent.

Likewise, Cutler and Otake ran studies on the perception of Japanese accent by native speakers of Japanese who came from Tōkyō (Cutler and Otake, 1994) and from a region of Japan with an unaccented variety (Otake and Cutler, 1999). In these experiments, participants were presented with fragments of the first mora of words in which that mora could be H or L, and asked whether they thought the mora came from a word with a high first mora or from one in which the first mora is low. Surprisingly, participants in both studies had comparable and very high rates of overall correct responses (74% for listeners from Tōkyō, and 74.2% for speakers of the unaccented variety).

Japanese stimuli in this dissertation

Words used in this dissertation as stimuli in Japanese will be composed exclusively of monomoraic CV syllables, making the discussion regarding the TBU in Japanese irrelevant for the topic at hand. No more will be said of the different roles of moras and syllables in Japanese phonology, and in particular regarding the accent. This did not impose a serious limitation in word choice, since most syllables in Japanese are indeed monomoraic, as reported by Otake, Hatano, et al. (1993), with more than 70% of the syllables in their corpus being CV.

The distinction between the terms for moras and syllables will be maintained throughout the text, however, so even if they will most commonly be referring to the same unit, they should not be interpreted to have the same meaning. To make it possible to use the same language when discussing the Japanese and the Spanish data, "final accented" should be understood as meaning "with the accent on the final syllable" unless otherwise stated.

1.4 Second language acquisition

Added to the relative shortage of studies in suprasegmental L2 perception, there are surprisingly few studies focusing on Spanish and Japanese as a specific language pair. This is remarkable considering how massively spread both of these languages are, as well as their popularity both for L2 students and as study cases particularly with regards to their prosody. However, some studies do touch either on the specific pair or on aspects that are relevant for its discussion, and those will be briefly covered in this section.

Stress "deafness"

There have been some studies regarding in specific the perception of Spanish stress as an L2, and in particular from the perspective of French (Dupoux, Pallier, et al., 1997; Dupoux, Sebastián-Gallés, et al., 2008; Muñoz et al., 2009; Muñoz Garcia, 2010; Schwab and Llisterri, 2011a,b, 2012, 2014), and Italian speakers (Alfano, Llisterri, and Savy, 2007; Alfano, Savy, and Llisterri, 2008; Alfano, Schwab, et al., 2010).

Studies with French speakers became popular when Dupoux, Pallier, et al., briefly mentioned above, reported evidence of a difficulty in the perception of non-native segmental contrasts affecting French speakers but not Spanish speakers, which was referred to by the authors as "stress deafness".

It is clear that the word is ill-suited, since they themselves report that the magnitude of the performance gap between French and Spanish speakers depends on the type of task they are presented with. They report tasks with high phonetic variability and memory load as having the lowest performance, while tasks that had a lower memory load did not prove as problematic. In a later study, this was taken to mean that the processing problems faced by French speakers were the result of the lack of a phonological representation of stress (Dupoux, Sebastián-Gallés, et al., 2008). The higher performance rates in discrimination tasks, they reasoned, were explained by participants having access to acoustic representations instead of phonological ones.

However, these claims have also been contested. Muñoz et al. (2009) tested the performance of French speakers in a lexical identification task, with words in isolation as well as in sentences with falling and rising intonations, and showed that correct perception of lexical stress position depended on the language proficiency of the participant, with advanced learners reaching 95% correct identification across conditions. They also showed that perception was more difficult in rising intonations (although this affected oxytone accents the least), and that performance was significantly affected by word-length and stress position.

Similarly, in a series of studies Schwab and Llisterri (2011a,b, 2012, 2014) tested whether French speakers are able to learn to perceive the Spanish stress contrast, and they also found high performance rates for French speakers (although not as high as that of native speakers). Although Muñoz et al. (2009) embraces the possibility that this was due to the task chosen, this is unlikely to be the case, since the argument put forth by Dupoux, Sebastián-Gallés, et al. (2008) referred to discrimination tasks, and these results stem from word identification tasks. This is even clearer in the case of the work by Schwab and Llisterri, who used a matching task, in which participants were asked to identify abstract shapes using nonce words that had been linked to them; a task that resembles the ones where Dupoux, Sebastián-Gallés, et al. (2008) identified low performances.

These results are in fact quite similar to those found for Italian speakers (Alfano, Llisterri, and Savy, 2007; Alfano, Savy, and Llisterri, 2008; Alfano, Schwab, et al., 2010), not only in that they show that L2 learners are able to learn non-native prominence contrasts, but also in that they show that despite this, there are clear differences in the way that the acoustic correlates of stress are interpreted by native and non-native populations. Interestingly, although in Spanish F_0 is the strongest cue for prominence, both French and Italian speakers seemed to be more sensitive to changes in duration and (to a lesser extent) intensity than changes in the fundamental frequency (Schwab and Llisterri, 2012; Alfano, Savy, and Llisterri, 2008). Which means that, regardless of the side in the controversy surrounding the notion of "stress deafness", there is strong evidence that suprasegmental perception is also affected by language transfer effects.

Japanese learners

While not directly concerned with Spanish speakers, a study by Nishinuma, Arai, and Ayusawa (1996) examined the perception of what they called the Japanese "tonal accent" by American students of Japanese. Comparing their results with previous research done on French, Korean and Chinese students of Japanese, they reported that American students performed highest with unaccented words, and showed a significant effect of skill level for accents that occurred early in the phrase (in the first two syllables). Accents that occurred in the third or fourth syllables had significantly lower performances across their spectrum of L2 skill. However, this "skill level" is not automatically translatable to more general measures of language proficiency since it was defined post-hoc as the overall skill of the participants at performing the task itself.

Accent-type contrast as possible explanation

In attempting to explain this specific behaviour they say that

> accent in English — a "stress accent" language — is marked by an *increase* in the degree of "force". In terms of acoustic parameters, this corresponds to a pitch peak, a prominent intensity, and vowel lengthening. Tonal accent in Japanese involves the opposing type of phenomenon, *i.e.* a more or less abrupt *decrease* in F_0. The difficulty experienced by English speakers is predictable. (p 648; emphasis in the original).

The phrasing chosen by the authors also suggested that they believed the underlying reason for the poor performance was the contrast in accent types: Japanese had what they call a "tonal" accent, and English a stress accent. But the features they mention for the English stress (which did not include vowel reduction) also characterise the Spanish stress, which would mean that similar results should be expected for Spanish learners of Japanese. And if the explanation for the effect on the perception of English speakers lies in the contrast between accent types, then similar results should also be expected for Japanese learners of Spanish, since the contrast itself is independent of the direction of learning.

However, Kimura et al. (2012) reports the results of a perception study on the effects of sentence intonation on the perception of Spanish accent by Japanese learners, and what they found does not agree with the account presented in Nishinuma, Arai, and Ayusawa (1996). This study will be reviewed in more detail in section 1.5.

Research on Spanish as L2

As for the specific language pair of Japanese and Spanish, a dissertation by Cabezas (2009) presents results of a quantitative survey on the attitudes of Japanese students of Spanish highlighting the widespread impression in Japanese L2 learners of Spanish that pronunciation is the easiest aspect of the language. While rooted on the similitudes of the Spanish and Japanese phonological inventory of segments, there is ample evidence that the pronunciation of Spanish by Japanese learners is still heavily affected by their L1 (Ueda, 1977, 1978; Yasutomi, 1994; Martínez, 1995; referenced in Carranza, 2012; Carranza et al., 2014), and that most of the problems they have that lead to errors in communication stem from problems in their production of prosody (Martínez, 1995).

1.4.1 Non-native language proficiency

Like prominence, L2 "proficiency" is a difficult term to use because of the large number of ways in which it can be understood, and the large number of different factors that are involved in it. This includes the age at which learning began; factors related to language instruction (if there was any instruction), such as the proficiency of the instructor, and the quality of the classes (if any); those that are prior to the study of the L2, such as more general language skills and other developmental aspects; and environmental factors, such as the level of immersion the learner has in the L2, and the percentage (and type) of use they make of the language on a regular basis. Evidence also suggests that proficiency will vary depending on the linguistic modality and domain (Tremblay, 2011; Bachman, 1990).

There is extensive evidence of the effect of these factors on L2 production and perception. Numerous studies have examined the effects of language training has at a segmental level (*e.g.* Flege, 1995, and Iverson, Hazan, and Bannister, 2005; Iverson, Pinet, and Evans, 2011, for some recent comparisons of different training methods; see also the beginning of this chapter for additional coverage), and there is ample evidence that a similar effect exists also for suprasegmentals. In a study on the factors involved in the pronunciation accuracy of non-native learners of English reported on Purcell and Suter (1980), length of residence (LOR) was the third most significant factor; and a study on the effects of training on the perception of Mandarin tones found that performance on perception was improved by high-variability auditory training for children and adults of different ages (Wang, Spence, et al., 1999; Wang and Kuhl, 2003).

The studies by Schwab and Llisterri reviewed above are particularly relevant in this case as well, since their results showed that training was able to improve the learners' rate of correct identification of Spanish stress in nonce words (although more recent results suggest improvement might not depend entirely in the training itself).

Most of the studies on the benefits of proficiency in general in the perception and production of L2 sounds has focused on SLA, and other such settings in which learners receive large amounts of L2 input. In contrast, foreign language acquisition (FLA), the result of largely formal training taking place in classrooms in countries where the target language is not predominant, provides a much more impoverished L2 input. Because of this, results from one learning environment are not necessarily applicable to those of the other (Best and Tyler, 2007), which raises the question as to whether proficiency effects would show in these cases.

Kimura et al. (2012) used participants who were FLA learners, but they did not control for language proficiency, which makes it impossible to tell how much of the variation they found, if any, is due to differences in the proficiency of the participants. This is an important question because of the implications it has for prosodic training in general. Most L2 students will study under these impoverished conditions, and it is important to be able to say whether the training that does take place, however small, has an effect or not.

1.4.2 Measuring proficiency

For the purposes of this dissertation, proficiency will be used as an intentionally broad term to refer to the joint effect that all of the factors mentioned above (and others that may have been left out) have on the language learner's ability to use and make themselves understood in their L2. While these solves the problem of having to specify and control for the specific factors that might be of interest, it does introduce the problem of how something like this can be measured.

One way to encapsulate a complex notion like that is to use a collection of methods, each one addressing a specific dimension of the task at hand, dimensions which may or may not be of the same level of specificity. Tremblay (2011) presents a review of methods to measure proficiency used in SLA studies, and although the authors are strongly in favour of *cloze* tests, they themselves acknowledge that it will give a better overall picture when paired with both an auditory perception test, as well as the participants' self-rated proficiency measures. As measures becomes more complex, however, so does the actual test of which it is but a part, and there is a limit to how complex a task can be made before its complexity in itself starts interfering with the results. In the case of the studies in this dissertation the length of the test was a considerable concern, which motivated the decision to move away from proficiency tests like the ones described in Tremblay and Garrison (2010) and Tremblay (2011), and towards a measure of proficiency that, while less granular, would not task the participants as heavily.

The measure of choice was a self-reported proficiency questionnaire. Linguistic questionnaires have been use extensively in bilingualism studies, and they have been reported to show significant correlations between responses to theoretically grounded questions (like length of studies) and self-assessment measures of language ability. The questionnaire tested in those studies has been described in detail in Li, Sepanski, and Zhao (2006) and extended in Li, Zhang, et al. (2014). Those papers report the development of a generic web-based linguistic history questionnaire (the LHQ), made up of the most common questions found in a review of over 40 studies which used linguistic questionnaires for the assessment of their participants.

As can be expected, linguistic questionnaires in general are not without shortcomings. MacIntyre, Noels, and Clément (1997), for example, reports that a speaker's *language anxiety* will introduce a bias in their self reporting of language proficiency; while Delgado et al. (1999) shows that the correlation with other independent tests is better for reading and writing than for listening and speaking. In an attempt at improving the self-reported questionnaire, Marian, Blumenfeld, and Kaushanskaya (2007) shows that widening the scope of questions in the questionnaire, so that they take into account different aspects of the student's linguistic history, can provide a valid and reliable assessment of the student's language profile.

Despite the pleads of Li, Sepanski, and Zhao, however, this dissertation does not use their questionnaire, but one that is very similar in design and capabilities, a decision that was made to facilitate the handling of the collected data, and to offer the participants a unified style instead of asking them to switch over different online tools.

Based on the joint suggestions of Li, Sepanski, and Zhao (2006) and Marian, Blumenfeld, and Kaushanskaya (2007), the questionnaire included a number of questions regarding language use, and participants were asked to provide independent ratings for their use at home, with their friends, for entertainment purposes, for their work (or for their studies, when participants were students), and for personal enjoyment (*e.g.* writing a diary or personal journal, a common practice with language learners). In the case of proficiency, independent self-assessment ratings were obtained for the four main linguistic abilities of listening, speaking, reading and writing. A sample of the linguistic questionnaires used in this dissertation is shown in appendix A.

Other measures of proficiency: cloze tests

There are many other measures that have been used to gauge L2 proficiency with varying degrees of success. A review of a number of different assessment methods, published in Tremblay (2011), highlighted the need to improve the quality of the assessment in second language acquisition studies. They show that a *cloze* (fill-the-blank) test is a good predictor of overall L2 proficiency, which correlates well with other measures (such as those obtained from a linguistic background survey).

However, it it not clear that the benefits they describe would make that a suitable test in this situation.

First of all, cloze tests pose a rather high entry level requirement: they are written tests that require the participant to write-in their responses, which makes them not very well suited to low-level students. The authors in the above mentioned review actually acknowledge this as a limitation of the test. The fact that it is a written test also can be a problem, since it tests primarily reading and writing skills in that order, which makes it necessary to pair it with additional tests (such an aural perception test) to cover other linguistic abilities.

Particularly in the case of the studies reported in this dissertation, these are not minor concerns. First of all, the length of the experiments meant that whatever tool was chosen to measure L2 proficiency would have to introduce the least amount of load on the participants, both in terms of the time it took to complete, and the difficulty of answering the questions. An entirely separate issue dealt with the relative gap existing between the familiarity of both populations (Spanish and Japanese speakers) with their respective L2's writing system (see Chikamatsu, 1996). Even if this is not so serious for J1S2 (since there are many channels through which they might

Meaningful words			Nonce words		
Número	Numero	Numeró	Ládebo	Ladebo	Ladebó
Médico	Medico	Medicó	Núlibo	Nulibo	Nulibó
Límite	Limite	Limité	Máledo	Maledo	Maledó
Válido	Valido	Validó	Lúguido	Luguido	Lugidó

Table 1.3: target words used by Llisterri et al. (2005, 2003) and Kimura et al. (2012).

acquire familiarity with the Roman alphabet used in Spanish), whatever measure of L2 proficiency that was used needed to be applicable for both tests. This meant that a writing-based proficiency test (or one that included questions of that kind) was unsuitable almost from the start, if one did not want to give S1J2 participants an unfair playing field.

1.5 Prior research

The work in this dissertation builds on the results of a pilot study which focused on the effect that sentence intonation had on the perception of different types of Spanish stress by Japanese learners (Kimura et al., 2012). Although the research scope has been greatly extended, by including a bidirectional perspective and examining the effects of cue weighting in non-native populations, a review of the original study is helpful to inform the discussion that will follow. This section provides such an overview.

1.5.1 Background

Speech materials

Target words were taken from the set designed and used in the studies by Llisterri and colleagues (see Llisterri et al., 2003, 2005; Alfano, Schwab, et al., 2010, among others) and shown in table 1.3. These were all tri-syllabic CV words in which the accent could occur in any of the last three syllables. A native Spanish speaker from the Spanish province of Cuenca was recorded producing all 24 meaningful and nonce target words in isolation, as well as one each of the meaningful and nonce word sets in a number of different sentence contexts: internal position in an interrogation (*qint*); final position in an affirmative statement (*sfin*); internal position in an affirmative statement (*sint*); internal position in a negative statement (*nint*); and final position in an interrogation (*qfin*).

Each token was repeated 5 times, which resulted in a sum total of 270 tokens ($24 \times 5 + 6 \times 5 \times 5$). Recordings took place in the phonetics laboratory at Sophia University.

Participants and procedure

Fifty one Japanese students of Spanish residents in Japan and 43 native Spanish speakers from Spain enrolled in Japanese language courses took part in the study reported in Kimura et al. (2012). Their level of proficiency was not controlled for. Participants listened to the recorded stimuli through headphones and were presented with a pen-and-paper 3AFC identification task. Each item consisted of a transcription of the sentence they had heard with the target word blanked out, and a list of possible words (the three members from the trio). All the text on the test was written using standard Spanish orthography. Students were told to mark on the list the word they had heard.

Results

The study showed a strong effect of sentence context, as can be seen in figure 1.4 which reproduces data from the original publication. The figure shows the percentage of correct responses obtained by both Japanese and Spanish participants for each of the six contexts studied, sorted in decreasing order for the performance of the Japanese participants. Performance for some contexts (*isol*, *qint* and *sfin*) was comparable for both the native and the non-native groups, and no significant difference was found between these conditions for the L2 learners. However, words in final positions of interrogations (*qfin*) had the lowest performance by a significant margin, reaching chance level.

Sentence context effect

Contexts for which Japanese students of Spanish (from now on J1S2) reached native-like performance had the target word in final positions of intonational phrases. On the other hand, the *qfin* context placed the target word in the middle of the Spanish rising interrogative toneme, making it difficult for the Japanese participants to tell the information from the lexical

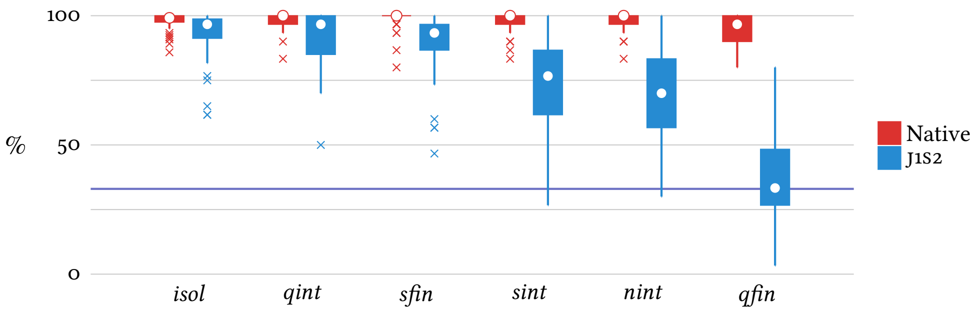


Figure 1.4: Percentage of correct identification of the prominent syllable in Spanish minimal trios in different sentence contexts. The violet ● line marks the chance level. Results from Kimura et al. (2012).

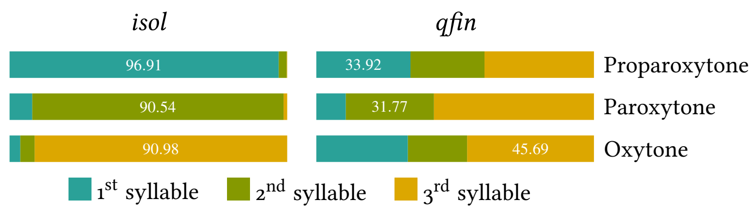


Figure 1.5: Spanish syllables perceived as accented by non-native participants in words in isolation (left) and in final positions of interrogative utterances (right). Results from Kimura et al. (2012).

intonation from that of the sentence intonation. This suggested that pitch was the primary acoustic cue used by the Japanese students.

Accent type effect

Significant differences were also found in the responses for words with the prominence in different positions, as can be seen in figure 1.5. The figure shows the distribution of responses for the non-native speakers in both the highest (*i.e. isol*) and the lowest (*i.e. qfin*) performing contexts, and highlights the extent of the difference, and the significantly lower performance for paroxytone words. The original report only showed the analysis for words in *qfin* context, but this finding remains significant across conditions, as confirmed by a repeated measures ANOVA run on the original data ($F_{(2,850)} = 21.996$, $p < 0.001$). As suggested by figure 1.5, correct responses for paroxytone words ($\bar{x} = 75.96$, $\sigma = 15.63$) were significantly lower than those for oxytone ($\bar{x} = 86.1$, $\sigma = 9.78$) and for proparoxytone words ($\bar{x} = 81.83$, $\sigma = 11.77$), and this was confirmed by a two-sample t-test run between the first set and the latter two ($t_{(151)} = 3.45$, $p < 0.001$).

1.5.2 Findings

These results showed that, despite the differences in accent type between Spanish and Japanese, J1S2 do not seem to have any difficulty with the perception of the Spanish stress at the end of intonational phrases, a fact that does not seem to agree with the conclusions in Nishinuma, Arai, and Ayusawa (1996). This means that either their results do not depend on the fact that English and Japanese have different accent types, or that they failed to account for some other factor which is responsible for their results. However, results from Kimura et al. (2012) also showed that the ability of J1S2 to perceive the Spanish stress is fragile, and very susceptible to differences in the intonational context of the word in question, suggesting that they relied excessively on pitch and were not capable of tuning in to the additional cues available in Spanish.

A re-analysis of the results also suggests that the performance for different accent positions is dependent on contextual differences in intonation, and consequently that specific accent realisations might bias participants towards the perception of the accent in different positions (*e.g.* the preference for oxytone accents in *qfin* contexts).

The design of this test was focused on exploring the effects of sentence context and accent type on the perception of non-native listeners, and not on making direct comparisons between native and non-native populations. Because of this, no measures were taken to avoid ceiling effects in the perception of native speakers, which made the analysis of interactions between both groups impossible. This because ceiling effects may elicit type-II errors for main effects, and type-I errors for interactions. Limitations

This limitation was due to the fact that this was an early study, looking at a comparison that had not been investigated before, and largely presented as exploratory. Since the first studies in this dissertation (reported in chapter 3 for Spanish and chapter 4 for Japanese) inherited the design from that study, they inherited this limitation as well, and did not look at between-group interactions.

1.6 Research objectives

1.6.1 A bidirectional approach

This dissertation is focused on the perception of L2 prosody, a field whose contributions to the understanding of language learning and speech perception have still not been thoroughly explored. While the studies presented herein will look at the specific cases of Spanish and Japanese, the fundamental questions they try to answer are not specific to these languages: they deal with the perception of suprasegmental phonological categories across language barriers, and with the interaction of accent systems that differ phonologically (see section 1.3) but have similarities at a phonetic level.

Language transfer effects are a phenomenon which emerges from the interaction between two distinct linguistic systems. Therefore, the specific ways in which they will intervene in language learning will depend on what the L1 and the L2 of the learner are, and the way in which the features from those languages interact with one another.

In the case of the study by Nishinuma, Arai, and Ayusawa (1996), however, they suggest that their findings are the result of the interaction not necessarily of two specific languages, but of languages that have different accent systems. In short, that this particular effect at least partly emerges from the contrast between the phonological definition of prominence in both languages, and its effects on the interpretation of the acoustic signal. However, the results shown in Kimura et al. (2012), from what ostensibly is a language pair that presents the same accentual contrast, do not agree with this explanation. How can we account for this discrepancy?

One possibility is that those findings are not due to a contrast of accent types, but the result of some particular interaction that depends on the direction of language learning. To test this, this dissertation employs a bidirectional approach, to study one feature across languages using the direction of learning as one of the conditions. Studying the same linguistic features from both directions, with a shared methodology and focus —

understanding that an exact reproduction of the methodology will be impossible in most cases — should make it possible to do more than simply study the transfer effects that exist in a particular language pair. It should provide novel insights into the circumstances under which a given feature can improve or hinder the perception of non-native categories, which will have implications for the more general understanding of speech perception.

This is what motivates this otherwise unorthodox design.

1.6.2 Research questions

Throughout this chapter, several theoretical and methodological aspects have been reviewed. Some of these serve to motivate the questions that lie at the core of this dissertation, while others were presented because they were considered to be relevant to understand the discussion of those questions.

Four main research questions lie at the core of this dissertation:

1 Do the effects of perceptual interference in L2 learning shown at the segmental level occur as well at the suprasegmental level in the acquisition of L2 lexical stress?

2 Is the weighting of L2 acoustic cues of suprasegmental categories affected by that of L1 cues, as it happens for segmental categories?

3 Does L2 language proficiency have an effect on the perception of L2 lexical prominence in FLA learners?

4 Are the transfer effects affecting cross-linguistic perception of lexical prominence independent of the direction of learning?

This motivated two sets of mirrored experiments which will be described in detail in the following chapters. The first experiment is designed as an extension of the paradigm in Kimura et al. (2012), and focuses primarily on the interaction between the position of a target word in the sentence and the position of the prominence in that word, and its effects on the perception of L2 prominence. Chapter 2 explains the design and collection of the speech materials used for both Japanese and Spanish, and shows the results of an cross-linguistic acoustic analysis of prominence; the empirical manifestation of the contents of sections 1.3.1 and 1.3.2. Chapters 3 and 4 describe the results of the first experiments for Spanish and Japanese respectively, and explore the implications of their results in the context of this dissertation.

The second experiment, described in part 2, explores the sensitivity of non-native learners to secondary acoustic cues for the perception of L2 accentual categories, and the differences in their evaluation of those cues. Chapter 5 explains the process behind the synthesis of the stimuli for those tests, and chapters 6 and 7 show the results of those studies for Spanish and Japanese respectively.

Chapter 8, at the end of the dissertation, provides a general discussion of the results of the tests in parts 1 and 2, their implications, and the way in which they relate to the themes discussed in this chapter.

Part 1

Factors involved in the perception of second language prominence

Chapter 2

Design and analysis of speech materials

The previous chapter provided an overview of the ways in which Spanish and Japanese mark lexical prominence, and of some of the research on the factors involved in non-native perception of these cues. It also described the results of Kimura et al. (2012), a study which looked at the effects of sentence intonation in the perception of Spanish stress by Japanese learners, and which motivated the first set of studies in this dissertation. Before proceeding to those studies (which are described in chapters 3 and 4, dealing with L2 Spanish and Japanese respectively), this chapter describes the aspects in the speech materials from Kimura et al. (2012) that were subject to improvement, explains the design and collection of new materials for Spanish and Japanese, and provides details of their acoustic properties.

2.1 Materials

2.1.1 The source Spanish material

Keywords

The materials used in the Spanish tests were based from the same set of keywords that were used in Kimura et al. (2012), and initially in Llisterri et al. (2003, 2005). These words were originally selected based on the following set of requirements: that they were real words with 3 syllables each, all of which had a CV structure, and whose only difference was the position of the stress. This resulted in 4 sets of accentual minimal trios (for a total of 12

Accent	Minimal sets			
	/limite/	/mediko/	/numero/	/balido/
Proparoxytone	Límite limit N	Médico medic N	Número number N	Válido valid N
Paroxytone	Limite limit V-1SG-PRES	Medico prescribe V-1SG-PRES	Numero number V-1SG-PRES	Valido validate V-1SG-PRES
Oxytone	Limité limit V-2SG-PAST	Medicó prescribe V-2SG-PAST	Numeró number V-2SG-PAST	Validó validate V-2SG-PAST

Table 2.1: Spanish target words used by this study. Compare with the Japanese keywords shown in table 2.4, on page 78.

Accent	Minimal sets			
	/limite/	/mediko/	/numero/	/balido/
Proparoxytone	76.85	273.49	769.99	38.15
Paroxytone	0.0044	0.0044	0.0199	0.0044
Oxytone	-	-	-	-

Table 2.2: Occurrences per million words of each of the target words in Spanish, as reported by LIFCACH for Chilean Spanish. Compare with table 2.5, on page 79.

keywords), which are shown in table 2.1 along with their glosses.

These requirements left only a limited number of choices, to the extent that the selected words are the only ones in standard Spanish that fulfil all conditions. This meant that equality in their rates of occurrence could not be ensured, and indeed proparoxytones were far more common than the rest in all trios. Table 2.2 shows the extent of this disparity.

Despite the large differences shown in the table, this was not considered to be a serious problem, since studies that have used these trios before showed no difference between them and similar sets of nonce words (Llisterri et al., 2003, 2005; Kimura et al., 2012).

Carrier sentences

To examine the effects of changes in the intonational context surrounding these words, they were framed in low-predictability carrier sentences which imposed few requirements on the framed words. This also served to minimise other possible effects of word frequency, and allowed the trios to be composed of a combination of different parts of speech, with two thirds being verbs in different conjugations, and the remaining being nouns. To preserve comparability with the source study, the carrier sentences represented largely the same sentence contexts as in Kimura et al. (2012)[1].

Carrier sentences

Four carrier sentences were used to cover the combinations between 2 different types of statements (declarative and interrogative), and 2 positions of the keyword within those statements (final and non-final). Keywords were also presented in isolation for a total of 5 different conditions.

While the selection of the keywords was not changed, the text of the sentences did undergo some changes. In the original stimuli, the complexity and length of some of the utterances caused the speaker who produced them at the time to break them into two independent intonational phrases, often with a short pause in between. As this phrasal boundary tended to fall immediately after the target word, the position of that word changed from non-final to final, meaning one of the conditions was not faithfully represented.

The resulting sentences shown below (where _____ indicates the position of the target word) are an attempt at solving these problems while keeping the original structure of the sentences unchanged.

sfin Dijo la palabra _____

(He/she) said the word _____

sint Dijo la palabra _____ de nuevo

(He/she) said the word _____ again

qint ¿Dijo la palabra _____ de nuevo?

Did (he/she) say the word _____ again?

qfin ¿Dijo la palabra _____?

Did (he/she) say the word _____?

1 The only exception was the non-final position in a negative statement (*nint*) context, which was discarded since it was originally intended to serve as a contrast with the non-final position in a declarative statement (*sint*) context to study possible effects of focus.

2.1.2 Japanese speech materials

In keeping with the bidirectional approach set forth in the introduction, whatever modification was made to the set of materials used in the Spanish test needed to have an equivalent for Japanese. This was essential to ensure the maximum degree of comparability between both studies.

Keywords

Like Spanish, Japanese also has word-level prominence, and it also varies freely in that it can fall on any of the syllables in the word (see section 1.3.2). But since Japanese has unaccented words, some words can lack an accent altogether. This does not happen in Spanish, where all content words have lexically-determined prominence. Because of this, one of the main points of interest of these studies was whether Spanish speakers would be able to correctly map those words into appropriate accentual categories, and what factors would be involved in this process. Therefore, while the Spanish keywords were chosen to test the different positions of the Spanish prominence, the Japanese words had to make it possible to compare between accented and unaccented as well as between words with different accent positions. To keep the number of conditions equal to those of the Spanish test, the Japanese materials would focus on the following 3 accent types: unaccented words; initial-accented words; and final-accented words[2].

While the original materials all had 3 syllables, a study on the role of Japanese accent in the discrimination of minimal sets of Japanese words (Sibata and Shibata, 1990) found no words fitting these constraints (sets of 3 words with 3 syllables, each with a different accent as their only difference). However, it did report the existence of 12 accentual minimal trios of 2-mora words, which could in principle be used for this study (since in Japanese n-mora words accommodate $n+1$ accent patterns, including unaccented).

In order to find these sets (which were reported, but not speci-

2 These accent types are entirely equivalent to what is sometimes reported as *type-0*, *type-1*, and *type-2* accents respectively. In Japanese, they are also sometimes respectively called *heibangata* 平板型, *atamadakagata* 頭高型 and *odakagata* 尾高型.

r	*n*														
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1643	351	113	42	13	10	7	1	1	1	1	0	1	0	0
2	516	189	79	41	22	9	5	2	2	1	0	1	0	0	1
3		15	5	4	3	5	0	1	2	0	0	0	0	0	0
4			0	1	0	0	0	0	0	0	0	0	0	0	0

Table 2.3: Distribution of Japanese accentual minimal contrasts, based on data reported in Sibata and Shibata (1990). Each cell represents the number of *n*-member *sets of homophones* distinguished by *r* accent patterns.

fied), the analysis described in Sibata and Shibata (1990) was repeated by hand with an earlier edition of the same accent dictionary used in their study (Kindaichi, 1981). This process yielded a similar number of words that became likely candidates, but closer examination revealed most of these would have to be discarded because they

1. had items which were made up of at least one mora which was not cv;
2. had items that were used almost exclusively in compounds, where the accentual difference disappeared;
3. had items which were so infrequent that they were unfamiliar to some of the speakers during the recording session (at least as realised with the desired intonation); or
4. had items which belonged to different grammatical categories.

The final requirement might seem out of place, considering none of the Spanish minimal trios meet it. However, as will be explained in more detail below, these more stringent criteria were needed to accommodate the Japanese carrier sentences, and were considered to be a minor compromise.

After discarding the offending candidates, two trios were left for which these problems were solved to a satisfying degree. This provided half of the 12 stimuli that made up the set of Spanish materials.

Inclusion of minimal pairs

As is shown in table 2.3 (adapted from Sibata and Shibata, 1990), accentual minimal pairs are much more common than minimal trios in Japanese. To equalise the number of stimuli on the sets used for both languages, three minimal pairs where added to the set of words, each representing a different contrast between the three accent types to be examined: one

Accent	Minimal sets				
	/kaki/	/haçi/	/hana/	/momo/	/kami/
First $\mu^{\downarrow}\mu$	牡蠣 oyster N	箸 chopsticks N		腿 thigh N	神 god N
Last $\mu^{\uparrow}\mu^{\downarrow}$	垣 fence N	橋 bridge N	花 flower N		髪 hair N
Unaccented $\mu^{\uparrow}\mu$	柿 persimmon N	端 end N	鼻 nose N	桃 peach N	

Table 2.4: Japanese target words used by this study. A $^{\downarrow}$ indicates a fall in pitch while a $^{\uparrow}$ indicates a pitch rise. μ indicates a mora. The top two rows show the triplets while the lower three show the minimal pairs. Compare with the Spanish keywords shown in table 2.1, on page 74.

between initial-accented and unaccented words (*i.e.* /kami/); one between final-accented and unaccented words (*i.e.* /momo/); and one between initial- and final-accented words (*i.e.* /hana/). The final set of 12 target words used in the experiment — including trios and pairs ($2 \times 3 + 3 \times 2$) — are shown in table 2.4.

Like with the Spanish stimuli, similar word frequencies had to be sacrificed to meet the keyword requirements . Table 2.5 shows the frequency of occurrence for each of the final words as reported by the `internet-jp` corpus of Japanese (see Sharoff, 2006, 2007). However, once again as with the Spanish test, this was not considered to be an issue based on previous studies with similar corpora (Llisterri et al., 2003, 2005; Kimura et al., 2012).

Word counts in Japanese are made difficult by the fact that Japanese does not use any graphic separation between words in running text, does not have any accent marks, and is written using a combination of multiple sets of characters. The final point is particularly relevant in this case since the lack of graphical accent marks in Japanese means that cases in which the word was written using the moraic *hiragana* alphabet are impossible to disambiguate without access to each instance of the word. The rightmost column in table 2.5 reports the number of such ambiguous oc-

Accent	Minimal sets				
	/kaki/	/haɕi/	/hana/	/momo/	/kami/
First	-	58.16		-	106.48
Last	-	21.23	76.48		20.78
Unaccented	2.79	13.61	17.35	6.17	
Hiragana	-	3.30	3.94	4.88	-

Table 2.5: Occurrences per million words of each of the target words in Japanese, as reported by the `internet-jp` corpus of Japanese (see Sharoff, 2006, 2007). This table only counts entries in which the words appeared isolated, not as part of compounds. The last column lists the frequency of relevant words that appeared only as spelled in *hiragana*, making it impossible to disambiguate their meaning. Compare with table 2.2, on page 74.

currences found for each word set.

The "third" syllable

As stated in the introduction, a final-accented Japanese word will be indistinguishable from an unaccented one when these words are presented in isolation (*cf.* Vance, 1995), because the difference between the two lies in the shape that the pitch contour takes after the end of the word. To overcome this problem, a grammatical particle was added to each 2-syllable word, and it was this 3-syllable phrase that was used as the "target word". This grammatical particle made it possible to see the shape the pitch contour took after the word ended, and thus allowed for the discrimination between final-accented and unaccented words.

The binding particle /mo/ も (en. *also*) was chosen since it presents little contextual requirements; it is able to create phrases that can exist isolated as well as within a sentence; it consisted of a single cv syllable in which the consonant is a nasal — which would not interfere with F_0 measurements; and most importantly, it does not affect the accent of the preceding word (see page 53). This explains the additional grammatical category constraint that was included only for Japanese, since this particle does not bind equally to all parts of speech.

Carrier sentences

Changes were also needed in the carrier sentence in order to make them adapt better to the structural constraints of Japanese — which is an sov language, as opposed to Spanish which is svo. The final carrier sentences are shown below, once again with _____ marking the position of the target word.

sfin ano _____ mo

That _____ too

qfin ano _____ mo?

That _____ too?

sint ano _____ mo to iimashita

(He, she) said "that _____ too"

qint ano _____ mo to iimashita ka

Did (he|she) say "that _____ too"?

A pitch fall was expected to occur following the final syllable of the keyword compound (*i.e.* the particle /mo/) in all but the isolated context, meaning that sentences with unaccented words would effectively have an accent in this "third" syllable, which would facilitate the parallel between Japanese and Spanish sentences significantly.

2.2 Data collection

Recordings for both languages took place primarily in London, at the Speech Sciences Laboratories at UCL, where samples were recorded in a soundproof room using a Røde NT1-A large diaphragm condenser microphone. All recordings were made digitally with 44100 16-bit samples per second using professional recording software. A subset of the Spanish recordings took place in Chile under the supervision of personnel of the phonetics laboratory at Universidad Católica de Chile[3]. Recordings there also used a soundproof room, but used instead an AudioTechnica AT3035 large diaphragm condenser microphone.

Participants were provided with a printed list of sentences written using the standard orthographic rules for Spanish or Japanese as appropriate. In the case of Japanese, words were written using logograms (for semantic disambiguation) with a pronunciation guide written on top. The pronunciation guide was spelled out using the *hiragana* moraic alphabet, with

3 Special thanks to professor Domingo Román for his help.

the accented syllable marked in red to serve mostly as a mnemonic for the speakers. No such special mark was needed in Spanish, so only the standard accent mark (´) was used.

Recording sessions were monitored from outside the recording booth. The researcher in charge would monitor the speakers and prompt them to read out a specific item of the list by number, setting the pace of the recording and ensuring an appropriate spacing between each utterance. If a particular problem was identified, speakers were asked to repeat the sentence either at that time, or later, to avoid repetition effects. The list contained 4 repetitions of each sentence in a random order. The recorded materials were later rated for naturalness by two trained native speakers of each language, and the best rated utterance was selected to be part of the final set of materials.

2.2.1 Spanish speakers

Four Chilean native speakers of Spanish (2 female, 2 male) were asked to record the list of materials. Two of the speakers were around 25 years old (24 for the female; 26 for the male), and two were aged around 40 (43 for the female; 37 for the male). All speakers were phonetically trained. Three of the speakers were born and raised in the metropolitan area of Santiago, and were recorded in Chile. The older female speaker was born and raised in the area of Concepción, in the south of Chile, and had resided in the United Kingdom for 3 years when the recording took place in London. Despite this, no differences between the speakers were apparent to the researcher, and no effects of speaker place or origin[4]were found in the data.

2.2.2 Japanese speakers

Two female and 2 male speakers were used, of an age range similar to that of the Chilean speakers (24 and 36 years old for the female; 23 and 41 years old for the male), and all of them were raised in the Kantō area (which for the purposes of this study includes the prefectures of Tōkyō, Kanagawa,

4 Throughout this dissertation, the terms "place of origin" will be understood to mean, not necessarily the place of birth, but the place where an individual was initially raised during childhood.

Saitama and Chiba). The older female speaker had been living in the United Kingdom for a period of 6 years at the time of recording, but the Japanese raters reported no significant effects on her pronunciation. The rest of the speakers had all been in London for periods shorter than 6 months.

Both younger speakers were in London as overseas students, while the older speakers were working as professional actors, some of their work being in voice-acting in particular. In the end, however, the older male participant had to be discarded from the study because his productions of some stimuli did not follow standard accent placement rules for Tōkyō Japanese. This meant the Japanese materials were spoken by 3 speakers, for a total of 180 stimuli ($12 \times 3 \times 5$).

2.3 Acoustic analysis

2.3.1 Fundamental frequency

Figures 2.1 and 2.2 show the smoothed and normalised F_0 tracks for the stimuli across all speakers, for Spanish and Japanese respectively. The process behind plotting these tracks involved a number of steps in order to minimise meaningless microprosodic differences and to ensure that different utterances from different speakers and different content words were comparable.

1 Individual F_0 tracks for each of the utterances in the final sets were first calculated in Praat using an implementation[5] of De Looze and Hirst's algorithm for automatic per-utterance estimation of pitch range (De Looze and Hirst, 2008; Hirst, 2012).

2 These pitch contours where then simplified using Momel target points as provided by the `momel-intsint` plug-in for Praat (Hirst, 2007). These points were hand-corrected by the researcher to fix any perceivable deviations from the original contour.

3 A piece-wise quadratic interpolation was used to trace pitch curves across the Momel target points. the interpolation function was based on the one

5 The implementation has been made available as a Praat plug-in available at http://cpran.net/plugins/twopass.

described in Hirst (2012), which is defined as

$$t_i \in [t_1 \dots t_k] : h_i = h_1 + \frac{(h_2 - h_1) \cdot (t_i - t_1)^2}{(t_k - t_1)(t_2 - t_1)} \tag{2.1}$$

$$t_i \in [t_k \dots t_2] : h_i = h_2 + \frac{(h_1 - h_2) \cdot (t_i - t_2)^2}{(t_k - t_2)(t_1 - t_2)} \tag{2.2}$$

such that the interpolation between points t_1 and t_2 results from the combination of two half parabolas which meet at a midpoint t_k, the point at which the slope is maximal.

4 Pitch values derived from these curves were converted to semitones in reference to the mean F_0 of each utterance for normalisation. Time values were likewise normalised by setting the beginning of the utterance to zero and its end to 1, and making each syllable in the keyword occupy a third of the keyword's duration[6].

5 The aggregated data, with information from the best-rated repetitions of all speakers and all keywords, was then used to fit Loess curves using the `loess.smooth` function available in the `R` statistics package.

6 The resulting Loess models were finally used to predict pitch values for each time point, producing continuous representations of the data despite the irregularities present in the original utterances (*e.g.* from voiceless phonemes). These predicted values, and the prediction intervals, are what is shown in shown in figures 2.1 and 2.2.

This method makes it possible to construct a detailed, normalised, and continuous aggregate representation of the changes in pitch. By tracking the spread of the prediction intervals — which will increase in sections of the recordings with devoicing or voice qualities that interfere with correct measurements — this method is also transparent about the relation between

6 The time warping that was needed for normalisation was implemented linearly, with anchor points at the beginning and end of the utterance, and at the boundaries of each of the syllables in the keyword. Alternative methods for normalisation exist, and Lucero et al. (1997) makes the case for nonlinear scaling between specific anchor points (what they refer to as *nonlinearly normalised averaging*). However, as they themselves acknowledge, "for short stretches of behaviour the assumption of linear scaling may be acceptable", which meant implementing their nonlinear scaling algorithm would have likely offered little additional gain at the expense of a significantly more complex computation.

this approximation and the underlying data.

Using ToBI for comparisons: a disclaimer

In the following sections, the analysis of the tone patterns in both Spanish and Japanese will be done using primarily the ToBI standard (Silverman et al., 1992). Although this standard has been used mostly for phonological descriptions, and the system itself makes some assumptions regarding the basic tones available as compositional units (a H and a L tones), the system itself is not internally tied to the phonological realm, and even in its original inception it was presented as a "standard for transcribing prosody, analogous to IPA for segments" (Silverman et al., 1992, p 867). Indeed, the ToBI system has been used in the past as a means to annotate and describe intonation separate from the phonology: one such example is the work by Fernández Planas and Martínez Celdrán (2003) referenced in chapter 1, which used an automatic method to assign ToBI labels to intonation patterns, and then interpreted these as the surface realisations of a smaller number of underlying phonological tones.

This is relevant for this dissertation because it should be stressed that, although ToBI is being used, this is only in order to have a common language for the description of both Spanish and Japanese (as was explained in chapter 1), and not to make any claims regarding the phonological status of the tones described below.

Spanish

For each sentence context, the shown curves share a similar shape outside the boundaries of the keyword. This confirms that carrier sentences are behaving as expected, and that their most relevant differences are constricted to the part that contains the keyword. The pronounced peak at the beginning of all sentences is aligned with the accented syllable in the word *dijo* (en. *he/she said*) which bears the most semantic weight.

Three different realisations of stress

Curves within keyword boundaries for both *isol* and *sfin* show a clear predominance of L+H* pitch accents (see page 46 for the inventory of Spanish pitch accents), while *qint* and *sint* mainly had L*+H accents. In the latter, the effects of peak displacement are also clearly evident. Keywords in the

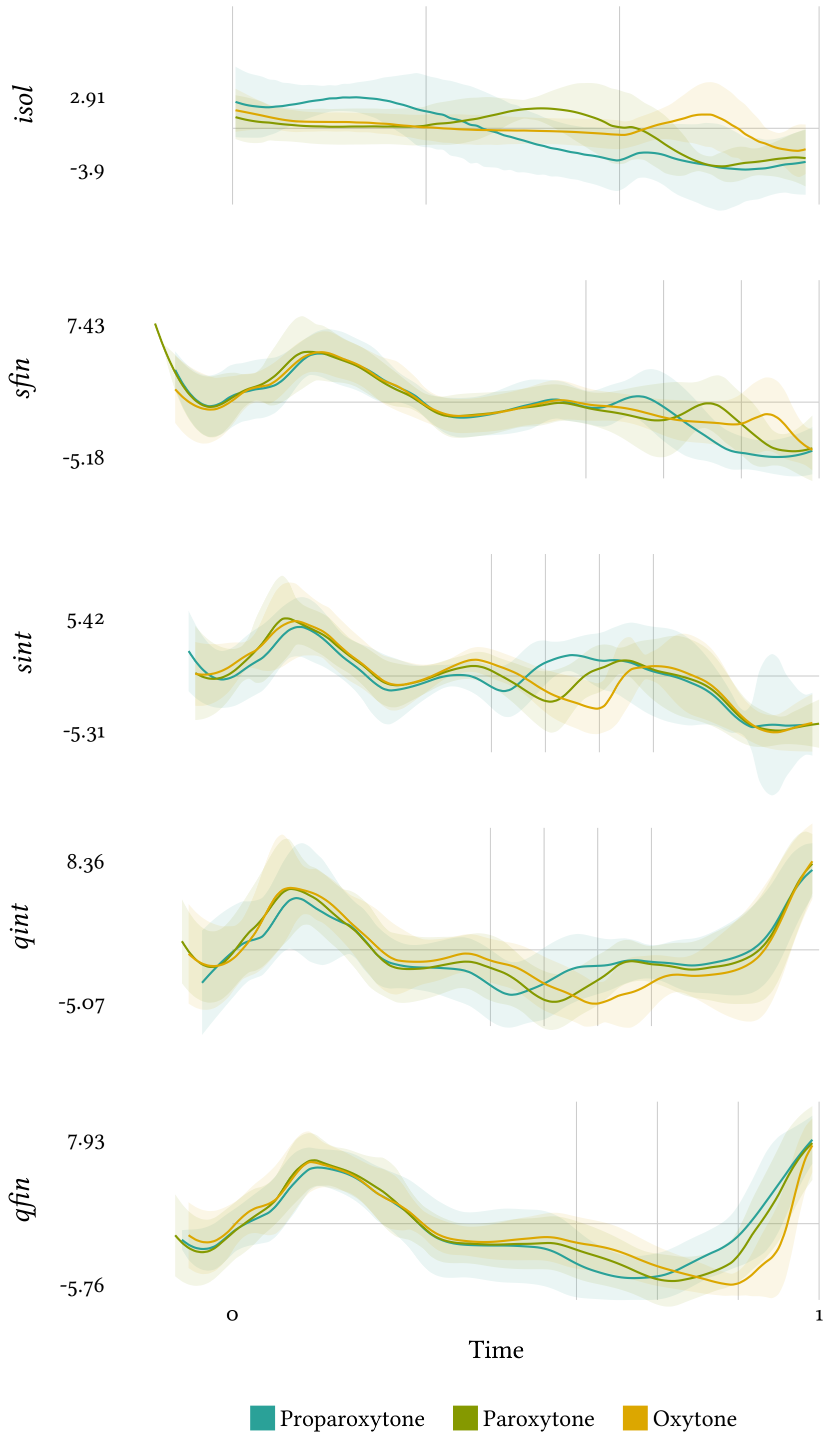


Figure 2.1: Normalised F_0 tracks of stimuli used per context. F_0 values have been converted to semitones in reference to the utterance mean (marked by the horizontal line) and plotted against normalised time values ranging from zero at the beginning of the utterance, to 1 at its end. Syllable boundaries in the keyword are shown by vertical lines. Shaded areas mark the prediction intervals (see the text). Compare with Japanese stimuli in figure 2.2.

qfin context, on the other hand, had primarily L̊ pitch accents immediately followed by a H% boundary tone.

Comparing these F0 tracks and those of the stimuli used in Kimura et al. (2012), which are shown in figure B.1 following the same methodology, reveals that the changes made to the carrier sentences worked as expected, providing distinct intonational contexts for final and non-final positions, and maintaining the steep final rise in the *qfin* context. These differences, although intentional and desirable, do make direct comparisons between both experiments less straightforward.

Japanese

Like with the Spanish test, figure 2.2 shows that the differences in the shape of the pitch contour of the Japanese stimuli is largely constricted to the keyword. Likewise, the data confirms that the stimuli were well designed, in the sense that there were clear differences between the contour for each accent type, that there was as expected no clear ill-effects of adding the final particle, and that the shapes of the different accent types follow what was predicted in the literature.

The sharp increase in the vertical area of the prediction intervals towards the end of *isol* and *sfin* shows that measurements in these regions did not result in reliable data. The final rises in the interrogative sentences (*i.e. qint* and *qfin*), on the other hand, have comparatively much narrower prediction intervals, showing that these were reliably calculated.

The curves also show the robustness of the Japanese accent under different prosodic conditions, an aspect that is particularly clear in the case of the *qfin* context. More generally, while in the Spanish test comparable sentence contexts elicited three altogether different realisations of stress, Japanese accented words remained relatively unaffected by the changes in the position of the word in the sentence, and by the type of sentence in question.

Differences for unaccented words

Flat at end of declaratives

The largest differences are seen in the behaviour of the pitch contour of unaccented words. When they are in final position of a declarative state-

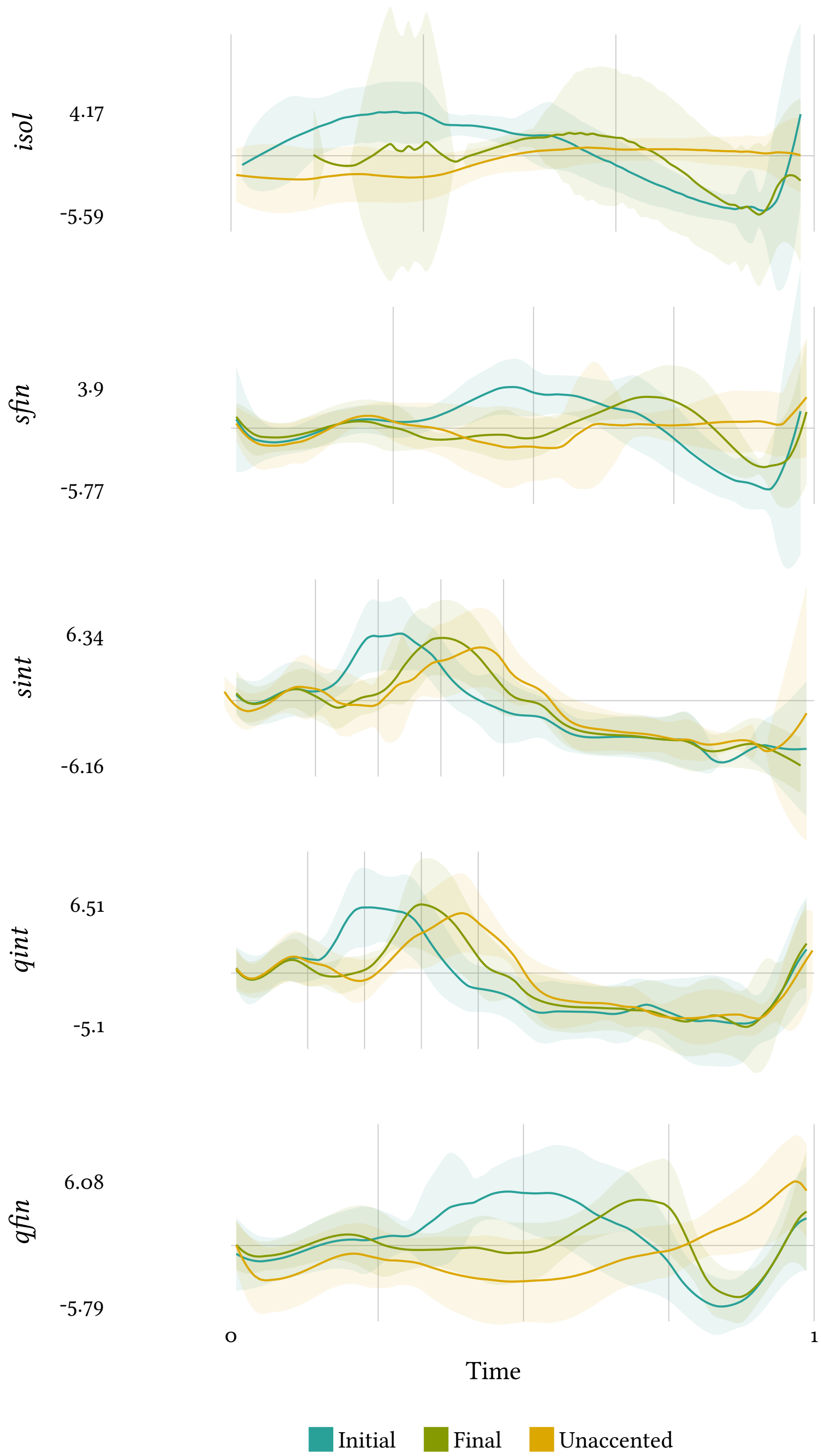


Figure 2.2: Normalised F_0 tracks of stimuli used per context for the Japanese test. F_0 values have been converted to semitones in reference to the utterance mean (marked by the horizontal line) and plotted against normalised time values ranging from zero at the beginning of the utterance, to 1 at its end. Syllable boundaries in the keyword are shown by vertical lines. Shaded areas mark the prediction intervals (see the text). Compare with figure 2.1.

ment (*i.e. isol* and *sfin*), they show a flat tone that spans the entire word. If any change is evident, it is in the small rise that takes place in the first half of the second syllable, as predicted by the LH(H) pattern reported in the literature. When unaccented words are in final position in a question (*i.e. qfin*), they exhibit a steady rise that begins at the end of the first syllable (the position of the L) and ends at the H% tone, at a height that is comparable for words of all accent types. This is a pattern that is very different from the shape of accented words in this same context (with a fall and a rise in quick succession), and of the flat shape that unaccented words had in final position in declaratives. Finally, when the word was in a non-final position, regardless of whether it was of an interrogative or a declarative statement (*i.e. sint* and *qint*), there were no obvious differences between accented and unaccented words, other than the position of the fall (and therefore, of the F0 peak-like shape).

Steady pitch rise at end of interrogatives

Resembles accented patterns in non-final positions

Unaccented H tones are lower

As reported by *e.g.* Vance (1995), the height of the H tones in unaccented words in final position of declaratives was also significantly lower those in accented words in the same contexts, particularly considering the pitch range of the word in question. This height difference is not evident when the word is in final position of a question (*qfin*), but this H is not the tone of the word, but the boundary tone of the phrase, which explains the lack of difference between different accent types.

2.3.2 Duration

For the analysis of duration, each utterance was hand annotated using Praat, with annotations at a segmental and syllabic level. Syllable duration was measured in milliseconds and then normalised by calculating the proportion of the word that the syllable occupied. These duration ratios were calculated per utterance, and results were later averaged together across all speakers.

The result of this calculation, comparing accented and unaccented syllables for each of the three syllables in the recorded target words, is shown first in tables 2.6 and 2.7 for both languages. These tables show how many times longer or shorter on average each individual syllable became when accented, both overall and by sentence context. When calculating these values, care was taken to only compare like with like: *e.g.* in Spanish, duration values for the first syllable of /ˈlimite/ were compared with the first syllables of /liˈmite/ and /limiˈte/, and so on.

The calculated duration ratios were also used as the dependent variable in an ANCOVA with main factors for syllable-prominence, accent type, syllable position, and sentence, as well as their interactions. Since some measurements were produced by the same speaker, this was added as a covariate.

A note on terminology

In the analysis of the Japanese data, the effort was made to maintain comparability with the Spanish data described above by using largely the same methods and vocabulary. However, it bears reminding that the same words in both sections will have slightly different interpretations depending on the language they refer to. When referring to syllable positions, the word "ultima" will refer in Spanish to the final syllable in the 3-syllable target word; whereas in Japanese it will refer to the syllable corresponding to the grammatical particle that was added after the end of the 2-syllable target, and therefore to the final syllable of the 3-syllable stimulus. Likewise, when referring to unaccented words, the term "prominent syllable" might still be used even if in these there is strictly speaking no prominence. In this case also, the term will refer to the syllable corresponding to the grammatical particle, which is not part of the target word itself. Besides simplifying the analysis, the rationale for talking about the "accented syllable" in unaccented words is that, as explained in section 2.3.1, this syllable was expected to be followed by a fall, and therefore to be accented. This is indeed what was evident in words in non-final positions within the utterance (see figure 2.2).

Spanish

Duration is one of the three main acoustic cues of Spanish, with accented syllables being traditionally reported as being longer than comparable un-

Context	Syllable		
	First	Second	Third
isol	1.23	1.32	1.21
sfin	1.45	1.38	1.16
sint	1.43	1.36	1.28
qint	1.43	1.33	1.34
qfin	1.39	1.42	1.30
$\bar{x}$	1.39	1.36	1.26

Table 2.6: Syllable duration for the Spanish materials as a ratio of accented to unaccented, for all sentence contexts. Compare with data for the Japanese materials in table 2.7, on page 92.

Stressed syllables are longer

accented ones. Results from the ANCOVA showed that this was the case with the recorded materials, with a significant main effect of syllable accentedness ($F_{(1,681)} = 678.63$, $p < 0.001$) and accented syllables occupying on average 10% more of the keyword than unaccented syllables ($\bar{x}_{\text{accented}} = 0.4$, $\sigma = 0.07$; $\bar{x}_{\text{unacc.}} = 0.3$, $\sigma = 0.07$). This did not mean that accented syllables were always the longest syllable in the word, only that they were longer than the same syllable when unaccented. In fact, the only case where accented syllables were always the longest within the word was with oxytone words ($\bar{x}_{\text{accented}} = 0.46$, $\sigma = 0.05$; $\bar{x}_{\text{unacc.}} = 0.27$, $\sigma = 0.05$), and in this case the length of the ultima is influenced by other factors as well, like final lengthening.

Final syllables are longer...

...and more so in utterance-final positions

The effects of final lengthening indeed made final syllables longer than syllables in other positions ($F_{(2,681)} = 279.03$, $p < 0.001$; $\bar{x}_{\text{final}} = 0.39$, $\sigma = 0.07$; $\bar{x}_{\text{rest}} = 0.3$, $\sigma = 0.07$). A significant 2-way interaction between syllable and word position also showed that the effect of final lengthening was stronger for words in final positions (*i.e. isol*, *sfin*, and *qfin*; $F_{(8,681)} = 58.84$, $p < 0.001$). This interaction is shown in figure 2.3.

Syllable position and accent type interaction

A 2-way interaction between syllable position and accent type was also significant ($F_{(3,681)} = 3.24$, $p < 0.05$), and is shown in figure 2.4. The figure shows that in proparoxytone and paroxytone words, the accented syllable and the ultima had similar durations, while the unaccented syllable had a significantly shorter duration. It also shows how in the case of oxytone words, the length of the ultima combines the effects of its position

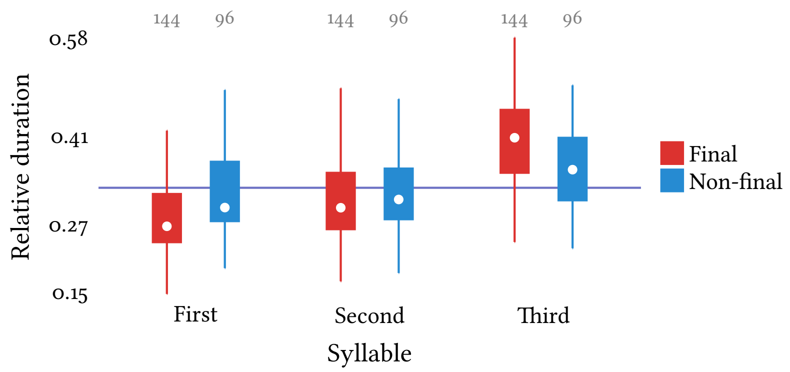


Figure 2.3: Proportion of the Spanish words in each syllable for words in utterance final and non-final positions. The violet ● line indicates the point at which syllables last for a third of the whole word. Numbers above each box indicate the corresponding *n*.

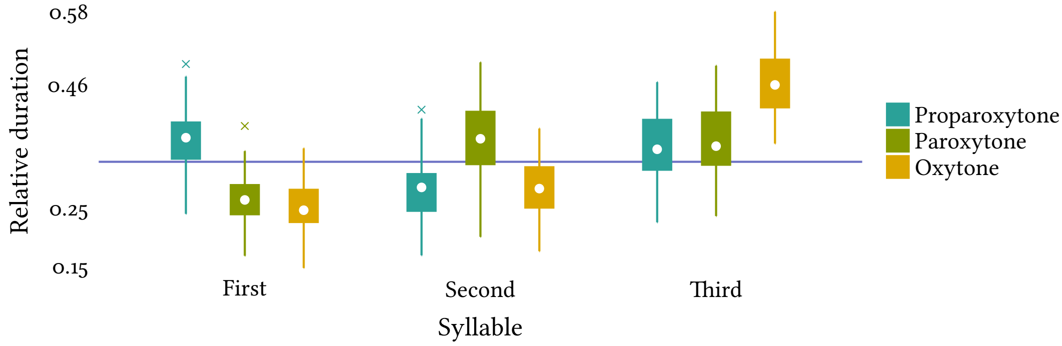


Figure 2.4: Proportion of the Spanish words that belonged to each syllable broken down by accent position. Accented syllables are longer than unaccented ones, but those of the first two syllables are as long as the ultima when unaccented. When accented, the ultima is longer than any other. The violet ● line indicates the point at which syllables last for a third of the whole word. (n =80 per box). Syllable and accent

at the end of the word, and that of the prominence.

Syllable prominence and sentence interaction

Finally, a 2-way interaction between sentence type and syllable-prominence was also significant ($F_{(1,681)} = 4.27$, $p < 0.05$), with the difference between the duration of accented and unaccented being greater in words in questions ($\bar{x}_{\text{accented}} = 0.41$, $\sigma = 0.07$; $\bar{x}_{\text{unacc.}} = 0.3$, $\sigma = 0.06$) than in words in declaratives ($\bar{x}_{\text{accented}} = 0.39$, $\sigma = 0.07$; $\bar{x}_{\text{unacc.}} = 0.3$, $\sigma = 0.07$). This interaction is shown in figure 2.5.

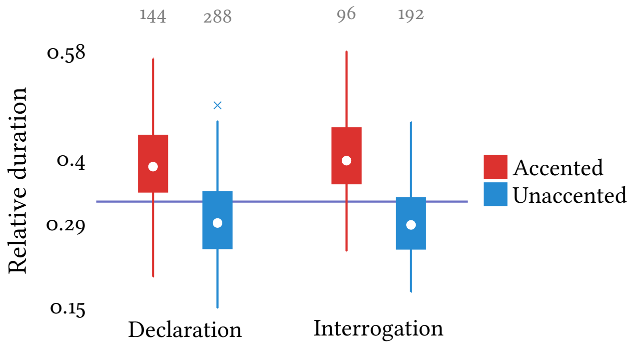


Figure 2.5: Proportion of the Spanish words that belonged to accented and unaccented syllables broken down by sentence type. The violet ● line indicates the point at which syllables last for a third of the whole word. Numbers above each box indicate the corresponding *n*.

Context	Syllable		
	First	Second	Third
isol	0.99	1.01	0.96
sfin	0.93	0.97	0.95
sint	0.95	0.98	0.96
qint	0.99	0.99	0.99
qfin	0.98	0.95	1.02
$\bar{x}$	0.97	0.98	0.98

Table 2.7: Syllable duration for the Japanese stimuli as a ratio of accented to unaccented, for all sentence contexts. Compare with data for the Spanish stimuli in table 2.6, on page 90.

Japanese

Accented syllables marginally *shorter*

Although there is a duration contrast in Japanese, it exists at a segmental level and is not used in the discrimination of lexical prominence, so no significant duration differences between accented and unaccented syllables were predicted. However, a significant main effect of syllable accentedness was found ($F_{(1,502)} = 12.66$, $p < 0.001$), but in the opposite direction of Spanish, with accented syllables being actually marginally shorter than those that bore the accent ($\bar{x}_{\text{accented}} = 0.33$, $\sigma = 0.07$; $\bar{x}_{\text{unacc.}} = 0.34$, $\sigma = 0.07$).

As with the Spanish data above, table 2.7 shows how many times longer each syllable was for each sentence context as well as overall, and provides an overview of the duration difference between accented and unaccented syllables in the Japanese materials. The table shows that the difference that turned out to be significant is, for lack of a better word, of little significance, since there was practically no difference in syllable duration regardless of whether they were accented or not. This pattern exists across all sentence contexts, with the exception perhaps of the duration of the ultima in *qfin* context.

Final syllables are longer...

A main effect of syllable position was found ($F_{(2,502)} = 213.44$, $p < 0.001$), with final syllables overall being longer than the rest ($\bar{x}_{final} = 0.38$, $\sigma = 0.08$; $\bar{x}_{rest} = 0.31$, $\sigma = 0.05$). The factor for syllable position also interacted significantly with that of word position ($F_{(2,502)} = 512.1$, $p < 0.001$). The interaction shows that the lengthening of final syllables only takes place in words in utterance-final positions, while words in non-final positions show a decrease in syllable length as the word progresses, from $\bar{x} = 0.37$, $\sigma = 0.04$ for the initial syllable to $\bar{x} = 0.3$, $\sigma = 0.03$ for the ultima. A 3-way interaction between these factors and sentence-type was also found to be significant ($F_{(2,502)} = 43.34$, $p < 0.001$). Figure 2.6 shows the proportion of the keyword in each syllable broken down by word position and sentence type, and reveals that, while there is no difference for non-final words in declaratives and interrogatives, words in final positions of questions have significantly longer final syllables than final words in declaratives, which is also evident in table 2.7.

...but only in utterance-final positions...

...and more so for interrogatives

An interaction between syllable and sentence type was also significant, but added nothing to the explanation of the data.

2.3.3 Intensity

Intensity was measured as the mean energy in dB in the nuclear vowel in each syllable using Praat. All sounds were normalised to a common RMS value using the software `normalize`[7] before analysis took place.

Direct comparisons of intensity were complicated by differences in the recording of the stimuli. In order to account for this variation, segment-level annotations were used to extract the nuclear vowel in each syllable.

7 Version 0.7.7, available at http://normalize.nongnu.org/.

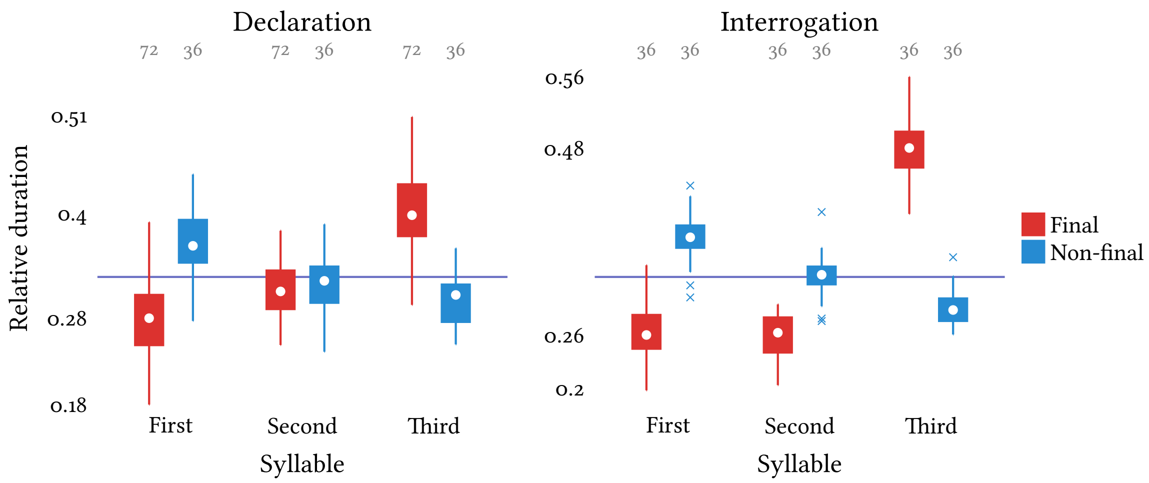


Figure 2.6: Proportion of the Japanese words that belonged to each syllable broken down by word position and sentence type. The violet ● line indicates the point at which syllables last for a third of the whole word. Numbers above each box indicate the corresponding *n*.

An utterance-specific reference intensity value was then calculated from the mean energy of the concatenation of these 3 fragments. Measures were limited to the vowels to minimise the variability that could come from segmental differences in the various minimal sets of keywords.

Mean intensity values per vowel were then divided by the reference value to obtain relative energy measures, which were used as the dependent variable. Comparisons were made using an ANCOVA with the same main factors as used in the analysis of duration. Since different vowels have different inherent intensity values, and the total number of each vowel was unbalanced in the set of stimuli, vowel was included as an additional covariate.

Spanish

Stressed syllables were louder...

...but only for *isol* and *sfin*

Results from the ANCOVA for intensity showed a significant main effect of syllable prominence ($F_{(1,677)} = 183.982$, $p < 0.001$, with stressed syllables having marginally higher relative energy values ($\bar{x}_{\text{accented}} = 1.01$, $\sigma = 0.05$; $\bar{x}_{\text{unacc.}} = 0.99$, $\sigma = 0.06$). This factor had significant 2-way interactions with word position ($F_{(1,677)} = 60.89$, $p < 0.001$) and sentence type ($F_{(1,677)} = 115.35$, $p < 0.001$), as well as a significant 3-way interaction with both of these ($F_{(1,677)} = 46.88$, $p < 0.001$). Figure 2.7 shows the data for this interaction, and shows that the difference between accented and un-

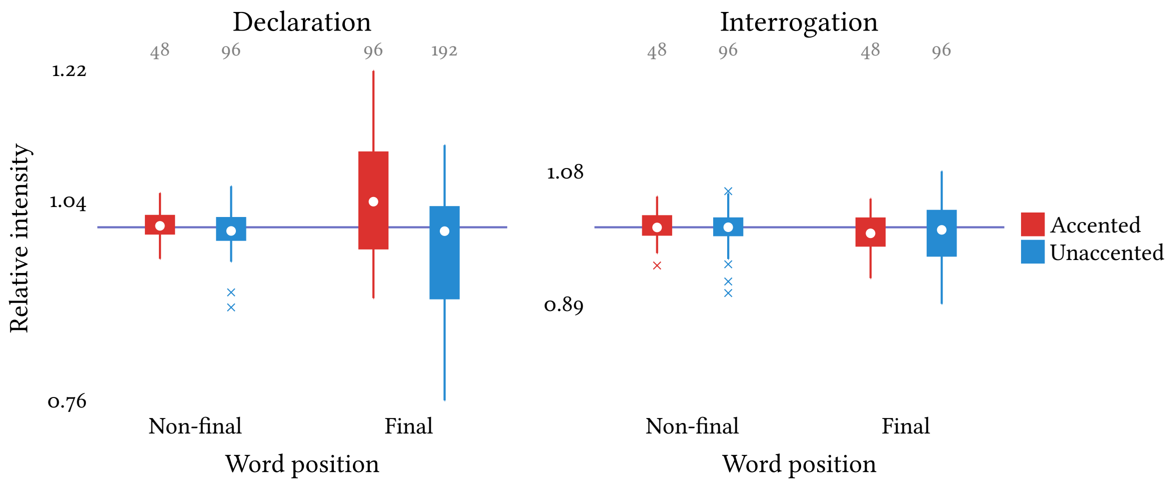


Figure 2.7: Differences in the distribution of relative energy for accented and unaccented syllables in the Spanish materials depending on word position and sentence type. Energy is calculated as the mean energy of the syllable nucleus, divided by the mean energy of the three nuclei in the keyword (both in dB). The violet ● line indicates this reference value, set to 1. Numbers above each box indicate the corresponding *n*.

accented syllables exists only for words in final positions of declaratives (*i.e. isol* and *sfin*). All other sentences show no difference between the intensity of accented and unaccented syllables. A significant main effect of word position was also found ($F_{(1,677)} = 14.43$, $p < 0.001$).

There was also a significant main effect of syllable position ($F_{(2,677)} = 161.44$, $p < 0.001$), since there was a steady decrease in syllable intensity as the word progressed, from $\bar{x} = 1.03$, $\sigma = 0.04$ for the initial syllable to $\bar{x} = 0.96$, $\sigma = 0.06$ for the ultima. However, this is only true for words in final positions, as shown by a significant 2-way interaction between syllable and word position ($F_{(2,677)} = 354.52$, $p < 0.001$). A significant 3-way interaction between these two and sentence type ($F_{(2,677)} = 65.75$, $p < 0.001$) showed that the intensity decrease in words in final position of declaratives was much steeper ($\bar{x}_{\text{first}} = 1.06$, $\sigma = 0.05$; $\bar{x}_{\text{last}} = 0.9$, $\sigma = 0.05$) than that of words in final position of questions ($\bar{x}_{\text{first}} = 1.02$, $\sigma = 0.03$; $\bar{x}_{\text{last}} = 0.96$, $\sigma = 0.03$). This is shown in figure 2.8. A significant interaction between syllable position and sentence type was also found to be significant ($F_{(2,677)} = 126.58$, $p < 0.001$).

Initial syllables were louder...

...in words in final positions...

...and more so for those in declaratives

The source of the large spread in the intensity values for accented syl-

Syllable × accent interaction

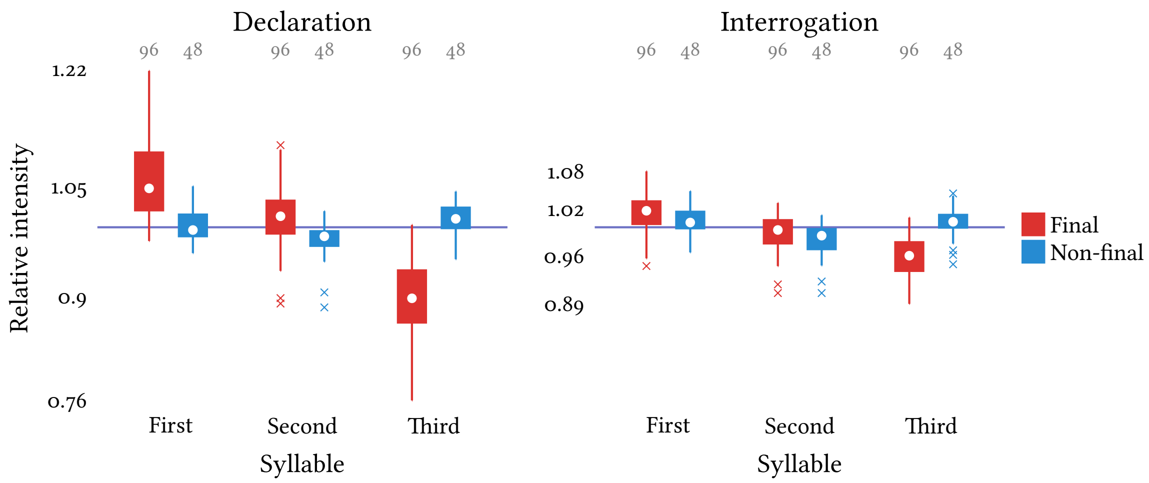


Figure 2.8: Differences in the distribution of relative energy for each syllable in the Spanish materials depending on word position and sentence type. Energy is calculated as the mean energy of the syllable nucleus, divided by the mean energy of the three nuclei in the keyword (both in dB). The violet ● line indicates this reference value, set to 1. Numbers above each box indicate the corresponding *n*.

lables in declarative-final positions shown in figure 2.7 is easier to understand by breaking that data by syllable. Figure 2.9 shows the data for the intensity interaction between syllable and accent position by position within the sentence as well as accent type. This 4-way interaction was found to be significant ($F_{(3,677)} = 4.98$, $p < 0.01$), and as the figure shows, this is due to syllables always being loudest when accented, but in a significant way only for words in final positions of declaratives. The interaction between this and the steady decrease in intensity as the word progresses explains the concentration of data points around the reference value in figure 2.7.

Lower-level interactions between these factors were also found, but do not add anything to the explanation of the data.

Japanese

Accented syllables louder

Results from the ANCOVA for the Japanese intensity data showed a main effect of accentedness ($F_{(1,500)} = 12.84$, $p < 0.001$), with accented syllables having greater intensity values than unaccented syllables ($\bar{x}_{\text{accented}} = 0.99$, $\sigma = 0.11$; $\bar{x}_{\text{unacc.}} = 0.97$, $\sigma = 0.1$). The difference, however, is very small.

Lower intensity for initial syllables

There was also a significant main effect of syllable position ($F_{(1,500)} = 27.84$, $p < 0.001$), with initial syllables having a significantly

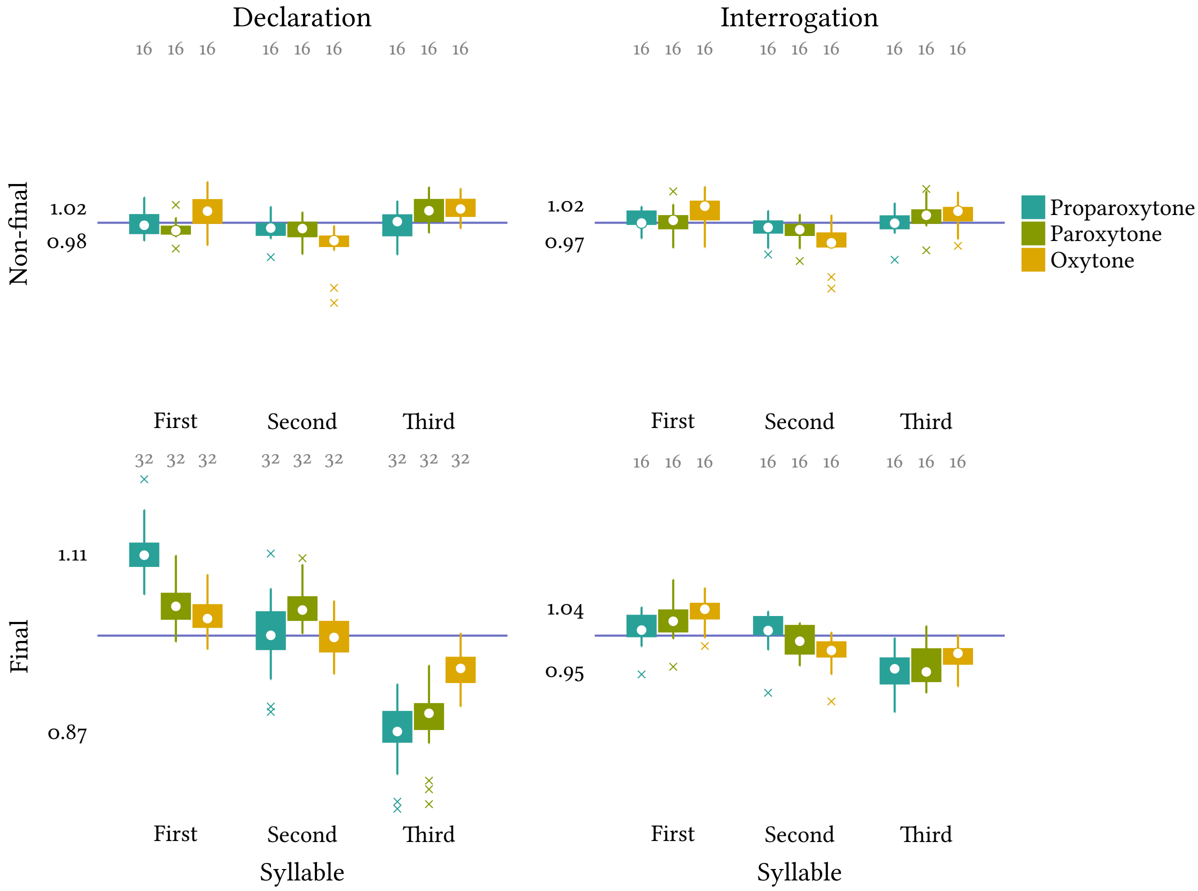


Figure 2.9: Differences in the distribution of relative energy for each syllable in Spanish words of each accent type. Data is broken down by word position and sentence type. Energy is calculated as the mean energy of the syllable nucleus, divided by the mean energy of the three nuclei in the keyword (both in dB). The violet ● line indicates this reference value, set to 1. Numbers above each box indicate the corresponding *n*.

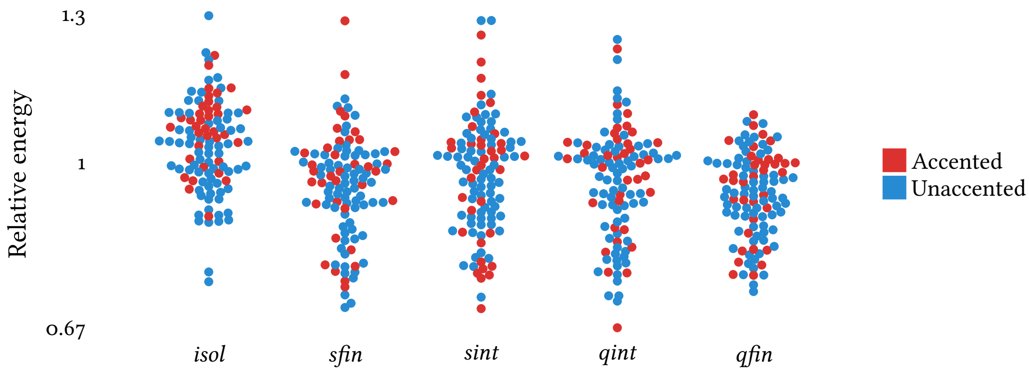


Figure 2.10: Distribution of relative energy for accented and unaccented syllable nuclei per sentence type. The reference value is set to 1. Energy is calculated as the mean energy of the syllable nucleus, divided by the mean energy of the two nuclei in the keyword plus that of the final particle (all in dB). The horizontal axis is only used to separate data points.

lower intensity than the rest ($\bar{x}_{\text{initial}} = 0.93$, $\sigma = 0.1$; $\bar{x}_{\text{rest}} = 1$, $\sigma = 0.1$).

Lower intensity in questions

A main effect of sentence type was also found ($F_{(1,500)} = 25.19$, $p < 0.001$), with interrogative sentences having slightly lower intensity values ($\bar{x}_{\text{interr.}} = 0.95$, $\sigma = 0.1$; $\bar{x}_{\text{declar.}} = 0.99$, $\sigma = 0.11$). Breaking these data down for each of the five sentences provides some additional information, by showing that most of that variation is due to isolated words having significantly greater intensity values overall, as shown in figure 2.10. This explains also the lack of a main effect of word position, since the combination of the data from the isolated context and that of *sfin* sentences masked this increase. There was a 2-way interaction between word position in the sentence, and sentence type ($F_{(1,500)} = 6.7$, $p < 0.01$), which largely confirms what is shown in figure 2.10, in that words in final positions of declarative sentences had the greatest intensity values ($\bar{x} = 1$, $\sigma = 0.11$), while words in final positions in interrogatives had the smallest ($\bar{x} = 0.94$, $\sigma = 0.08$). There was no difference between sentence types for words in non-final positions, which had intensity values between the extremes reported above ($\bar{x}_{\text{declar.}} = 0.98$, $\sigma = 0.11$; $\bar{x}_{\text{interr.}} = 0.96$, $\sigma = 0.11$).

Greater intensity for isolated words

Intensity difference for questions by position

2.4 Discussion

2.4.1 Spanish

Based on the results from the statistical tests, the collected speech materials can be grouped in three distinct categories, largely determined by the three different realisations of the stress that were found. Words in final position in declarative statements can be said to be the prototypical example of a Spanish accent, and should therefore be the easiest for non-native listeners. They have F_0 contours that trace primarily L+H* pitch accents and show a clear distinction in both the relative energy and the duration of accented and unaccented syllables. Since earlier syllables are significantly louder than the rest, L2 learners who rely on intensity cues should find it easier to perceived accents occurring towards the beginning of the word, and might even be biased towards perceiving accents as earlier than they are. Words in non-final positions tend to have accents marked by L*+H F_0 patterns and do not show an intensity contrast between accented and unaccented syllables (although duration remains present). The combination of a non-standard pitch accent and the lack of intensity contrasts should make these contexts harder than the first group, but the persistent presence of a F_0 peak-like contour (albeit displaced from the location of the accent) should counter some of this difficulty. Finally, words in final position in interrogations have accents marked by a single L* target followed by the H% boundary tone and also lack an intensity contrast (although duration is still available). This F_0 pattern, which is not only radically different from that used in Japanese — but also far from the standard realisation of Spanish stress — should pose the highest challenge for L2 learners (as it did in Kimura et al., 2012). Since this context does not show an intensity contrast between accented and unaccented syllables, the possible bias that was predicted for earlier accents in the other statement-final contexts is not predicted to occur in this case.

isol+sfin
L+H*
+intensity
+duration

sint+qint
L*+H
−intensity
+duration

qfin
L* H%
−intensity
+duration

Table 2.8 presents a summary of these findings for Spanish.

Accent	Feature	Sentence				
		isol	*sfin*	*sint*	*qint*	*qfin*
Oxytone	Pitch	L+H*	L+H*	L*+H	L*+H	L* H%
	Duration	SSL	SSL	SSL	SSL	SSL
	Intensity	HML	HML	-	-	-
Paroxytone	Pitch	L+H*	L+H*	L*+H	L*+H	L* H%
	Duration	SLL	SLL	SLS	SLS	SLL
	Intensity	HHL	HHL	-	-	-
Proparoxytone	Pitch	L+H*	L+H*	L*+H	L*+H	L* H%
	Duration	SSL	LSL	LSS	LSS	LSL
	Intensity	HML	HML	-	-	-

Table 2.8: Spanish acoustic profiles per accent position and sentence context. Values for duration indicate long (L) and short (S) syllables, while those of intensity do so for low (L) and high (H). Values for F_0 use AM notation.

2.4.2 Japanese

Acoustic analysis of the Japanese data shows that, as expected, the largest differences in the realisations of the various accent types were found in the shape of the F_0 contour. Of particular interest was the variation in the shapes of unaccented words, which seemed to be significantly more susceptible to the effects of intonational context than accented words.

Indeed, although no differences in the F_0 contour of accented words was found across sentences, the data for unaccented words makes it possible to separate the materials into the same three groups that were found with the Spanish materials: final position in declaratives, final position in interrogatives, and non-final positions.

Flat contour in *isol*+*sfin*

As accented in *sint*+*qint*

Rising in *qfin*

When in final position of declarative statements, unaccented words were produced with either a completely flat tone, or a tone in which the only difference was a small increase between the first and the second syllables, from an L to a lowered H (see Vance, 1995). When in non-final positions, on the other hand, the differences that existed between accented and unaccented words disappeared, not only in the general shape of the F_0 curve, with a much clearer peak-like shape, but also with H tones that now reach the same level as those in other accented words. And when unaccented words were in final positions of interrogative statements, they started from an L tone and steadily increased to an H% target at the end of the word. This target was the same for unaccented and accented words, which makes sense since it is the target of the sentence, and not the word. But it is interesting that while accented words maintained their shape despite the need to reach that H% including the cases were a final fall was required, unaccented words showed no tendency to maintain the flat contour that was so characteristic in the declarative utterances.

The robustness of the Japanese accent has been mentioned before in this dissertation, but unaccented words, in contrast, show a remarkable malleability, in the sense that they seem to offer very little resistance to the effects of sentence intonation.

Other cues

Traditionally, pitch has been described as being the only meaningful acoustic cue of the Japanese accent (Labrune, 2012), and the data presented here certainly supports the idea that pitch concentrates the largest part of the variance. But it also showed differences in both intensity and duration, that while small, were statistically significant. Because those differences are small, these results should not be interpreted to mean that duration and intensity should also be considered part of the inventory of acoustic cues of Japanese accent, nor should they be taken to imply that native speakers' perception is necessarily influenced by them in any way. But the fact that they were significant suggests that the amount of variation that existed, and that in all likelihood is entirely ignored by native speakers, is not entirely negligible. This is relevant because, although native speakers know to ignore these differences, nothing guarantees that non-native speakers will know to do the same, particularly those for whom these cues are meaningful.

Predictions

Based on the results presented in this section, two main predictions can be made.

1 Since in non-final contexts unaccented words would appear to behave much

like another kind of accented word, this will render them generally less acoustically salient. By doing so, differences in the contrast will become generally harder to perceive, which would result in non-native participants showing lower performances overall for these contexts.

2 At the same time, as the difference between accented and unaccented words is reduced, the effects of that difference, whether beneficial or detrimental (see the research questions above), should also diminish.

2.4.3 Speaker variability

To simplify the presentation of the results, this chapter has not considered the issue of speaker variability in the analysis of these productions, but the topic deserves some attention. Even with the limited sample sizes used in the preparation of these materials, there were aspects that showed considerable amounts of variation. Figures C.1 to C.5, in the appendix, show the effect of this variation on F_0 for the Spanish materials, while figures C.6 to C.10 do so for Japanese.

Particular among these are the differences in the preferred Spanish pitch accent used by some speakers for specific sentence types. As explained above, productions of *qint* sentences had a predominantly $\overset{*}{\text{L}}$+H pitch accent, and this was true for all participants. However, the younger female speaker produced very clearly marked F_0 valleys between the H in that accent and the H% at the end of the utterance. This is shown in figure C.4, on page 270. Likewise, the $\overset{*}{\text{L}}$+H pitch accents that were common for words in *sint* sentences, did not occur for productions of the older female speaker, who instead chose to systematically produce L+$\overset{*}{\text{H}}$ (or perhaps L+>$\overset{*}{\text{H}}$) pitch accents. This is shown in figure C.3, on page 269.

There were also differences in duration. An example of this in Spanish is the productions of isolated words by the older male speaker, who showed significantly less distinct duration contrasts with accents towards the end of the word.

These cases are notable in that they exemplify the type of variation that can exist, and they will become relevant in chapter 3, when their effects on the perception of non-native participants is examined. But they also serve to illustrate the fact that this variation has a limited scope, in that while these speakers showed a non-standard behaviour in these contexts, they did not show any significant difference when producing words in final positions. Nor was there any similar systematic difference for the Japanese speakers. This highlights the need to include larger number of participants when collecting materials, and the importance of considering what sort of variation can be expected under the specific conditions that are being tested.

2.4.4 Cue reliability and perceptual plasticity

In more general terms, the differences that were found in the realisations of Spanish stress, even in such a limited sample as the one used in this case, highlights the trade-off that takes place between cue availability and cue reliability. In the case of Spanish, lexical prominence has three main acoustic correlates: pitch, duration, and intensity. As the number of available cues increases, so does the number of combinations that will be available to the speakers, and these combinations will depend on a number of factors including individual differences, the immediate prosodic context of the word, the specific communicative effect that the speaker wants to provoke, etc.

The higher variability in the realisations of the Spanish stress means that native speakers of that language will likely benefit from a higher degree of plasticity, that allows them to map a number of different realisations of stress to the existing accentual categories. For Japanese speakers, however, the opposite is true: since this language has relatively few meaningful acoustic cues, what will be beneficial for speakers of that language will be the ability to ignore the inevitable and meaningless variation in those cues. It is unclear at this point whether this difference will have a perceivable effect in the two populations, and if it does what its magnitude will be. However, the bidirectional approach used in this dissertation is a good tool to explore such questions.

Chapter 3

Perception of L2 Spanish lexical prominence

This chapter presents the results of a study on the perception of Spanish lexical stress by native speakers of Japanese using the materials described in chapter 2. As explained in the previous chapter, the methodology in this study is an extension of that used in Kimura et al. (2012), a summary of which was presented in chapter 1. That study will on occasion be referred to as "the original study", to avoid needless repetitions.

3.1 Design

Research questions

Like in the original study, the study presented here also focused on a population of Japanese learners of Spanish, and attempted to answer the following questions:

1 How robust are the results found in Kimura et al. (2012)?

1.1 Are their results specific to the variety of Peninsular Spanish used in that experiment? Or will they generalise to at least one other variety?

1.2 Considering a single speaker was used in the original study, would those results generalise to other speakers?

2 What is the relationship between participant proficiency in the L2 and the results they found?

An early version of the results presented in this chapter were presented at the 59th meeting of the Congress of the Japanese Society of Hispanists (AJH) in Tōkyō, Japan, in the fall of 2013. The results presented here offer a correction on the contents of that presentation. Digital materials related to this study can be found in the study's repository at http://www.pinguinorodriguez.cl/research/phd/sentence-context/spanish.

Research variables

The answer to the first questions required an extension of the methodology of Kimura et al. (2012) rather than a simple replication. Methodological decisions were kept whenever possible, but the number of speakers was increased to 4, and a different variety of Spanish was used. See chapter 2 for a full description of the changes that the speech materials underwent, and the properties of the final stimuli.

As for the second main question, a measure of L2 proficiency was necessary. This was obtained through a linguistic background questionnaire in which participants provided an estimation of their own linguistic proficiency. The questionnaire covered a number of aspects of the learner's linguistic background (such as their place of origin, length of studies, and length of residence in a place where the L2 was spoken), but contained two questions that directly focused on the participant's L2 proficiency. The first asked the participant to give a self-rating on what they perceived to be their language proficiency in a 5-point scale (with 1 being "beginner" and 5 being "advanced") for writing, reading, speaking and listening per separate. The second reproduced the 5-point scale to ask participants to give information on their L2 use in a number of different contexts: at home, with their friends, when consuming media, at their place of work or study, or for personal pleasure.

An illustration of the structure and items in this questionnaire can be seen in appendix A on page 253, while a discussion of the possible shortcomings of this method is presented in section 1.4.1. It must be noted that, since the questionnaire was presented to the participants in their native language, this illustration (presented in English for the reader's convenience) differs from the questionnaire actually presented to the participants. However, the items shown and their order in the questionnaire is the same.

The testing platform

Recruitment of participants in the populations targeted by this study presented a number of difficulties. There were, of course, the geographical limitations of recruitment in Japan and Chile, two countries that are almost exactly on different sides of the globe. But another and perhaps greater difficulty was that the populations themselves are not so large, meaning that recruitment would have to draw from a reduced pool of possible participants.

In order to facilitate the recruitment process and the access to the largest possible number of participants in both countries, the test was designed to be hosted online and administered (either locally or remotely) through a computer connected to the internet. Participant responses would then be automatically collected in a central location in Germany, regardless of where the actual test took place, enabling participants to take the test in their own countries with or without external supervision.

LimeSurvey

This was done using the free software LimeSurvey[1]. This PHP-based software allows for the development of cross-platform tests that run consistently on any graphical internet browser, which, considering their massively widespread use, would greatly improve the accessibility to the test. This also meant that participants would not be forced to download any new software to whatever computer they happened to use for the test, and that the researcher would not have to distribute standalone versions of the test, as would have been the case with some of the other alternatives that were considered. Additionally, this web-based test proved to be more versatile, allowing for both remote and local data-collection, which was deemed to be a desirable feature in terms of future-proofing (*i.e.* if the procedures of the test were changed in the future).

Issues with online testing

Internet-based data collection for language sciences has seen a tremendous surge in popularity in recent years, which has understandably resulted in a number of studies regarding the validity of online data (*e.g.* Paolacci and Chandler, 2014; Sprouse, 2011; Fort, Adda, and Cohen, 2011, among others). Although most of their attention has been focused on specific recruitment platforms due to their popularity, some of the conclusions they reveal can be expected to apply to other sources of remote data collection.

In general, the choice between internet-based testing and traditional

1 Version 1.92+, available at `http://www.limesurvey.org`.

laboratory testing presents a trade-off between the ease of access to participants afforded by the former, and the increased control over the testing conditions provided by the latter. However, research has shown that results from both sources are comparable, despite the fact that an increase in data variability (Paolacci and Chandler, 2014) and a fractionally higher participant rejection rate can cause the results of online testing to have a marginally lower statistical power (Sprouse, 2011), when participant numbers are equal.

Taking this into consideration, data collection for the study reported in this and the following chapter was done exclusively using remote testing, in the hopes that that would increase the number of participants. As an aside, since one of the objectives of this study was to examine the results of previous experiments under more variable conditions to test the robustness of those effects, the increased variability of data from online sources was not seen as a strong enough methodological deterrent. Still, to further address some of the problems of remote data collection outlined above, additional questions were added to the proficiency questionnaire to collect as much information as possible on the participants and their test environment, including their place of origin, length of residence in the target areas, etc.

3.1.1 Participants

Japanese participants: J1S2

Japanese participants were primarily recruited from the student body enrolled in the 4-year Spanish programs at Seisen University and Tōkyō University of Foreign Studies, both located in Tōkyō. The test was advertised by local professors, and students were encouraged to forward the test invitation to others they thought suitable, which reduced the control over the precise place of origin of the participants. Collaborating professors were allowed freedom to choose the best method to motivate their students to participate, which included extra credit or other such benefits. A total of 35 (25 female and 10 male) participants completed the test. Participants reported their age in terms of age categories, since it was thought that this

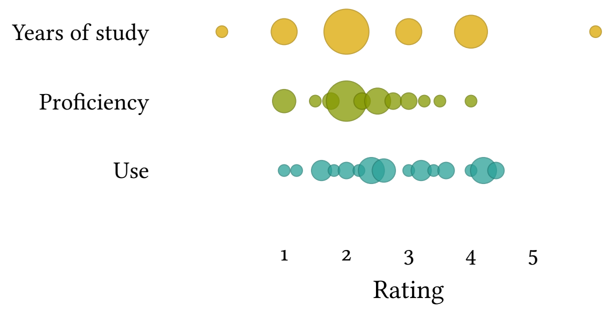


Figure 3.1: Self-reported measures of L2 proficiency of students of Spanish in the first experiment. The horizontal position of each circle marks the mean score reported by each participant to the three scales. In the case of proficiency and use, participants were given a 5-point scale (1 being "beginner"; 5 being "advanced"), while they were free to enter how many years they had studied for. The area of each circle reflects the number of participants for that particular score.

would increase the willingness of participants to report their age accurately, which meant that it was not possible to provide mean values. In lieu of that, table 3.1 shows the distribution of participants by age bracket, and shows that the large majority reported being between 15 and 24 years of age, as expected of a population sampled predominantly from university students.

Figure 3.1 shows the distribution of participants according to the 3 self-reported measures of second language proficiency: length of studies, overall proficiency, and frequency of use. While participants seemed to be spread more or less evenly in terms of length of studies and self-reported language use, most of them reported having an intermediate-to-low level of proficiency. The questionnaire also collected information about the participants' place of origin, and although the original intent was to limit participants to those whose place of origin was in the Kantō region to control for regional variation, preliminary analysis of the results showed this not to be a significant factor.

Chilean participants

Chilean controls were recruited from a broader pool, including students from Universidad Católica de Chile and Universidad Andrés Bello, both of them in Santiago, and from Universidad de Concepción in Concepción. A total of 28 (14 female and 14 male) Chilean participants completed the test.

Age group	Native	J1S2
15–19	3	14
20–24	6	18
25–29	8	
30–34	5	1
35–39	1	
40–44	1	
45–49	1	
>50	3	2
Total	28	35

Table 3.1: Age distribution for participants of the first Japanese test, according to age categories.

Compared to the Japanese participants, members of the control group had a larger age spread reflecting the broader call for participants who did not in this case need to be language learners. Despite this, most of them reported being in their late twenties, as shown in table 3.1.

An additional 24 participants (from both groups) signed up for the test but did not complete it.

3.1.2 The test procedure

Participants were instructed to wear headphones and sit in a quiet room where they wouldn't be disturbed for the 30 minutes the test was expected to take. Because the test was administered remotely, however, it was impossible to completely enforce these instructions. To sidestep this, the questionnaire tried to gather as much information as possible about the specific testing environment, so later filtering was possible. Participants were allowed to set the volume of their equipment to a comfortable level before the test began.

Stages in the test

The test had four main stages:

1 The questionnaire mentioned above, where each participant had to answer questions on their demographic and linguistic background, as well as on the environment under which the test had taken place (*e.g.* what headphones they were wearing; the condition of the room; etc).

2 A practice stage, in which trial versions of the stimuli were presented to

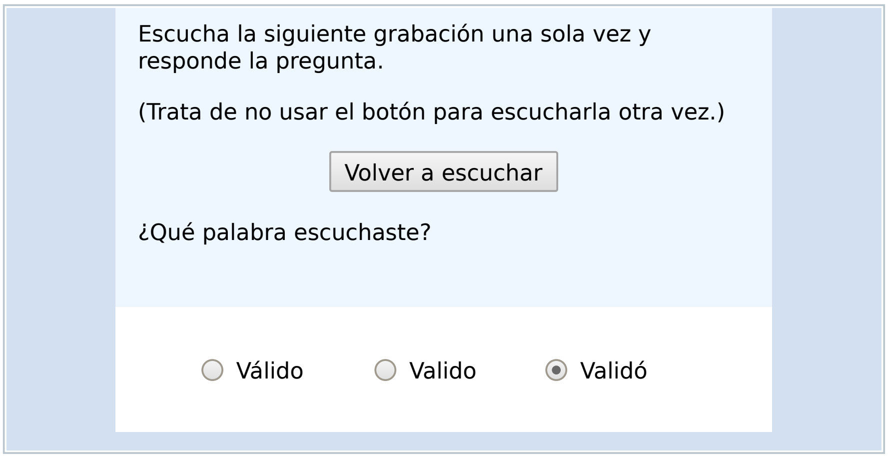


Figure 3.2: Example of a question in the Spanish test. The question text was shown in Spanish for all participants. The replay button can be seen in the middle of the question overlay. In the example, the third alternative is selected.

make sure the procedure was clearly understood.

3 The testing stage, separated into 4 different test blocks with breaks in between.

4 A final stage, during which participants were given feedback on their overall performance in terms of number of correct responses per sentence context and accent type.

Once the test was completed, participants were offered the chance to register their contact details for further contact.

Question format

Questions were presented in the same way in both the trial and the testing stage, following the example shown in figure 3.2. Participants were presented with a multiple forced choice question written in Spanish[2] using standard Spanish orthographic rules, which include a graphic mark (´) for the position of the stress[3]. The recording would play automatically as soon as the question finished loading. Responses were disabled until the sound had finished playing and participants were not allowed to continue until they had marked their response.

A button in the centre of the screen with the text "listen again" al-

2 The question text says: "Listen to the following recording once and answer the question (try not to use the replay button). What word did you hear?".

3 Depending on the letter with which a word ends in Spanish, an unmarked word can have its stress in the ultima or the penult (cf. table 1.1). However, since all keywords in this test ended with a vowel, the unmarked position was always the penult.

lowed them to replay the sound they had heard. Including this button introduced the possibility of abuse, but due to the lack of control over the testing environment and the possibility that participants might get distracted by events beyond their control, it was decided that this feature was a necessity. To curb excessive use, the button would become inactive after being used twice.

The trial stage

The trial stage consisted of six test items using stimuli that were not part of the actual testing stage. These were taken from the set of nonce words recorded for Kimura et al. (2012) to make sure they had the same basic structure as the rest of the stimuli. The first three trials were words in isolation, while the latter three were in a selection of the different sentence contexts. Participants were given feedback on their response after each trial question and allowed to continue.

The testing stage

Randomisation scheme

The testing stage had four testing blocks, with breaks in between to allow participants to rest if they wanted. During each block, stimuli were presented in a pseudo-random order, with stimuli in the isolated context presented at regular intervals and in a randomised order. In between any two isolated words, one item of each of the other sentence contexts would appear, also in a random order. This was done to make contiguous repetitions of similar contexts less likely.

A further reason to do this was to test for loss of attention in what was a relatively long and repetitive task. Since Japanese participants had been shown to have very high performance with isolated words (Kimura et al., 2012), a constant distribution of these relatively "easy" items would make it possible to identify cases in which their performance decreased throughout the test[4].

In each testing block, all 60 combinations of the carrier sentences and the keywords appeared once as spoken by one of the 4 speakers described in chapter 2. Not all stimuli from the same block necessarily belonged to the same speaker, but the sorting of speakers and stimuli guaranteed that

4 It didn't.

by the end of the fourth testing block, each stimulus had been played once with each speaker. The order in which these were presented was determined per-participant, to avoid any possible order effects.

Digital sound management

To ensure sound playback was as stable as possible, which was important since the researcher had no control over the devices used to run the test, SoundManager2[5], a web-based sound library which uses HTML5 and falls back to Flash when unavailable, was used to play the samples in each question page. This solution was tested in the largest number of computer set-ups available to the researcher at the time, and was deemed to be stable enough for the purpose of the test. Despite this, sound playback remained problematic under some set-ups, which lead to some participants reporting not being able to run the test.

Sound samples were compressed as MP3 files at 192kB/s using `lame`[6] to minimise the amount of data transferred during the test and ensure greatest compatibility. The sampling rate was left unchanged.

Hosting

The test itself was hosted using the server solution offered by LimeSurvey. Their servers hosted both the test and the stimuli, and provided the secure storage of the response database. The test servers were based in Germany, and ensured compliance with the data protection regulation in place in the European Union.

The complete XML-based definition file used by LimeSurvey, which contains all the questions and materials of the test, as well as the logic that controls its behaviour, can be found together with the rest of the relevant digital files in the study's online repository (see page 105).

3.2 Results

Participants provided responses to a 3AFC categorical identification task. In order to facilitate analysis, results were re-coded as a binary response stating whether the accent had been correctly identified in each case or not.

5 Version 2.97a, available at http://www.schillmania.com/projects/soundmanager2/
6 Version 3.99.5, available at http://lame.sourceforge.net/. Encodings were made using its `-h` flag, which ensures the highest quality available.

Data filtering

Preliminary analysis of the results showed that testing conditions among participants had been homogeneous, and even the responses of those participants that had *e.g.* used different kinds of headphones showed no significant differences to the rest of the participants. Likewise, no significant effects of participant fatigue were shown by comparing the responses given to isolated words across testing blocks. A comparison of the variance in the data from this study and those reported in Kimura et al. (2012), together with the aspects mentioned above, strongly suggested that the noise introduced into the results by testing remotely was not a cause for concern.

A generalised linear mixed-effects model, as implemented in the `lme4` package for `R`, was fitted to the data. The final model was built using a bottom-up approach, in which a baseline model with no factors is used as a baseline against which increasingly complex models are compared. When the comparison revealed the newer model to be better, this became the new baseline and the process continued until an optimal model was found. This approach was used for both fixed and random factors.

Model comparison was based on their AIC values, which penalise the fit of a model by the number of predictors that have been introduced. This means that when comparing two models that fit equally well, the one with fewer factors will be considered to be "better fitting". Additionally, the result of a χ^2 test provided information on the statistical difference between the models. A model was therefore better fitting when it provided a lower AIC value than the baseline, and was significantly different from it. These results were provided by the `anova()` function available through the `stats` package for `R`.

Results overview

Overall results for both the control and the experimental groups are shown in figure 3.3a broken down by sentence context, and in figure 3.3b broken down by position of the prominence within the word.

As can be seen, responses are similar to those reported in Kimura et al. (2012), with identification rates for native speakers showing ceiling effects throughout, and responses from non-natives varying significantly across contexts. Like in the original study responses for words in isolated contexts were highest, and those for words in the *qfin* context were lowest. A similar pattern can be seen for words with different accent positions, with native speakers reaching ceiling for all of them, and responses for non-natives showing differences depending on the position.

However, unlike the original study, J1S2 responses are considerably higher ($\bar{x}$= 79.01%overall; $\bar{x}$= 61.77%for *qfin*), never reaching below 41%.

Control responses

Responses from the control group show extremely good performance rates for all sentence contexts and accent types. Although responses for *qfin* sentences and paroxytone words show slightly more of a spread and are generally lower than the rest, these differences did not turn out to be significant.

Focus on non-native responses

Because of the great separation between both populations, and to simplify the large number of results to report, the following sections focus only on the results of the non-native participants. Likewise, unless otherwise specified, all the following figures show J1S2 responses.

3.2.1 Main effects

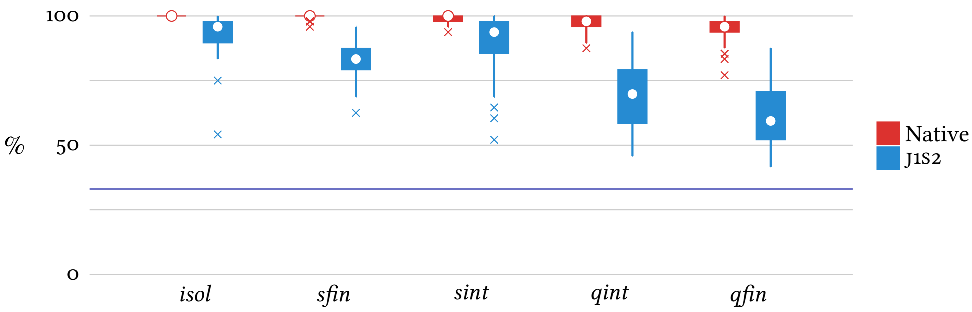
Significant factors:
accent
sentence
speaker
position×sentence
accent×sentence
accent×speaker
position×speaker
sentence×speaker
3-way interactions
4-way interaction

Models were built using accent position, word position within the sentence, sentence type (interrogative or declarative), and speaker as fixed factors. Their interactions were also considered in the model building. Apart from word position, the remaining three were found to be significant as isolated factors. Significant 2-way interactions were also found between speaker and, in turn, accent position, word position, and sentence type; and between sentence type and both accent and word position. Three-way interactions were found for sentence type and word position with speaker on one hand, and accent position on the other; and a 4-way interaction between these was also found. An overall of the results broken down by sentence and accent position, including their interaction, is shown in figure 3.3. The only random factor in the model was that for participant.

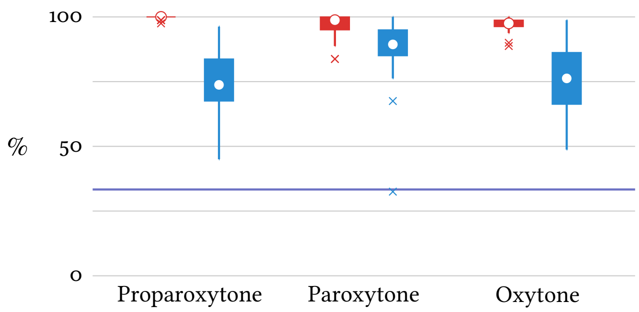
Random factor

No effect of proficiency

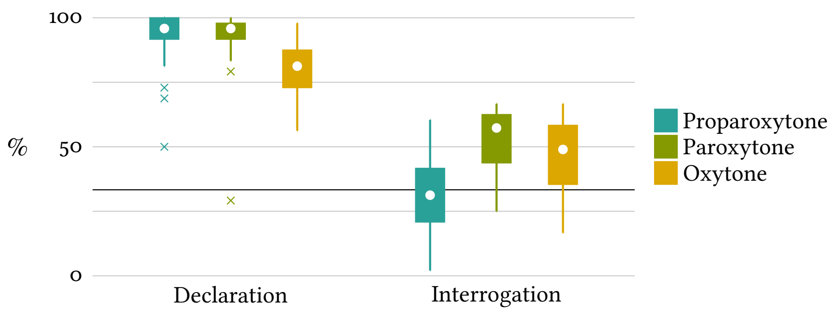
The attempt was made to introduce measures of proficiency to the model. However, this did not significantly improve the model, neither when including the measures as isolated factors nor when considering their in-



(a) Results per sentence (n =25/box for natives, n =34/box for S1J2)



(b) Results per accent (n =25/box for natives, n =34/box for S1J2)



(c) Interaction between sentence and accent (n =34/box)

Figure 3.3: Overall results of the Spanish test, broken down by sentence (top), accent (middle) and their interaction (bottom, only for non-natives). In each case, the violet ● line marks the chance level. Compare with Japanese results in figures 4.3 and 4.4, on pages 138 and 139. (n =25/box for natives; n =34/box for non-natives).

teractions, nor when aggregate measures of proficiency were used.

Results from this final model were later interpreted using Wald χ^2 tests, as reported by the `Anova()` function in the `car` package.

Effects of accent position

Accent position was significant ($\chi^2_{(2)} = 245.42$, $p < 0.001$), with correct responses for paroxytone being significantly higher ($\bar{x} = 87.9$, $\sigma = 12.19$) than those for the rest ($\bar{x} = 74.56$, $\sigma = 12.34$), among which there was no difference. These overall results are shown in figure 3.3b.

Effects of sentence type

A main effect of sentence type ($\chi^2_{(1)} = 375.44$, $p < 0.001$) was found to be significant, with declarations having significantly higher percentages of correct responses ($\bar{x} = 88.39$, $\sigma = 9.88$) than questions ($\bar{x} = 65.13$, $\sigma = 12.74$). Figure 3.3a shows these results, broken down for each individual carrier sentence used. Results in the figure show an outcome that is similar to what was expected from Kimura et al. (2012), with isolated words having the highest percentage of correct responses ($\bar{x} = 92.77$, $\sigma = 8.92$), and words in final positions in questions having the lowest ($\bar{x} = 61.77$, $\sigma = 12.4$). The main difference — apart from participants having performed better in general in this study — is in the correct responses for *qint*, which are much lower than what would have been expected from the results in the original study. This was expected due to the changes in the sentence structure that were introduced to ensure that keywords presented in this sentence were framed in a non-final context (see page 75).

Speaker differences

Speaker differences proved to be significant ($\chi^2_{(3)} = 169.79$, $p < 0.001$), and are shown in figure 3.4. Although native participants showed no difference at all among the speakers, non-natives showed overall lower performance with stimuli from the older male speaker ($\bar{x} = 69.51$, $\sigma = 11.51$) as compared to all the rest ($\bar{x} = 82.17$, $\sigma = 8.73$). Results for all other speakers were not significantly different overall even for non-native participants.

3.2.2 Interactions

Sentence type × word position interaction

A significant interaction between sentence type and word position was found ($\chi^2_{(1)} = 7.33$, $p < 0.01$), with words in questions having significantly lower numbers of correct responses when in final position ($\bar{x} = 61.77$, $\sigma = 12.4$) than when they were in non-final positions ($\bar{x} = 68.51$, $\sigma = 12.35$. Responses for words in declarations, on the other hand, were higher and showed almost no difference save for their variance ($\bar{x}_{\text{final}} = 87.99$, $\sigma = 7.17$;

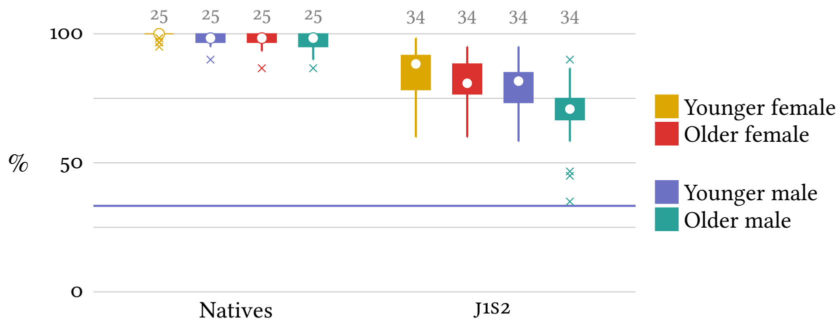


Figure 3.4: Percentage of correct non-native responses per speaker. Correct responses for the older male speaker were significantly lower than those for the rest of the speakers. The violet ● line marks the chance level. Numbers above each box indicate the corresponding *n*.

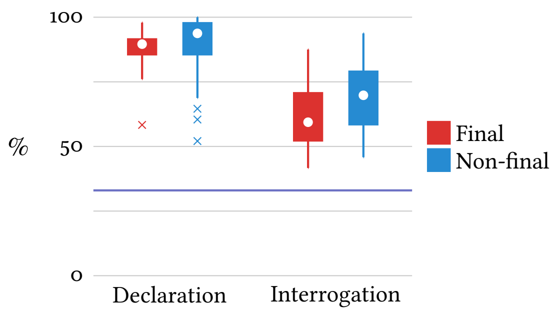


Figure 3.5: Percentage of correct non-native responses per sentence type and word position within the sentence. Words in non-final positions were less affected by their being in a question than those in final positions. The violet ● line marks the chance level. (n =34/box).

$\bar{x}_{\text{non}}$= 88.79, σ= 12.09). This is shown in figure 3.5.

Sentence type × accent position interaction

Correct responses for questions were significantly lower than those for declaratives (see page 117). Breaking these results down further between different accent positions revealed a significant interaction ($\chi^2_{(2)} = 164.55$, $p < 0.001$), with words in questions having significantly lower correct responses when they were accented in the first syllable ($\bar{x}_{\text{first}}$= 31.01, σ= 14.25; $\bar{x}_{\text{rest}}$= 49.63, σ= 12.66); while words in declarative statements had lower correct responses for words accented in the final syllable ($\bar{x}_{\text{final}}$= 79.04, σ= 10.7; $\bar{x}_{\text{rest}}$= 92.86, σ= 11.38). This interaction is shown in figure 3.3c, which shows the number of correct responses per condition; and in figure 3.7, which shows the confusion patterns per sentence context

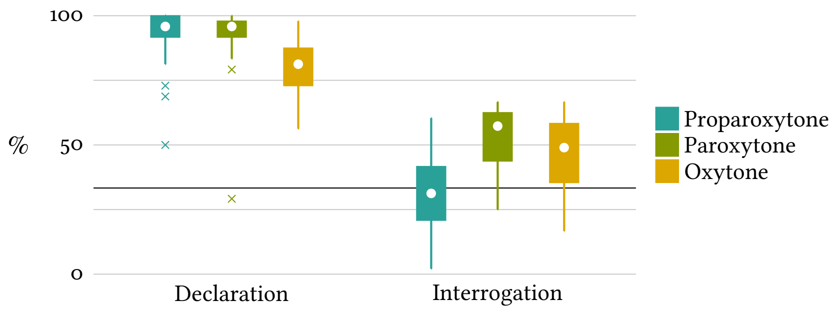


Figure 3.6: Percentage of correct non-native responses for both types of sentences per accent position. Oxytone words had the lowest number of correct responses in declarative statements, but this was reversed in questions, in which the lowest were proparoxytone words. The violet ● line marks the chance level. (n =34/box).

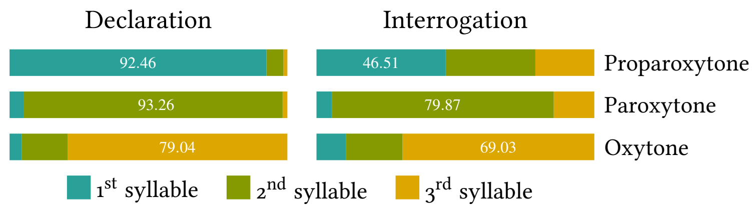


Figure 3.7: Non-native confusion patterns for accents in both types of sentences. The plot shows the syllable that was perceived as accented for each of the three accentual classes used as stimulus. The white number indicates the correct responses. Results in this figure are aggregated across all participants.

for both native and non-native participants.

Speaker × accent position interaction

Correct responses per accent position varied according to which of the speakers had produced the stimulus ($\chi^2_{(6)} = 15.62$, $p < 0.05$). Figure 3.8 shows the overall confusion patterns for each speaker, and shows that, while the response bias towards paroxytone words was present for all speakers — as shown by the presence of paroxytone responses regardless of the actual position of the prominence — this bias was significantly more pronounced for the older male speaker.

Speaker × sentence type interaction

The magnitude change in performance between declarative and interrogative statements described in section 3.2.1 also varied depending on the speaker ($\chi^2_{(3)} = 43.06$, $p < 0.001$), as is shown in figure 3.9. While words in questions had a lower number of correct responses for all speakers, the

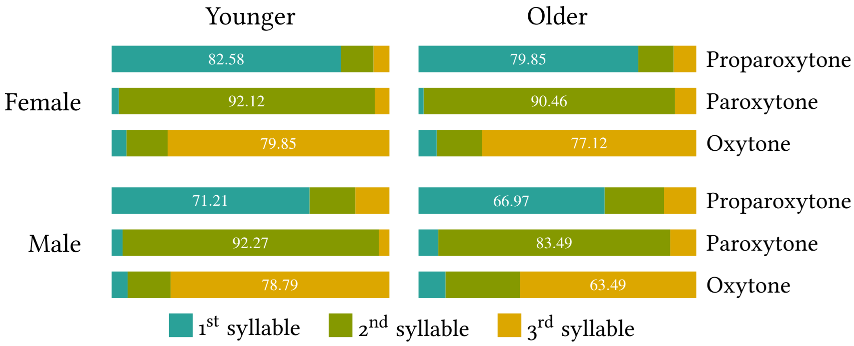


Figure 3.8: Non-native confusion patterns for accents produced by each of the four speakers. The white number indicates the correct responses. Results in this figure are aggregated across all participants.

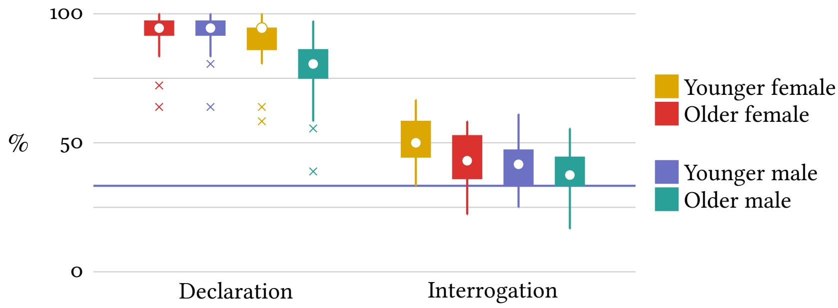


Figure 3.9: Percentage of correct non-native responses for both types of sentences broken down for each speaker. The decrease in the number of correct responses in questions affected the stimuli produced by the older female speaker and the younger male significantly more than the rest. The violet ● line marks the chance level. (n =34/box).

decrease in the number of correct responses in questions affected the stimuli produced by the older female and the younger male speaker significantly more than the rest: taken as an aggregate, responses for these two speakers showed a decrease of 25.97 percentile points; while responses for the other two speakers only fell by 14.61 percentile points.

Speaker × word position interaction

Effects of word position within the sentence also affected different speakers in different ways ($\chi^2_{(3)} = 25$, $p < 0.001$). As shown in figure 3.10, a change in word position reduced the number of correct responses for the same speakers as above (*i.e.* the older female and the younger male speakers), but increased the percentage of correct responses for stimuli from the younger female speaker. Correct responses for stimuli from the older male speaker

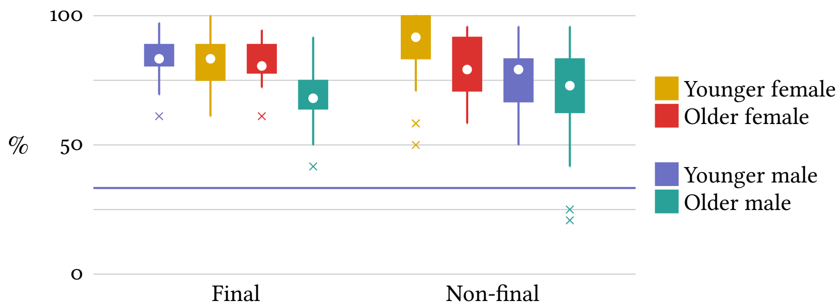


Figure 3.10: Percentage of correct non-native responses for words in final and non-final positions within the utterance, broken down for each speaker. Stimuli from the older female speaker and the younger male elicited fewer correct responses in non-final positions, while the opposite was true for those from the younger female speaker. The violet ● line marks the chance level. (n =34/box).

did not change on average, but they did increase their spread.

Sentence type × word position × accent position interaction

A 3-way interaction between sentence type, word position, and accent position was also found to be significant ($\chi^2_{(2)} = 115.1$, $p < 0.001$). Figure 3.11 shows that the different response patterns found for both sentence types (seen in section 3.2.2) are more clearly marked for words in final positions. Word in non-final positions, on the other hand, share a different pattern that is similar to that shown first in figure 3.3b, with paroxytone words having a higher number of correct responses than the rest of the positions.

These results follow those predicted in section 2.4.1 based on the acoustic profiles of the Spanish stimuli. Words in final position of declarations were predicted to elicit higher correct responses for word-initial accents, with the possibility of participants being biased towards earlier accents. This is precisely what is shown in figure 3.12, which shows the confusion patterns for this interaction. This word-initial bias was predicted not to occur in words in final position of questions, which were instead predicted to elicit the lowest number of correct responses in general. Like before, this prediction is also reflected in the data. In this context correct responses were lower, but those for proparoxytone words were significantly lower than the rest, reaching chance levels. Finally, words in non-final positions were predicted to elicit an intermediate number of correct responses, which was

Final position of declarations

Final position of questions

Non-final positions

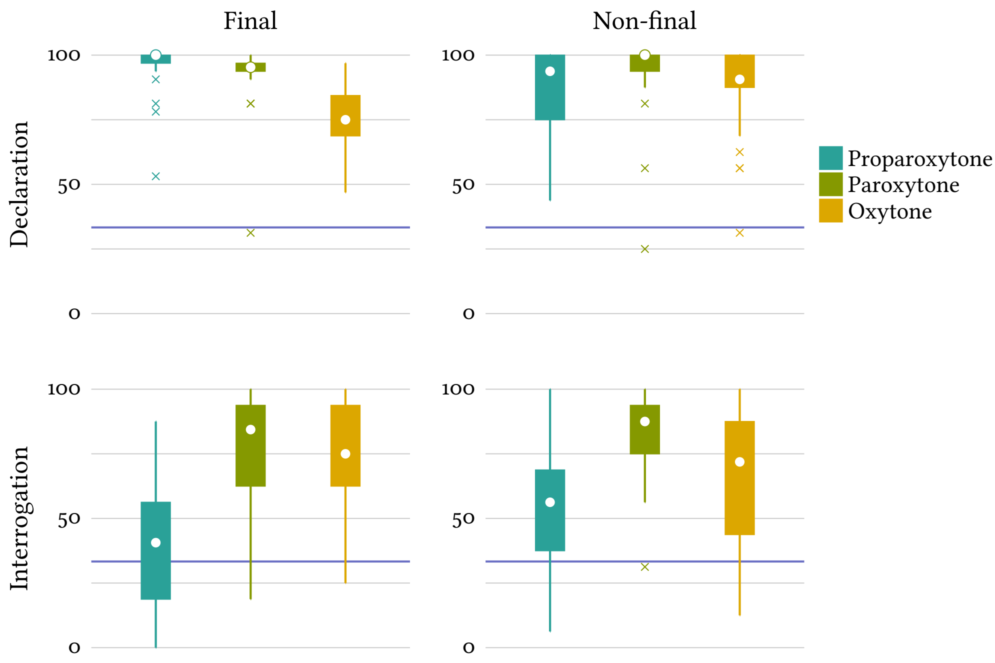


Figure 3.11: Percentage of correct non-native responses for each accent, broken down by sentence type and word position. Words in non-final positions show a similar response pattern, although shifted downwards in questions. For words in final positions, response patterns in declaratives and interrogatives were dramatically opposed. The violet ● line marks the chance level. (n =34/box).

also true. The data in figure 3.12 also shows participants were affected by a strong bias towards paroxytones. This bias appears stronger in questions, but ceiling effects in declaratives makes it impossible to prove this.

Speaker × sentence type × word position interaction

An interaction between between speaker, sentence type, and word position was also significant ($\chi^2_{(3)} = 8.67$, $p < 0.05$), and is shown in figure 3.13. As the figure shows, correct responses for words in final and non-final positions of declarations differed only slightly. And although words in questions showed also little change for the older female and the younger male speakers, words for the other two speakers showed a significant increase.

Speaker × sentence type × accent position interaction

A 3-way interaction between speaker, word position, and accent position was also highly significant ($\chi^2_{(6)} = 24.66$, $p < 0.001$), and is shown in figure 3.14. The figure shows that the effect of a change in word position from final to non-final not only affected different speakers differently (see page

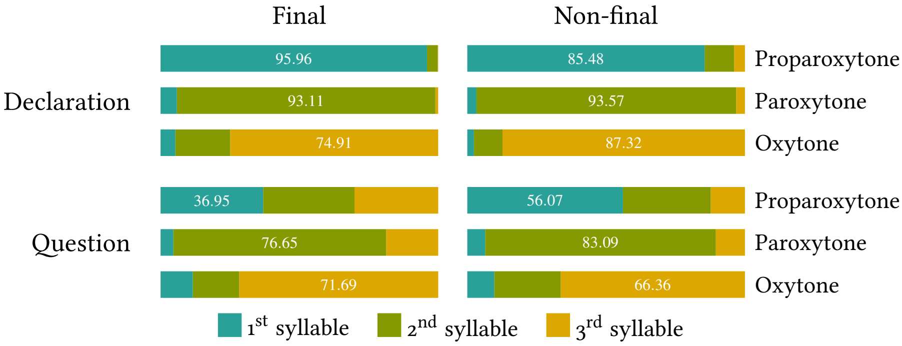


Figure 3.12: Confusion patterns per accent category for each sentence type and word position. The white number indicates the correct responses. Results in this figure are aggregated across all participants.

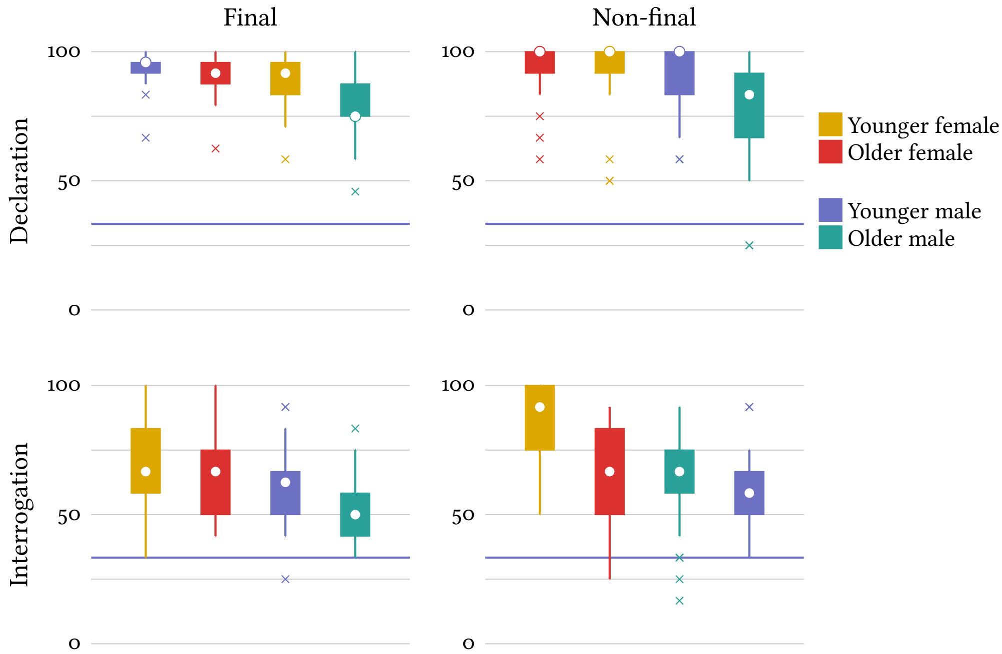


Figure 3.13: Percentage of correct responses for each speaker per word position and sentence type. In questions, the number of correct responses for the younger female and the older male speakers increased significantly when the word was in non-final position, despite there not being a significant difference for the rest of the speakers. The violet ● line marks the chance level. (n =34/box).

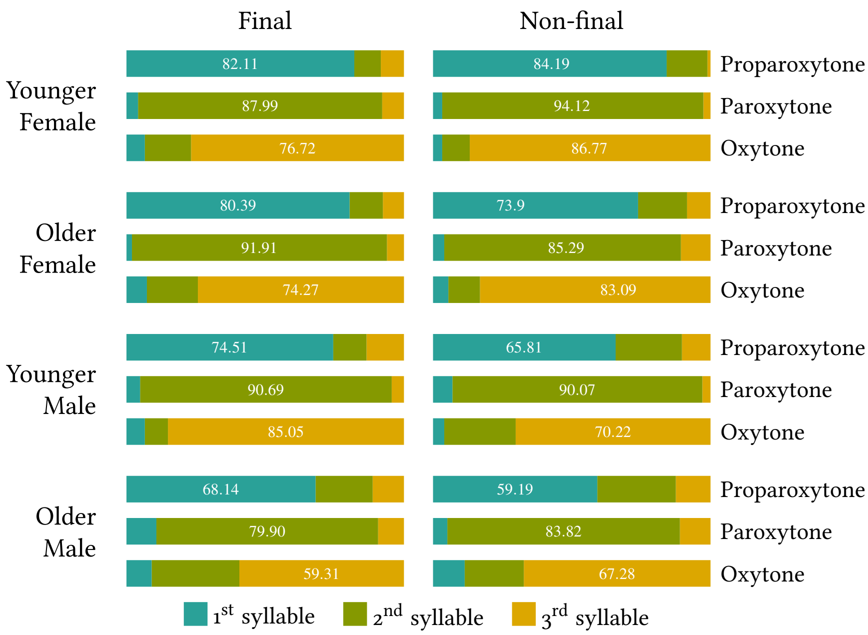


Figure 3.14: Confusion patterns for accents per word position, broken down by speaker. The white number indicates the correct responses. Results in this figure are aggregated across all participants.

118), but was also different for different accents. One such difference is that while correct proparoxytone responses decreased in non-final positions, the change had the opposite effect for oxytone responses. However, this was not the case for the younger female speaker, for whom proparoxytone responses only marginally increased; nor was it true for the younger male speaker, for whom oxytone responses decreased.

Four-way interaction

Similarly, a highly significant four-way interaction between all the factors considered was found ($\chi^2_{(12)} = 34.52$, $p < 0.001$). Figure 3.15 shows the confusion patterns for all accent types broken down by speaker and each of the five carrier sentences, and thus provides the most detailed view of the data. There are multiple possible sources of variation that can explain the four-way interaction, among them the particularly low performance for oxytone stimuli by the older male speaker in the *sfin* context, and the also particularly low number of correct responses for oxytone words in the *qint* sentences as spoken by the younger male speaker.

Figure 3.15: Syllables perceived as accented for stimuli from each speaker for each of the five sentences used. Cells corresponding to correct responses have their magnitudes labelled. Compare with figure 4.8, on page 146.

3.3 Discussion

The main objective of this study was to confirm the results reported in Kimura et al. (2012), and examine whether they would exist under more variable conditions. Results show this to be broadly the case, providing definite support to the idea that the ability of Japanese learners to perceive the stress in L2 Spanish is heavily dependent on the sentence context in which the keyword is framed.

3.3.1 Effects of intonational context

The original experiment grouped the sentences in three according to overall levels of performance alone, with words produced with a descending F0 (*i.e. isol*, *sfin*, and with their materials, *qint*) having the highest performance; words produced with a rising intonation (*i.e. qfin*) having the lowest performance; and the remaining contexts having an intermediate number of correct responses.

The acoustic analysis of the materials used in this study, described in detail in section 2.4.1, made similar predictions in terms of overall difficulty for non-native participants, but extended those predictions to include the way in which intonational contexts would be accent-specific.

Effects were mostly accent-specific

These extended predictions were shown to be accurate: earlier accents had higher performances when in words in statement-final positions, while later accents were easier to perceive in words in question-final positions. In the case of words in non-final positions regardless of statement type, participant performance was intermediate and subject to a bias towards paroxytone responses.

Role of acoustic cues

This presents strong evidence for the sensitivity of J1S2 to a multiplicity of acoustic cues, not all of which exist in Japanese. The design of the current experiment makes it difficult to specify what the precise role of each cue is. However, it does provide sufficient evidence in support of a primary sensitivity of J1S2 to F0 (in line with the conclusions put forth by Kimura et al.), and a secondary sensitivity to duration and/or intensity.

Primary sensitivity to F_0

The primary reliance on F_0 is expected from the fact that this is the only cue for word-level prominence in Japanese. But it is also supported by the response patterns described above, which follow what are primarily differences in the general shape of the F_0 contours of those sentences.

This is further supported by examples in which there are high correct response rates even in the relative absence of other cues. This is the case for intensity, with participants being able to accurately identify the position of the accent in *sint* sentences which showed no intensity contrasts; and also for duration, since there are clear differences in the ability of non-native participants to perceive oxytone accents despite the fact that these always have a similar duration profile (see table 2.8, on page 100).

Sensitivity to duration

To a lesser degree, participants were also sensitive to duration, as evidenced by the responses for isolated words from the older male speaker. His productions of words in isolation had noticeably weaker and more highly variable duration contrasts than those for the rest of the speakers for all accent types save proparoxytone (see section 2.4.3). Participant responses follow this same pattern, with a high number of correct responses for proparoxytone words and significantly poorer performance for the rest, suggesting an effect of duration on their perception.

Sensitivity to intensity

Likewise, the generally higher performance for sentences with an intensity contrast (*i.e. isol* and *sfin*) than for those without one (*i.e. sint*, *qint*, and *qfin*) supports a sensitivity to this cue.

Further support comes from the low performance for oxytone words in these otherwise high-performing contexts. As explained in section 2.4.1, words in *isol* and *sfin* showed a constant decrease in vowel intensity for each syllable regardless of the position of the stress; a decrease which is more pronounced for words with initial accents. This means that earlier accents would have accented syllables which were markedly louder than the unaccented ones; while accented syllables that came later in the word had a good chance of not only not being louder, but of having the lowest intensity in the word. This not only explains the low performance for final accents, but also the perceptual shift of the accent towards the preceding syllables, which were louder.

Speaker variability

The main objective of this study was to examine the effects of the original experiment under more variable conditions, and the main source of that variability came from the larger number of speakers used for the production of the materials. In analysing the materials collected, chapter 2 focused on the general trends needed to describe the data, but showed an overview of the scope of that variability in section 2.4.3.

The large number of significant interactions in which speaker was one of the participating factors is testament to the magnitude of the effect of speaker variability, and once again highlights the need for high-variability testing and training in linguistic research in general, and in L2 research in particular.

As described in section 2.4.3, most of the variation in Spanish could be traced to systematic differences in the choice of pitch accents and their production by different speakers. Speaker-specific normalised F_0 tracks for each sentence are shown in Figures C.1 to C.5 for Spanish and figures C.6 to C.10 for Japanese (in the appendix), and provide an explanation to some of the speaker-related effects presented above.

In specific, this is the case for the high performance of non-native participants for the older female speaker's *sint* sentences and the younger female speaker's *qint* sentences. In both cases, the general tendency to produce L*+H accents was changed, favouring instead more peak-like productions: L+H* accents in the case of the older female speaker, and something akin to either L+H* or L+>H* accents caused by the introduction of a L target between the keyword and the H% by the younger female speaker. This supports what is perhaps the most pervasive tendency throughout the results in this study: that pitch accents that have more peak-like shapes result in higher numbers of correct responses for J1S2 participants.

3.3.2 Limitations

A note on remote testing

It must be noted here again that the data reported in this study (as well as in the study reported in the following chapter) were collected using remote tests, under considerably different testing environments. The fact that similar effects can be seen despite this, and that as explained above they align with the predictions from section 2.4.1, suggests that the data collection procedure did not introduce unmanageable amounts of noise.

Still, there is the possibility that some of the differences that were found in the data sets, in particular regarding the significantly higher overall number of correct responses in this study, could be explained by the procedure chosen for data collection. Although it is unlikely that changing the data collection procedure would result in significant differences in the main effects shown in this study, a finer examination under more controlled conditions and with participants spanning a larger range of linguistic proficiency could have shed some light on some aspects of the unresolved issues in this experiment such as the lack of effects of proficiency and L2 use.

Differences in participant recruitment

Related to the issue mentioned above, the same concerns that promoted the choice of online tests to increase the number of participants, also promoted the use of different participant recruitment strategies to make use of the largest possible participant pools. This resulted in different sources for participants from the control and experimental groups, which places a certain amount of doubt regarding the comparability of both.

The measures that were being examined in this case were the ability of the populations to perceive the position of the Spanish prominence, and this was not an ability that was expected to change dramatically with the differences that this recruitment method introduced in both populations (*e.g.* age, educational level, etc). This means that, like before, and particularly because of the magnitude of the differences found, it is unlikely that repeating this study with a more stringent and balanced recruitment process would result in significant changes in the results. But such a replication would be welcome.

3.3.3 Summary of the results

In Kimura et al. (2012), results were explained as the effect of language transfers between Japanese and Spanish. Japanese participants would focus on pitch information and interpret it much in the same way as they did in their native language (but see a possible re-analysis in Kimura et al., 2015). However, this does not seem to be the case.

Results from the present study cannot be explained by simply assuming that J1S2 participants attend to pitch as they do in their native language, since they are evidently being influenced by a more complex interaction between available acoustic cues. Participant responses are clearly motivated by differences in duration and intensity, and the way in which these influence their perception was accent-specific: biased towards the beginning of the word in *e.g. isol* (perhaps by the presence of high intensity early syllables), and towards the end of the word by the effect of a rising intonation in *qfin* sentences. The scope of the current study, however, does not make it possible to thoroughly evaluate what cues are being considered in each case nor to what extent, so this remains an unanswered question.

Further implications and possible extensions of the results from this study will be discussed together with the results from the study in the next chapter, in section 4.3.7.

Chapter 4

Perception of L2 Japanese lexical prominence

Chapter 3 reported the results of an extended replication of the study in Kimura et al. (2012, and see a summary on page 64), which focused on the effect the intonational context had on the perception of L2 lexical prominence by Japanese students of Spanish (J1S2). As part of the bidirectional approach explained in the introduction, this chapter presents the results of an experiment whose design mirrors the Spanish replication mentioned above, but focused this time on the perception of lexical prominence in Japanese by Chilean students of Japanese (S1J2). Since the same linguistic features that cause language transfer effects for L2 learners are manifest in the productions of their L1, this bidirectional approach makes it possible to look at the effects of those features from both ends.

4.1 Design

Sections 1.3.1 and 1.3.2 respectively describe the aspects of Spanish and Japanese that are relevant to the discussion of the realisation of lexical prominence in both languages. To summarise, Japanese and Spanish belong to different language families and rhythm classes, use different sets of cues for the marking of prominence, and the prominence that is marked is of a

---

An early version of the results in this chapter were presented at the 20th meeting of the Congress of the Chilean Society of Linguistics (SOCHIL) in Concepción, Chile, in the winter of 2013. Digital materials related to this study can be found in the study's repository at `http://www.pinguinorodriguez.cl/research/phd/sentence-context/japanese`.

different type for each language: while Spanish has a stress accent, Japanese has a pitch accent. And despite all these differences, there are some specific contexts in which the realisation of lexical prominence bears an acoustic resemblance. In more specific terms, there are cases in which the F_0 contour, the main acoustic cue for lexical prominence in Spanish and the only one in Japanese, behaves similarly across languages; results from both Kimura et al. (2012) and chapter 3 confirm that at least J1S2 benefit when this is the case.

The questions that this present study is trying to resolve, then, focus broadly on the other side of the story, to see to what extent the partial acoustic similarities between the Spanish and the Japanese prominence are also beneficial for S1J2. Even if these similarities are largely restricted to the behaviour of F_0 (as described in chapter 2), since F_0 is the primary cue for Spanish stress this might still prove beneficial. If not, the aim of this study is to try to understand why.

4.1.1 Research questions

Broken down in more detail, this study attempts to answer the following questions:

1 Will the performance of S1J2 be affected by accent type, such that some accent positions are easier to perceive than others? For J1S2 this was not only the case, but they also showed a particularly complex set of interactions, according to which it was possible to predict what accents would be difficult in different types of sentences.

1.1 How will S1J2 perform with the Japanese unaccented pattern, particularly considering unaccented lexical items do not exist in Spanish? They could be predicted to perform worse with unaccented words based on their lack of an unaccented category in Spanish and the even lower availability of acoustic cues (since the lack of a F_0 fall will also mean the lack of anything resembling a F_0 peak). But if they are perceived to be from a non-existing category, or even if they are perceived as poor-exemplars of normally accented words, existing perception models (*e.g.* PAM, Best, 1995; Best and Tyler, 2007) would predict higher discriminability.

2 Would the proficiency of S1J2 at perceiving the position of L2 prominence be influenced by differences in sentence contexts, like it was for J1S2?

2.1 Will isolated words be the easiest for S1J2? Previous results found that non-native perception of lexical prominence in L2 Spanish is highest with words in isolation, in which the shape of the F_0 contour is the most similar, and in which there is the least amount of distracting information.

2.2 Will the *qfin* context be similarly difficult for S1J2 as it was for J1S2? J1S2 suffered particularly from the effects of intonational context when the word was at the end of a question, and therefore in the middle of a steep final rise. As has been shown in figure 1.3, Japanese does not behave the same way in these contexts, but it is unclear if this robustness will help the perception of L2 prominence by S1J2 or if they will be confused by what would be an atypical (but possible) final pitch movement.

To achieve results that would be comparable across tests languages, the design of the test presented in this chapter closely followed that of the Spanish test. However, some changes needed to be introduced to account for the unavoidable structural and lexical differences between the two target languages. To different degrees, these changes affected the inventory of target words to be used, and. the structure of the carrier sentences (although not the sentence contexts). See chapter 2 for a complete description of the design and development of the speech materials used in this study.

4.1.2 Participants

Chilean participants: S1J2

Chilean participants were recruited mostly from 1st to 5th year students at the Instituto Chileno Japonés de Cultura and the Japanese language department at Universidad de Santiago de Chile. Chilean students were offered monetary compensation in the form of a raffle for prize money to take place upon completion of data collection.

Japanese participants

Controls, on the other hand, were recruited through internet-based social networks and local contacts in Japan. A total of 28 (20 female and 8 male) participants from Chile and 17 (9 female and 8 male) participants from

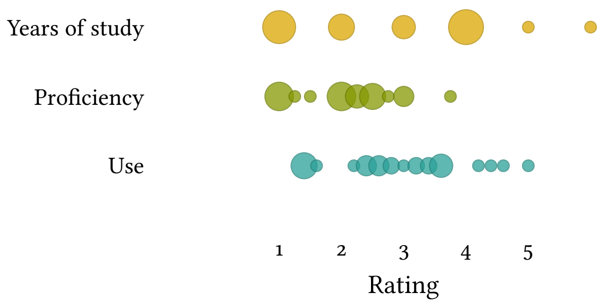


Figure 4.1: Self-reported measures of L2 proficiency of students of Japanese in the first experiment. The horizontal position of each circle marks the mean score reported by each participant to the three scales. In the case of proficiency and use, participants were given a 5-point scale, while they were free to enter how many years they had studied for. The area of each circle reflects the number of participants for that particular score.

Age group	Native	S1J2
15–19		2
20–24	2	19
25–29	4	4
30–34	9	1
35–39		2
40–44	1	
45–49		
>50	1	
Total	17	28

Table 4.1: Age distribution for participants of the first Japanese test, according to age categories.

Japan completed the test. An additional 33 participants (from both groups) signed up for the test but did not complete it. Like in the study reported in chapter 3, an original requirement for participants to have Kantō as their place of origin was dropped once it became clear this had no effect no the results. Also like in that test, participants registered their ages using age categories, which made it impossible to calculate accurate means. However, most Japanese participants reported being between 30 and 34 years old, while Chilean participants were mostly between 20 and 24, in accordance with the fact that most of them were students. Table 4.1 shows the details of this information.

4.1.3 The test procedure

The setup of the test shared most of its details with that of the Spanish test. Like in that case, the test used LimeSurvey as the testing platform to facilitate remote data collection, and discounting minor changes, the code that defined the stages and the logic of the test was the same: the general structure of the test, the overall content of each stage, and the randomisation strategy, they all remained unchanged.

Stages in the test

As a summary, the four main stages of the test were

1 A preliminary questionnaire, collecting data on each participant's demographic and linguistic background, as well as on the environment under which the test had taken place.

2 A practice stage during which participants received immediate feedback on their performance, to ensure they understood the task.

3 The testing stage, separated into a number of test blocks (see below) with breaks in between.

4 a final stage, during which participants were given feedback on their overall performance in terms of number of correct responses per sentence context and accent type.

Aspects that differed between both tests are detailed below. The specifics of all other aspects are detailed in section 3.1.2.

Testing blocks

The randomisation scheme used in the Spanish test required as many test blocks as speakers (to make sure that all tokens appeared as spoken by each speaker). However, as explained above, the Japanese test used only 3 speakers instead of 4, which meant that this test had one fewer testing block.

Graphical accent marks

Another difference between the tests related to the way in which the questions were presented. When presented in Spanish for the study in chapter 3, the question was presented using standard orthographic rules which include a graphical mark for the accent. If no orthographic mark had been used (*e.g.* if words had been identified with graphic representations of what they denote), then participant responses would have been affected by their

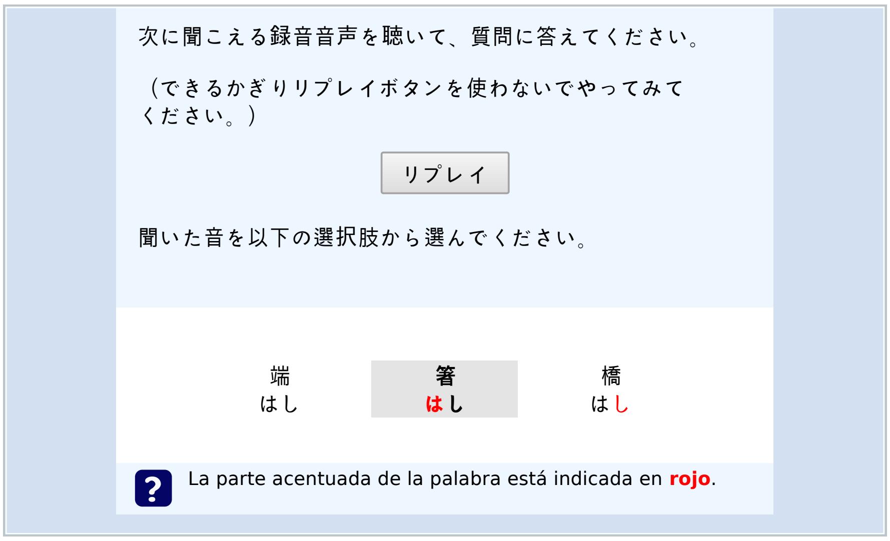


Figure 4.2: Example of a question in the Japanese test. The question text was shown in Japanese for all participants, but the explanation underneath was presented in the participant's L1. In the example, the second alternative is selected.

knowledge of the relevant lexical items, which would have been problematic because of the large differences in the rate of occurrence of the target words (see table 2.2). While this was not a problem in Spanish, in which the orthography is unambiguous regarding stress placement, standard Japanese orthography does not mark the position of the accent, and Japanese keywords had similar differences in terms of word frequency. To solve this, a non-standard accent marking system was added to the test.

Figure 4.2 shows a sample question from the Japanese test, and serves to illustrate the solution that was implemented. In the question text, the accented syllable in each of the alternatives was highlighted in bold red *hiragana* characters, and an explanatory line was added at the bottom of the screen stating this fact. Unlike the text in the rest of the question, which was written in Japanese, the explanatory line was presented in the participant's L1 to make sure that it was understood by all participants.

The trial stage

The trial stage was largely the same in both tests, the only difference being the source of the trial stimuli. To make both tests as similar as possible, a set of new Japanese nonce words was constructed with the same structure of the real keywords. They were recorded in London from a different speaker, who was not any of the speakers participants listened to during the testing stage. Like in the Spanish test, the trial keywords were first presented in isolation, and then framed in the same carrier sentences used in the actual stimuli.

The Japanese test was also hosted on servers provided by LimeSurvey, and like with the Spanish test, its complete definition file can be downloaded from the study's repository (see page 131).

4.2 Results

Participant responses were coded as correct or incorrect depending on whether they had accurately perceived the position of the accent in each stimulus. This allowed the data to be analysed as a binary response despite participants having responded to either a 2- or a 3-way categorical identification task. Since both trios and pairs have different chance levels, and because the tasks were considered to be fundamentally different, the analyses below treat both data sets separately. Like in the Spanish test in chapter 3, no differences in the testing conditions were found, and participants showed no effects of fatigue as implied by their responses to stimuli in isolation.

Data filtering

Generalised linear mixed-effects models, built using the same bottom-up approach used in the Spanish test, were fitted to the data using the `lme4` package for `R`. Also like in that study, model comparison was done on the basis of the models' AIC values, and the result of χ^2 tests to measure the significance of their differences. The only random factor in the models was that for participant. More details of the model-building process can be found in section 3.2.

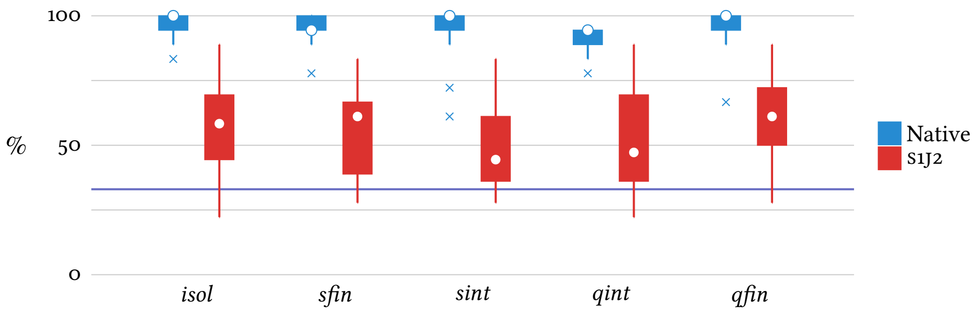
Mixed-models

Random factor: participant

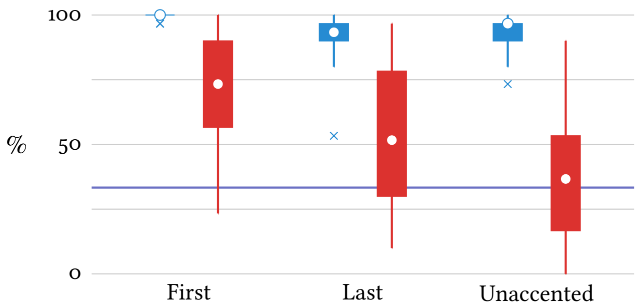
All models were built with the same combination of possible factors, and a detailed review of which were significant and under what conditions they were is given below. The factors considered were the position of the accent, the position of the word (in utterance final or non-final positions), sentence type (whether the word was a question or a declaration), the speaker, and the measures of proficiency obtained from the participant self-reports. Of these, neither sentence type nor proficiency (nor their interactions) were

Factors

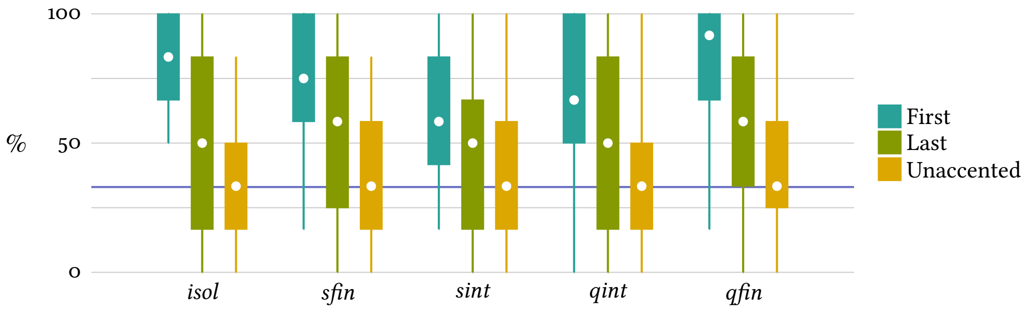
No proficiency or sentence type



(a) Results per sentence (n =17/box for natives, n =28/box for S1J2)

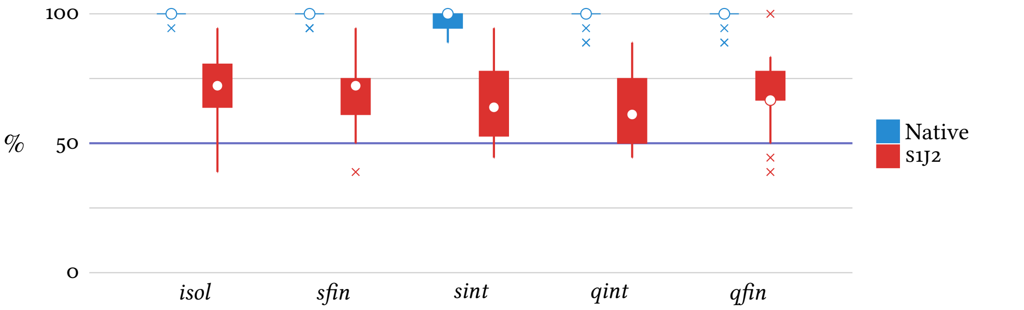


(b) Results per accent (n =17/box for natives, n =28/box for S1J2)

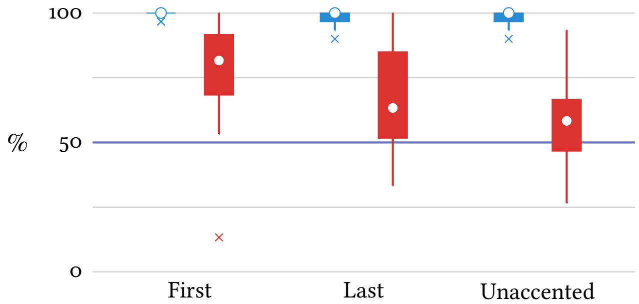


(c) Interaction between sentence and accent (n =28/box)

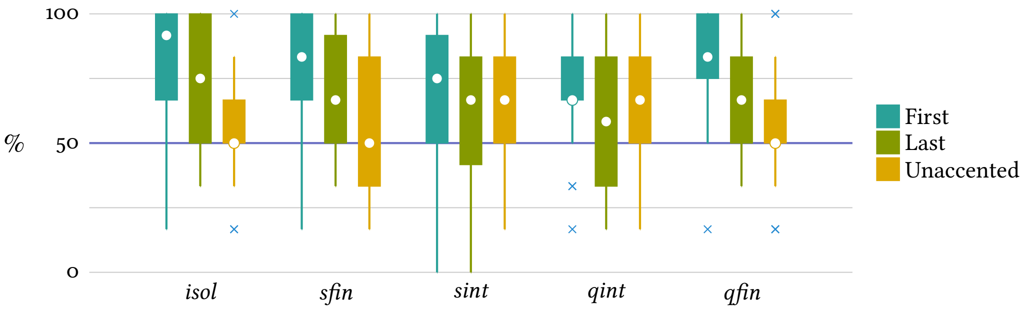
Figure 4.3: Overall results for minimal trios in the Japanese test, broken down by sentence (a), accent (b), and their interaction for S1J2 (c). In each case, the violet ● line marks the chance level. Compare with results for Japanese pairs in figure 4.4, on page 139; or with Spanish results in figure 4.3, on page 138.



(a) Results per sentence (n =17/box for natives, n =28/box for S1J2)



(b) Results per accent (n =17/box for natives, n =28/box for S1J2)



(c) Interaction between sentence and accent (n =28/box)

Figure 4.4: Overall results for minimal pairs in the Japanese test, broken down by sentence (a), accent (b), and their interaction for S1J2 (c). In each case, the violet ● line marks the chance level. Compare with results for Japanese trios in figure 4.3, on page 138; or with Spanish results in figure 3.3, on page 116.

significant under any conditions.

4.2.1 Overview

Trios: accent position accent×position accent×speaker position×speaker acc.×pos.×speaker

For trios, main effects of accent type and position within the utterance (final or non-final) were found to be significant. All three 2-way interactions between accent type, word position and speaker were also significant, as well as a 3-way interaction between these three factors. Figure 4.3 provides an overall view of these results broken down by sentence contexts, accent types, and their interaction. All significant main effects and interactions will be reviewed in more detail below.

Pairs: accent position accent×position

While considering the results of the pairs as a single set can provide meaningful information about general tendencies, the fact that each individual pair represents a different contrast means it is probably better to treat each of them separately. The results that are shown below reflect this decision, and in each case come from separate models, each fitted using the procedure explained above. An overview of the data for the minimal pairs can be seen in figure 4.4, which mirrors the structure of figure 4.3.

For overall model for pairs, main effects of accent type and word position were found to be significant, as well as the interaction between the two. This follows the predictions described in section 2.4.2.

First *v.* Last: accent position speaker accent×position

First *v.* Unaccented: position accent×position

Last *v.* Unaccented: accent accent×position

In the case of the contrast between initial- and final-accented words (*i.e.* /kami/), main effects of accent, word position, and speaker were found to be significant, as was the interaction between accent and position. For the contrast between initial-accented and unaccented words (*i.e.* /hana/), only a significant main effect of position was found, as well as a significant interaction between this and accent type. Finally, for the contrast between final-accented and unaccented words (*i.e.* /momo/), there was a significant main effect of accent, as well as a significant interaction between accent and position.

Detailed results for all significant main factors and interaction for trios are presented below, together with those for each of the individual minimal pairs and aggregated results across all pairs. These results are summarised

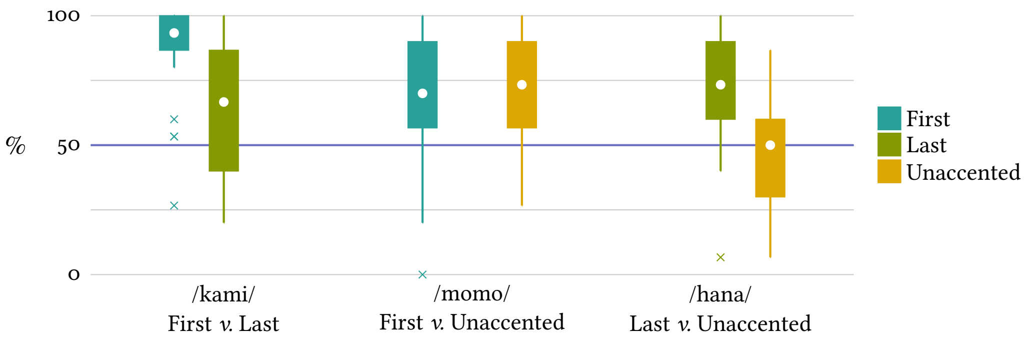


Figure 4.5: Percentage of correct non-native responses for each of the individual minimal pairs. The contrast in the /momo/ pair (between initial-accented and unaccented words) was the only case in which no main effect of accent type was found. The violet ● line marks the chance level. (n =28/box).

in table 4.2, at the end of this section.

4.2.2 Main effects

Accent type effect

A main effect of accent type was significant in almost all the tested cases, with the sole exception of the minimal pair contrasting initial-accented and unaccented words. In all other cases, the general tendency was for initial-accented words to be easier to correctly identify than unaccented words, which were the most difficult.

Not for /momo/

This was the case for minimal trios ($\chi^2_{(2)} = 204.99$, $p < 0.001$), in which correct responses for initial-accented words were the highest ($\bar{x} = 73.69$, $\sigma = 19.3$), and those for unaccented words the lowest ($\bar{x} = 37.74$, $\sigma = 23.22$); and for minimal pairs when taken as a whole ($\chi^2_{(2)} = 69.03$, $p < 0.001$): in this case too initial-accented words had the highest percentage of correct responses ($\bar{x} = 77.86$, $\sigma = 18.22$), and unaccented had the lowest ($\bar{x} = 58.93$, $\sigma = 16.83$). Figure 4.3 shows the overall results for trios, while figure 4.4 does so for minimal pairs.

Trios

Pairs

The effect of accent type was also significant for the contrasts between initial- and final-accented words ($\chi^2_{(1)} = 71.81$, $p < 0.001$) and between final-accented and unaccented words ($\chi^2_{(1)} = 43.52$, $p < 0.001$), and in both cases earlier accents had better performances. This was the case for the first contrast (/kami/; $\bar{x}_{\text{initial}} = 88.1$, $\sigma = 17.99$, $\bar{x}_{\text{final}} = 62.62$, $\sigma = 26.32$), and

Contiguous contrasts

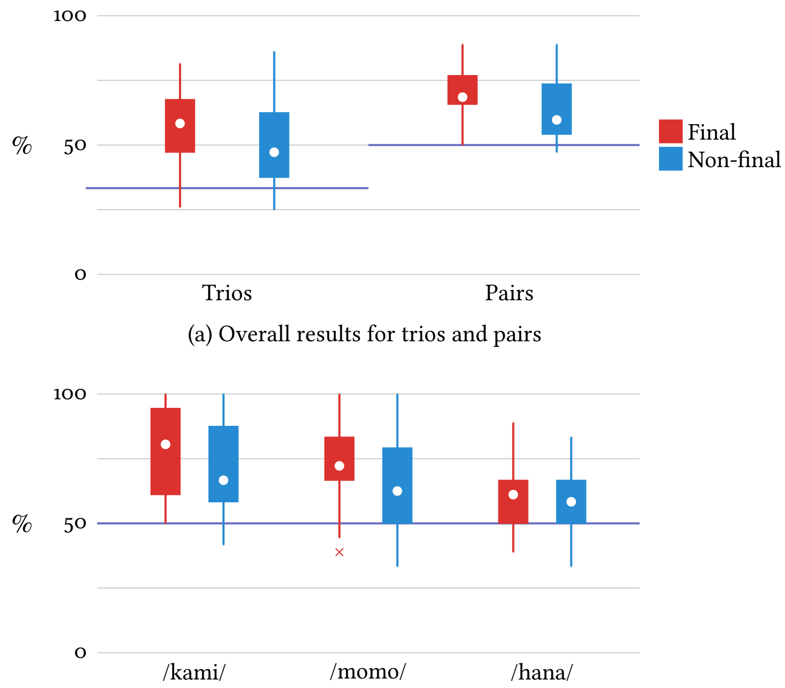


Figure 4.6: Percentage of correct non-native responses per word position within the utterance, for trios and pairs overall (a), and for each individual minimal pair (b). All displayed contrasts were significant, save for the /hana/ minimal pair. (n =28/box).

for the second (/momo/; $\bar{x}_{\text{final}} = 70.95$, $\sigma = 22.02$, $\bar{x}_{\text{unacc.}} = 47.14$, $\sigma = 23.23$).

No effect for First *v.* Unaccented

No significant effect of accent type was found for the contrast between initial-accented and unaccented words. Overall results for individual pairs are shown in figure 4.5.

For controls

A significant accent effect was also found for native Japanese speakers $\chi^2_{(2)} = 29.22$, $p < 0.001$, with correct responses for initial-accented words ($\bar{x} = 99.51$, $\sigma = 1.14$) being significantly higher than those for final-accented and unaccented words ($\bar{x} = 95$, $\sigma = 5.46$), although no difference was found between these two.

Word position effect

Acoustic analysis of the stimuli showed that unaccented words behaved differently in final (*isol*, *sfin*, and *qfin*) and non-final (*sint* and *qint*) positions within the utterance. As the difference between accented and unaccented words diminished, so would the salience of that difference, and this was predicted to negatively affect the performance of non-native participants in these contexts.

Trios

This was the case for trios ($\chi^2_{(1)} = 15.61$, $p < 0.001$), with responses for non-final contexts ($\bar{x} = 50.49$, $\sigma = 16.92$) being significantly lower than those for final contexts ($\bar{x} = 58$, $\sigma = 14.22$), and also for minimal pairs overall ($\chi^2_{(1)} = 8.91$, $p < 0.01$), with lower mean responses for non-final contexts ($\bar{x} = 64.58$, $\sigma = 12.87$) than for final ($\bar{x} = 70.04$, $\sigma = 10.19$).

Pairs

Contrasts against initial accents

When tested for individual minimal pairs, the effect was significant for the contrast between initial- and final-accented words (/kami/; $\chi^2_{(1)} = 4.21$, $p < 0.05$) and between initial-accented and unaccented (/momo/; $\chi^2_{(1)} = 5.87$, $p < 0.05$), but not for the contrast between final-accented and unaccented words (/hana/).

Last *v.* Unaccented

For controls

An effect of sentence context was also found to be significant for native Japanese speakers $\chi^2_{(4)} = 19.18$, $p < 0.001$, with correct responses for *qint* sentences ($\bar{x} = 94.44$, $\sigma = 3.68$) being significantly higher than those for the rest ($\bar{x} = 97.02$, $\sigma = 4.93$). No other significant differences were found for the rest of the sentences.

4.2.3 Interactions

Word position × accent type interaction

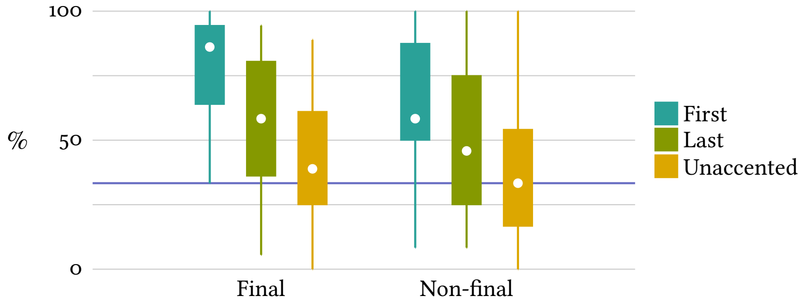
Trios and pairs overall

The realisation of unaccented words as accented in non-final positions was predicted to increase the number of correct responses for unaccented words and decrease them for the rest, at least for the contrasts in which unaccented words were one of the terms. This predicted interaction was found to be significant for minimal trios ($\chi^2_{(2)} = 8.37$, $p < 0.05$) and, to a much greater extent, for minimal pairs overall ($\chi^2_{(2)} = 19.67$, $p < 0.001$).

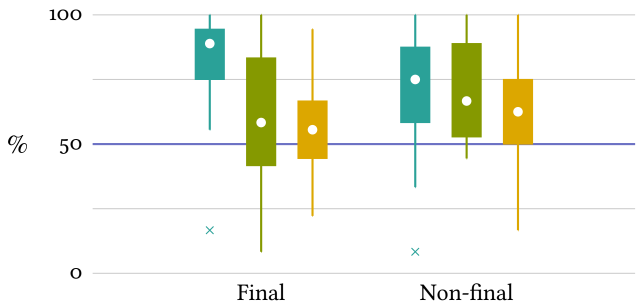
Initial-accented words saw the largest decrease in performance in the case of both trios ($\bar{x}_{\text{final}} = 79.37$, $\sigma = 18.14$; $\bar{x}_{\text{non}} = 65.18$, $\sigma = 24.12$) and pairs ($\bar{x}_{\text{final}} = 82.54$, $\sigma = 18.44$; $\bar{x}_{\text{non}} = 70.8\bar{3}$, $\sigma = 21.7$). These results are shown in figures 4.7a and 4.7b respectively.

Significant for all contrasts

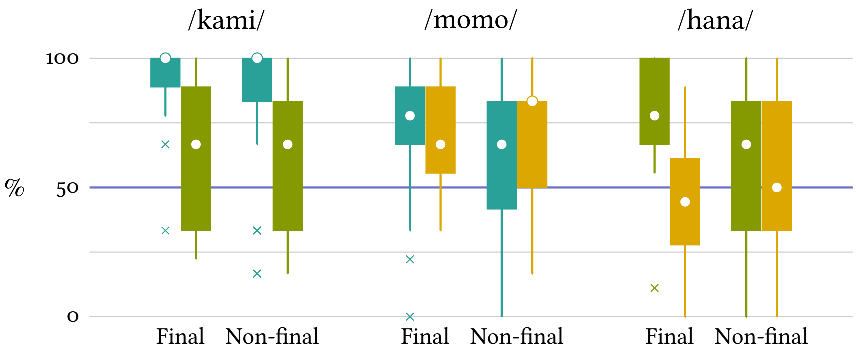
The interaction was also significant for each of the individual contrasts, results for which are shown in figure 4.7c. The effect was just under



(a) Results for trios.



(b) Results for minimal pairs overall.



(c) Results for minimal pairs, broken down by contrast.

Figure 4.7: Percentage of correct responses per word position for minimal trios (a) and pairs (b, c), broken down by position within the utterance. Responses are presented only for non-native participants. The violet ● line marks the chance level. (n =28/box).

the significance threshold for both contrasts involving initial-accented words ($\chi^2_{(1)} = 4.15$, $p = 0.042$ for the initial-final contrast, *i.e.* /kami/; $\chi^2_{(2)} = 6.43$, $p = 0.04$ for the initial-unaccented contrast, *i.e.* /momo/), and while the figure reveals it to also produce small differences, results for the /momo/ pair do tend to show the predicted tendency. On the other hand, the effect was highly significant for the contrast between final-accented and unaccented words (/hana/; $\chi^2_{(2)} = 20.34$, $p < 0.001$), in which the difference between the two accents effectively disappears in non-final positions ($\bar{x}_{\text{final}} = 60.12$, $\sigma = 31.54$; $\bar{x}_{\text{unacc.}} = 53.57$, $\sigma = 28.09$).

Effects of speaker variability

Speaker differences were a relevant factor both as main effects and as part of two- and three-way interactions.

Main effect only for /kami/

As a main effect, it was significant only for the contrast between initial- and final-accented words (/kami/; $\chi^2_{(2)} = 7.19$, $p < 0.05$), with stimuli from the male speaker having the highest number of correct responses. However, the effect was not particularly strong ($\bar{x} = 80$, $\sigma = 16.56$; $\bar{x}_{\text{rest}} = 73.04$, $\sigma = 17.15$).

Two-way interactions: speaker×position speaker×accent

Two 2-way interactions were found involving the factor for speaker, one with word position for trios ($\chi^2_{(2)} = 15.29$, $p < 0.001$) and specifically for the /hana/ contrast between final-accented and unaccented words ($\chi^2_{(2)} = 20.34$, $p < 0.001$); and the other with accent type for trios ($\chi^2_{(6)} = 29.58$, $p < 0.001$). In the case of word position, the change in mean performance for words in non-final positions (see page 142) was greatest in trios for stimuli from the older female speaker (-14.48 percentile points), and smallest for those from the male speaker (3.37 percentile points). Results for the /hana/ contrast are not as clear, since participant responses suffered from floor effects; but it is still possible to see that for the male speaker, the mean percentage of correct responses for words in final positions is $5.\bar{5}$ percentile points above corresponding responses in non-final positions, which were at chance level.

As for the interaction with accent type, correct responses for initial-

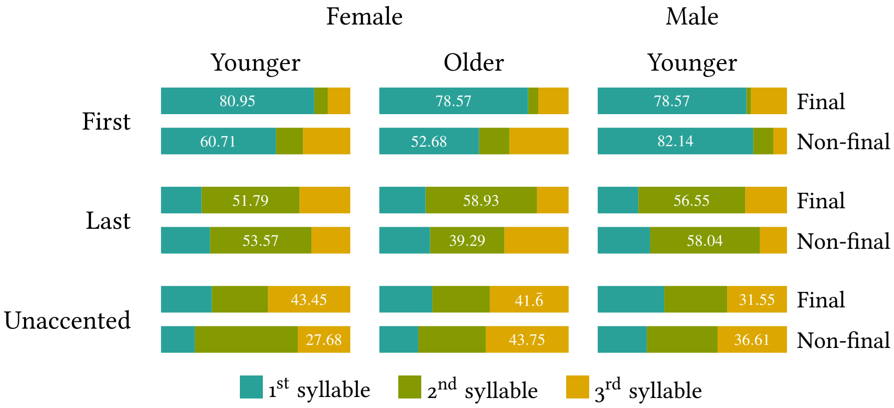


Figure 4.8: Percentage of syllables perceived as accented by S1J2 for stimuli from each speaker per sentence context. Cells corresponding to correct responses have their magnitudes labelled. Compare with figure 3.15 on page 125.

accented words were highest for the male speaker ($\bar{x}$= 80, σ= 19.63) and lowest for the older female speaker ($\bar{x}$= 68.21, σ= 22.62), but unaccented words showed the opposite tendency, having a higher performance with the female speaker ($\bar{x}$= 42.5, σ= 28.1) than they did for the male ($\bar{x}$= 33.57, σ= 24.22).

Three-way interaction only for trios

High variability in participant responses and little in terms of systematic differences in the production of individual speakers makes these results hard to explain, particularly if focus is paid to the lower-level effects in the presence of higher-level interactions. In this case, the clearest picture of the effects of speaker variability are probably shown by the significant 3-way interaction between word position, accent type, and speaker which was found for trios ($\chi^2_{(4)} = 16.57$, $p < 0.01$). Figure 4.8 shows the confusion patterns for this interaction as the distribution of responses (in percentages) for each of the accent types, broken down by position within the sentence and speaker. The figure shows that the differences responsible for the previous effects mostly reside in the particularly low correct responses for initial-accented words produced by the female speakers in non-final positions, and by low overall correct responses for unaccented words for the male speaker.

Effect	Trios	Pairs	/kami/	/momo/	/hana/
accent	***	***	***		***
position	***	**	*	*	
speaker			*		
accent×position	*	***	*	*	***
accent×speaker					
position×speaker	***			·	*
acc.×pos.×speaker	***				

Table 4.2: Summary of significant effects.

4.3 Discussion

The objective of this study was, like with the Spanish test presented in chapter 3, to examine the effects of sentence context and accent type on the perception of Japanese lexical prominence as L2, and to do so with multiple speakers. A secondary objective was to look for effects of language proficiency, to see if participants were improving in this task as they progressed through their studies.

4.3.1 Effects in the control group

Although small, significant main effects of sentence and accent type were found for the native group, with initial-accented words being identified correctly more often than the rest, and words in *qint* sentences having slightly smaller rates of correct identification. No other effects or interactions were found. Ceiling effects were found for all their other responses, which was the expected outcome. Since ceiling effects have the tendency to elicit type-II errors, hiding main effects that would otherwise be apparent, these differences in the responses from the control participants merit further study. However, they are outside the immediate scope of this dissertation, which focuses on the perception of L2 speech.

4.3.2 Accent effects and category development

Because of the differences between the way Spanish and Japanese lexical prominence is marked, non-native learners of Japanese were expected to show differences in their ability to perceive the L2 prominence depending on where it occurred within the word[1]. These differences involve not only the availability of acoustic cues (which in the case of Japanese is only F_0), but also their behaviour, since in Japanese the accent is marked by a F_0 fall, and in Spanish it is traditionally marked by a F_0 peak.

In short words, like the ones used in this study, the fall from a Japanese H tone tends to trace a contour that is similar to that of a F_0 peak, familiar to Spanish speakers. Because of this, non-native speakers were predicted to perform well when perceiving accents of types 1 or 2, which in this study were words with either initial or final accents. Unaccented words, on the other hand, were expected to prove problematic for non-native participants for a number of reasons:

1 The lack of a F_0 fall (and therefore, of a peak-like contour) and the production of what canonically amounts to basically a flat tone.

2 The lack of other acoustic cues that are normally available for Spanish speakers (*i.e.* duration and intensity), which together with an unexpected F_0 contour provided them with an even more impoverished stimulus.

3 The absence of similar accents in Spanish, which has no notion of unaccented content words.

Significant for all conditions save /momo/

For all tested conditions, with the exception of one pair, different accent types elicited different response patterns; and performance for unaccented words was always worse when it was part of the contrast. The sole exception was the /momo/ minimal pair, which represented the contrast between initial-accented and unaccented words. In this case, and only in this case, responses for both accent types was exactly the same.

The current study was not designed to provide a definitive answer on this issue, nor on the issue of why in this contrast identification rates for initial-accented words decreases and that for unaccented words increases. However, considering that the words in this contrast were the most dissimilar in terms of their accent patterns, and therefore should have the most salient auditory difference (HLL for the initial-accented word, LHH for the

1 Or, considering that unaccented words are of *type-0*, the type of the accent.

unaccented word), suggests that may be due to an auditory discrimination issue, and not one of categorical identification (see below).

Responses also differed in the case of the contrast between initial- and final-accented words (*i.e.* /kami/), in which initial accents were significantly easier to perceive for S1J2 participants than final accents. This was an unexpected result, since both accent types should have very similar realisations, and that is exactly what was found when examining the acoustic results (see section 2.4.2). However, as is shown in figures 4.3 and 4.4, there was also a slight (but significant: $\chi^2_{(2)} = 14.77$, $p < 0.001$) decrease in the number of correct responses provided by native speakers for initial ($\bar{x} = 99.21$, $\sigma = 1.31$) and final accents ($\bar{x} = 91.57$, $\sigma = 11.25$), even though there was no difference between their performance for final-accented and unaccented words ($\bar{x} = 91.96$, $\sigma = 7.27$). This suggests the possibility of a shared reason for the relative difficulty that both groups seemed to have had with accents of types 2 and 0. However, the magnitude of the difference between native and non-native participants, and the expected differences in their use of acoustic cues in general, makes this hard to demonstrate with the data collected from this study.

Difference between initial and final

Difference for native speakers

Spanish content words always have a lexical stress, so even in those cases in which words become unaccented because of sentence intonation, the stress is not lost but simply realised in a different manner. When using parenthetical intonation, for example, words are produced without a pitch accent, but in those cases lexical prominence is marked using alternative cues, and mainly duration (Ortega-Llebaria and Prieto, 2007). So even if flat tones are not entirely unfamiliar to Spanish speakers, they tend to occur in specific contexts (such as the one mentioned above), and when they do they are accompanied by these reliable secondary cues.

This is not the case in Japanese, in which F_0 is the only cue that varies in a systematic and predictable way, and is itself a sufficient marker of unaccentedness for native speakers. Secondary cues like duration and intensity do show variation both between and within speakers, but since it serves no purpose for prominence recognition in Japanese, native speakers are trained to ignore it. This lack of reliable secondary cues in the presence of unaccented words, and not necessarily the sole behaviour of the F_0 curve, is what renders these patterns unfamiliar for Spanish speakers.

Discriminability of unaccented words

Unfamiliar stimuli such as these are expected to be more acoustically salient, which in turn predicts a higher discriminability. Borrowing the terminology from PAM (introduced in chapter 1), unaccented Japanese words would either be categorised as a (very) poor exemplar of a Spanish accented word (CG), or if it was perceived to be sufficiently different, a member of a completely new category (TC). Both of these types of contrasts are expected to yield high discriminability, but this does not seem to be what was found in the results of this study.

However, the task that was presented to the participants was a categorical perception task, in which participants were asked to identify the word from a list with graphical representations of the accent, and this kind of test does not measure discriminability itself. It is entirely possible that S1J2 participants would have been able to auditorily identify unaccented words as a different sound, while still finding it difficult to map that sound to what on the screen was marked as unaccented (or more precisely, presented as unmarked). The decision to use such a test was made because participants were not naïve listeners but students of Japanese who were enrolled in university-level Japanese programs, so they were expected to have some knowledge of the language. The objective of this study was not to examine abstract auditory discriminability (which is likely to be high for Spanish speakers for the reasons stated above), but to answers questions about their ability for categorical mapping of L2 prosodic input.

The fact that they perform poorly in the chosen task says nothing about their ability to discriminate between these sounds. In fact, it would be surprising if they were not able to do so on the basis of their acoustic differences. But when they show difficulty in correctly mapping the acoustic stimuli to the appropriate linguistic categories, then the problem can be expected to lie with the development of those categories.

4.3.3 Effects of intonational context

A second objective of the study was to examine the effect differences in the intonational frame of the target words on the perception of S1J2 participants. Results from previous tests with J1S2 had shown them to be highly susceptible to the specific production of the L2 lexical prominence, which is itself heavily influenced by the word's intonational context.

This question was of interest because the Japanese accent is very robust against changes brought by immediate intonational context in particular. Figure 1.3 (from Vance, 2008) shows this, with final-accented words in question-final positions having final syllables which are almost twice as long, to accommodate the F_0 fall needed to mark the accent as well as the H% tone. This means that the context that was the most difficult for J1S2 participants in the Spanish test was not predicted to be any more difficult than the rest for S1J2. This is confirmed by the data, since the inclusion of a factor for sentence type (declarative or interrogative) never managed to significantly improve the fit of any of the models, neither on its own or in an interaction.

No effect of sentence type

There is a possibility that this might have been influenced by the low performance participants in general had with unaccented words, which often brought responses down to chance (and in some cases even further down). Perhaps using a different task the effects of the particular Japanese pitch movements at the end of questions on S1J2 participants, or the effects of the extreme lengthening of those final syllables, can be teased apart.

None of this must be taken to mean that the Japanese accent is immune to variation from the intonational context. Indeed, the current study used single words in quoted environments precisely to prevent one of the major sources of complexity in the Japanese accentuation system, which is the way phrase accents can modify the position and type of a words lexically-determined accent. But even with those considerations, figure 2.2 shows the extent of the difference found between words in final and non-final positions. As noted in the previous section, the difference is mostly limited to

Robustness is not immunity

the production of unaccented words, which in non-final positions are realised as if having type-3 accents, and are therefore produced in the same way as regular accented words (albeit with the accent in a different position).

Lower performance for non-final

This variation had a great impact on the perception of non-native participants, since it was with unaccented words that they were predicted to have particular difficulties, and non-final positions saw lower performances across the board. In fact, the only case in which no significant effect was found (the /hana/ minimal pair, see page 142) had responses that were very close to chance, and it is likely that using an adaptive test (or other such that ensured no floor or ceiling effects) would reveal this effect. Problematic in this case is the fact that the word position effect was also found for the /kami/ minimal pair, that had no unaccented words. However, in this case the significance of the effect that was found was marginal, and the difference between contexts (shown in figure 4.6b) was very limited.

This most likely explanation for the drop in overall performance for non-final positions lies with a decrease in the salience of unaccented words. This is consistent with the performance decrease in non-final positions for accented words, and with the increase for unaccented words, which also confirms that part of the difficulty that S1J2 participants had with unaccented words could be traced to the specific way in which unaccented words were being realised in final positions. Interestingly, the fact that in these positions performance for unaccented words did not rise to the top, shows that he acoustic realisation of the accent was not the only problem.

No difference for isolated words

In contrast with the results from both Kimura et al. (2012) and chapter 3, there was no particular benefit for non-native participants of words presented in isolation. However, the relatively short length of the carrier sentences meant that there was not a very large difference in length between the isolated and *e.g.* the *sfin* sentences. It is possible other carrier sentences might reveal a difference after all.

4.3.4 Speaker variability

One of the core interests of this study was to examine the issues described above under conditions of increased speaker variability. And despite this test using only a limited number of speakers, even smaller than originally planned, speaker differences turned out to be a significant factor, involved both in main effects (with the /kami/ contrast) and in interactions with word position in the case of /hana/, and with all other factors in the case of trios.

However, despite there being differences between the production of individual speakers, the lack of systematic differences (such as the ones observed in Spanish, in which specific speakers preferred different pitch accents) makes it difficult to interpret those results. A profile of the variation per speaker is captured in a series of figures in the appendix, with figures C.6 to C.10 showing differences in the F_0 contours per sentence for each speaker; and figures D.1 to D.6 and figures D.7 to D.12 showing the distribution of duration and intensity values respectively per speaker for each sentence.

Other than differences like the F_0 range of different speakers, a review of the data presented in those figures only makes it possible to point out the high variability, even within speaker, for certain acoustic features, specially intensity. In particular, the older female speaker tended to produce a louder third syllable in general for unaccented words (which might be responsible for the increased number of correct unaccented responses for her stimuli); while the male speaker tended to produce louder accented first syllables in non-final contexts, and syllables with lower intensities than those of the female speakers when that same syllable was in pre-tonic position (*i.e.* in final-accented words). These changes were undoubtedly not being controlled by the speakers, both because of the erratic nature of the variation, and because Japanese speakers have no motivation for careful intensity control. But since intensity is one of the acoustic cues used by Spanish speakers, it is not only possible but very likely that these differences, slight though they may be, are driving at least part of the significant differences

Possible intensity effects

in performance that were shown in the results.

4.3.5 Proficiency

No significant main effects of proficiency or L2 use were found, or any interactions between these two. These results mirror those of the Spanish test in chapter 3. This might be due to their proficiency in general being too low, since most of them reported being of a medium-to-intermediate level, or with there simply not being sufficient variance in their levels of proficiency. This is specially a possibility considering participants in the Spanish test showed a similar lack of an effect (but see the results in chapters 6 and 7).

Alternatively, this might be the result of a lack of appropriate training. Considering the relative lack of available acoustic cues, prosodic training is of even greater importance for S1J2 than it might be for J1S2, but language instructors are faced with similar problems in both cases, and a similar lack of motivation to tackle this issue.

4.3.6 Limitations

Like in chapter 3, participants from both the experimental and the control groups in this case were recruited by targeting different participant pools, and the call for native Japanese speakers was much broader than that of S1J2 participants. Much of the same concerns that are discussed in section 3.3.2 apply in this case as well, but like it is stated in that section, the magnitude of the effects found makes it unlikely that they are the result of differences in the participant recruitment strategies.

4.3.7 General summary

Results from this chapter and the previous one have shown that both in the case of Japanese students of Spanish and that of Chilean students of Japanese, certain contexts exist under which relatively high rates of correct word identification are possible based on suprasegmental information alone. However, the contexts under which this is possible are different for each language.

Results from chapter 3 confirmed the effects of sentence intonation on

J1S2 first reported in Kimura et al. (2012), and showed that they applied to at least one other variety of Spanish and that they remain in action across multiple speakers. Effects of prominence position were revealed to be at best secondary, and heavily dependent on the effects of sentence intonation, which accounted for most of the variance.

The results reported in this chapter, on the other hand, show that the ability of S1J2 to correctly perceive Japanese lexical prominence is not affected directly by the intonational context of the word at hand. Intonational context does have an indirect effect on their perception, however, since it modifies the production of unaccented words, and it is these words that pose the greatest challenge for Spanish speakers.

The linking thread between both sets of results is that of phonetic similarity. Those contexts in which a H tone aligned with the prominence (*i.e.* the L+H* accents in declarative statements in Spanish; and the H*+L accents in accented words in Japanese in general) were relatively easy to perceive, while those realised with different patterns — such as the L*+H or L* accents in Spanish, and the LH- in word-final unaccented Japanese words — posed an additional difficulty.

Phonetic similarities

These results support the notion that transfer effects similar to those seen in segmental perception are also at play: participants are better able to perceive those realisations of prominence that resemble their own, likely as an effect of a sensitivity inherited from their native inventories. These results are also in line with the accounts of perception that describe it as being a low-level phenomenon based on the acoustics (such as NLM), since participants do not seem to be affected by the phonological differences between these languages other than secondarily, in that they determine the contexts in which phonetic similarity is more likely to take place.

However, this interpretation might be coloured by the perceptual bias shown by S1J2 towards earlier accents, and their almost complete inability to correctly perceive unaccented words even in those contexts in which they were realised as accented (*i.e.* non-final positions). That these results coexist with those of the phonetic similarities discussed above, suggests that maybe these processes are the result of different but related mechanisms, one primarily phonetic and one based on the phonology.

It would be tempting to see in this bias towards earlier accents a transfer effect from the same lexical subregularities identified by Face (2006) for Spanish, but this is unlikely. While an initial-accented 2-syllable Japanese word *could* be interpreted to have a similar pattern as that of a Spanish paroxytone word, this would require parsing that word independently of the particle which, at least in this study, always followed it. This is not entirely impossible, but it is not at all clear whether S1J2 are indeed doing this, and further studies would be needed to answer this question.

There is also the possibility that whatever affected the perception of S1J2 participants, biasing them towards earlier accents, is also responsible for the increased initial-accented responses found for the control group (see section 4.3.1). However, responses from S1J2 participants declined steadily from initial-accented to unaccented words, a tendency which was not seen for the natives, so the likelihood that these two populations are being affected by different processes is high. This should also be studied further.

The low performance of S1J2 learners might alternatively be due to a bias against submitting unaccented responses even in those cases in which phonetic similarities would make them otherwise easier to perceive. This could be a consequence of the lack of an unaccented category in Spanish and, specially considering the lack of an effect of proficiency for both populations, general lack of prosodic training. This would be particularly important in this case, in which the task participants were presented with was one of category identification.

Follow-up studies

Having shown that both groups of participants are using a combination of acoustic cues for their perception of prominence, the studies in the next part of this dissertation will focus on further exploring these issues. In particular, the effect of secondary cues (and in particular duration) and the differences these groups might have on the weighting of these cues, as further confirmation of the links between segmental and suprasegmental perception. In particular in regards to Japanese as L2, the question of measuring the degree of category development of non-native speakers became relevant, so that will also be addressed in the following tests.

Part 2

Effects of duration in the perception of second language prominence

Chapter 5

Testing secondary cues: design and synthesis of materials

Part 1 reported the results of an experiment examining differences in the direction of language learning on the perception of L2 lexical prominence. This was done by looking at the performance of non-native learners listening to naturally-produced target words in different sentence positions and with different syllables as prominent, and repeating the experience with Japanese students of Spanish (J1S2) and Chilean students of Japanese (S1J2).

In both cases, conditions were found that allowed for relatively high performances for non-native listeners, showing that differences in the phonological definition of lexical prominence are not enough to predict a low performance in the perception of suprasegmental categories. Instead, it is more relevant to look at the effect of different contexts in the degree of phonetic similarity between native and non-native categories. Indeed, differences found between the performance of both language groups showed that the direction of language learning has a strong effect on what contexts will prove easy or difficult for non-native perceivers. These results support the notion that suprasegmental perception in general, and the perception of second language lexical prominence in particular, is subject to factors that are similar (if not the same) as those involved in perception at a segmental level.

The chapters in part 2 build on these results by examining whether or not non-native listeners are sensitive to secondary acoustic cues, and if their weighing of these non-native cues is affected by differences in their inventory of L1 cues. Evidence from chapters 3 and 4 suggested this was the case: in both cases, responses from non-native participants could not be entirely explained by the action of F0 alone, with evidence that both groups were basing part of their judgements in the behaviour of intensity and duration as secondary cues. But a more careful look at this sensitivity was necessary.

5.1 Materials

Choosing a secondary cue

Between intensity and duration, the latter was chosen as the focal point for the next set of studies because of the evidence that suggests duration has a particular status among acoustic cues, in that it is available as a cue even for listeners who do not natively attend to it (Bohn, 1995). Additionally, duration has been shown to be more relevant than intensity in the encoding of Spanish prominence in particular (Ortega-Llebaria, 2006; Ortega-Llebaria and Prieto, 2007; Ortega-Llebaria, Prieto, and Vanrell, 2008; Ortega-Llebaria and Prieto, 2009, and see Kim, 2015, for a review), and across languages in general (Chun, 2002). And although it is not used for accent discrimination in Japanese, it does have a contrastive role at a segmental level, which increased the likelihood that J1S2 would be able to tune in to it.

Intensity, on the other hand, has no role in accent discrimination in Japanese (Cutler and Otake, 1994); and requires a more involved process of acoustic manipulation, increasing the risk of introducing unwanted variation in the synthetic stimuli.

5.1.1 The test paradigm

To test the hypothesis that duration has an effect on the perception of Spanish and Japanese lexical prominence by non-native listeners, and to measure the degree of this effect, a modified trading relations paradigm (Nittrouer and Studdert-Kennedy, 1987; Mayo and Turk, 2004, 2005) was used. This paradigm makes it possible to test the effect of a secondary cue in the presence of a primary one in ways that are resistant to the neutralising effects or that predominant cue.

In general terms, the acceptable variation for the two freely varying target cues (c_1 and c_2) can be understood to define two independent dimensions which in turn defines a plane. In this plane, each point corresponds to a specific combination of both cues, and therefore to a different possible stimulus. Applied in a naïve way, a large number of stimuli would be necessary to cover the entire plane of possible cue combinations. But this number can be brought down by drastically reducing the resolution along one of these dimensions (*i.e.* the one corresponding to the secondary variable) and considering only the limits of this cue's variation. This process is schematically illustrated in figure 5.1.

This process not only reduces the plane to two lines corresponding to the opposite edges of the plane along the dimension of the secondary cue, but also maximises the effect of this cue. This is a particularly beneficial side-effect, since by definition a secondary cue will have a smaller effect than the primary cue with which it interacts. This procedure reduces the chances of masking.

There are other methods that also allow to more efficiently sample a complex stimulus space. One which has provided particularly interesting results is that of sampling by multidimensional vectors (see Iverson, Kuhl, et al., 2003, for an example). While this method remains tempting, the scenario in this case was considered too simple to merit the added complexity of such an approach, which involves adaptive testing, and is more well suited for higher-dimensional spaces.

Points along these edges define the stimuli, each of them representing one step along the continuum of variation for c_1 (with c_2 fixed at either of its extreme values). The difference between the two resulting continua will exist solely in c_2, so any difference in the responses obtained from these will derive from a sensitivity to this secondary variable, and the magnitude of the difference will correlate with the degree of sensitivity: as sensitivity increases, so will the effect of c_2 on participants' responses. This method allows for a considerable reduction in the number of tested stimuli while

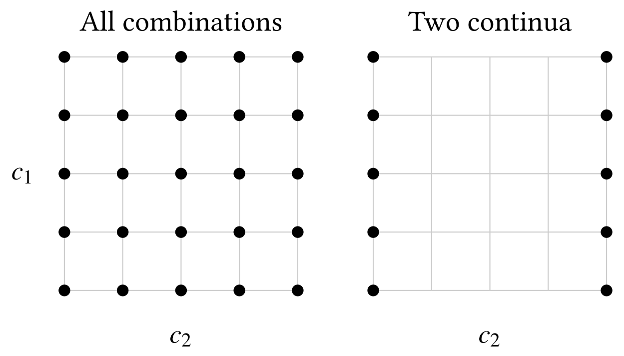


Figure 5.1: The plane of possible combinations of c_1 and c_2 can be reduced to two opposing edges by sacrificing resolution along the secondary cue (c_2).

still allowing for the precise measure of the contribution of *e.g.* duration in the presence of pitch differences. Figure 5.2 illustrates this part of the procedure, with the size of the shaded area representing the magnitude of the effect of c_2.

In contrast with the methods used for part 1, this method offers additional granularity, in that participants' responses are not simply a correct-or-incorrect binary response, or an aggregate of such, but a continuous variable represented by the position of categorical boundaries and the steepness of the response functions. This additional granularity will become particularly useful in chapter 7, when the question of the degree of categorical development in S1J2 participants becomes relevant.

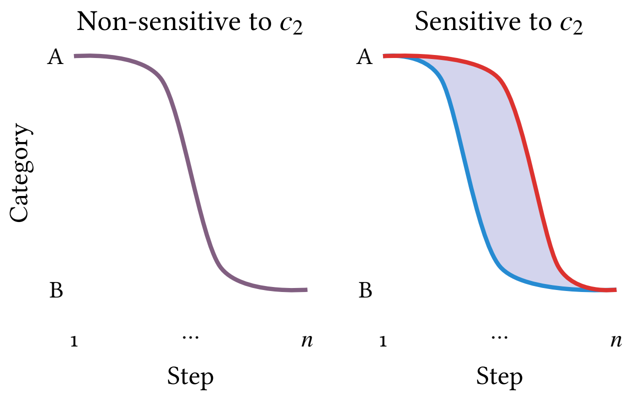


Figure 5.2: By using the two continua from figure 5.1, participants who are sensitive to c_2 will show a spread in their responses.

5.1.2 Design of the continua

To continue from the results in previous studies, the continua used in this case were developed based on a subset of the minimal trios used in chapters 3 and 4. Each of the three member words of those trios was arranged according to the position of the prominent syllable, from the antepenult to the ultima. In the case of Japanese, unaccented words were placed at the end of the continuum based on the similarities of the resulting F_0 progression between both languages, the confusion patterns shown in chapter 4, which showed Spanish learners more likely to interpret unaccented words as accented on the word's final syllable than on the first[1], and the steady decline in identification rate for S1J2 participants from initial-accented to unaccented words (see figure 4.3b). To simplify the description of the procedure, the following paragraphs will make reference to the members of the trios by their position in this sequence, regardless of the language they belong to. This means that when reference is made to words in the third or final position, they will refer to oxytone words in Spanish and unaccented words in Japanese.

The sequence of words from the minimal trios were separated into two separate contrasts: one between the first and the second word (*e.g.* proparoxytone and paroxytone in Spanish), and another between the second and the third word (*e.g.* final-accented and unaccented words in Japanese). Each of these contrasts resulted in a separate 10-step continuum, in which each step represented a different equidistant step from the pitch contour of the first term to that of the second. However, only the variation of the primary cue has been introduced up to this point.

Two versions of each of the 10-step continua were made, each one with the duration values fixed at the position of a different term in the contrast. In figure 5.1, these two continua correspond to the two sets of stimuli on the simplified plot, each at a different end of the stimulus plane along the

1 It is important to remember that while target words in Japanese were all 2-syllable words, the stimuli themselves were 3-syllable long, since they included a particle at the end to show the behaviour of the F_0 contour after the end of the word (see section 2.1.2). This means that the word's final syllable is the penult of the stimulus.

dimension of c_2. The continuum using the duration values of the first term was called continuum A, while continuum B had the duration values fixed on those of the final term. The combination of these two resulted in 4 independent continua:

A$_1$ Between the 1st and the 2nd words, with durations of the 1st

A$_2$ Between the 2nd and the 3rd words, with durations of the 2nd

B$_1$ Between the 1st and the 2nd words, with durations of the 2nd

B$_2$ Between the 2nd and the 3rd words, with durations of the 3rd

5.1.3 Source materials

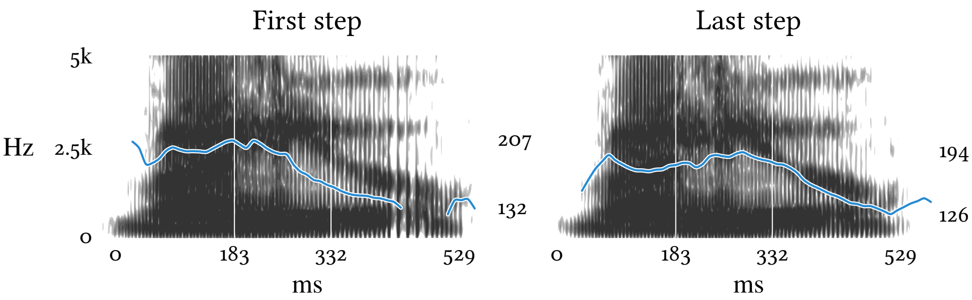
Stimuli for the continua were synthesised from recordings made from native speakers of both languages. The recordings were selected from the set of isolated words from Spanish and Japanese minimal trios that were originally made for the tests described in chapters 3 and 4 (see chapter 2 for details about the recording procedure and the acoustic characteristics of the source materials). The decision to only use isolated words was based on the results from chapter 3, which showed it to be the most beneficial for the perception of J1S2 listeners (who reached native-like levels). This was not the case for S1J2 participants, as was shown in chapter 4, but the same results showed that isolated words were among the high-performing phrases, which meant the decision would at the very least not affect S1J2 negatively.

5.2 Digital processing

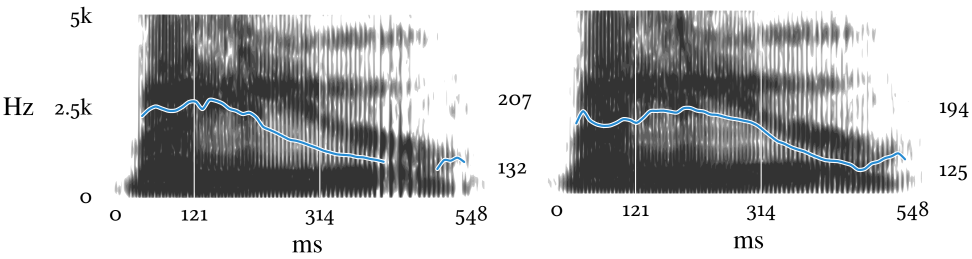
Manipulation

Manipulations were made using TANDEM-STRAIGHT (Kawahara, Morise, et al., 2008; Kawahara, Nisimura, et al., 2009; Kawahara and Morise, 2011), a system that improves the spectral calculations originally implemented in STRAIGHT (Speech Transformation and Representation using Adaptive Interpolation of weiGHTed spectrum; Kawahara, Masuda-Katsuse, and Cheveigné, 1999). Both the STRAIGHT vocoder and its improved version allow for the synthesis and modification of speech signals with very high quality output (Zen et al., 2007).

In these systems, speech is defined based on a combination of five attributes, each of which can be manipulated independently:



(a) Continuum A_1 for Spanish



(b) Continuum B_1 for Spanish

Figure 5.3: First and last steps for the continua between proparoxytone and paroxytone words. Sounds on the top row have the duration values of a proparoxytone; while those on the bottom row have the duration of a paroxytone. The displayed sounds are the female productions of the /balido/ set.

1 An aperiodicity map of the signal

2 An F_0 track, which includes voicing information

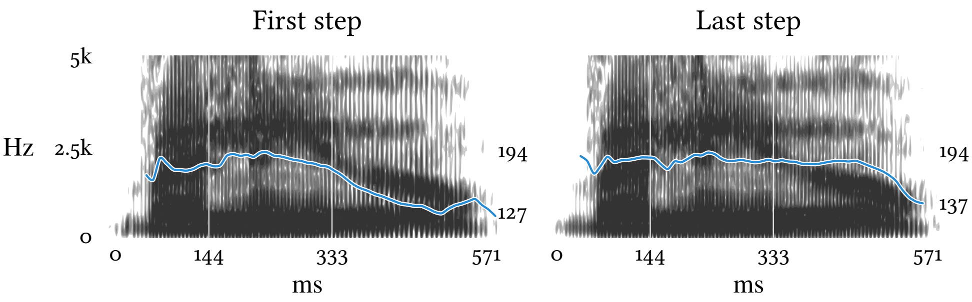
And three morphing rate parameters defining the interference-free spectrogram:

3 one for the spectrogram,

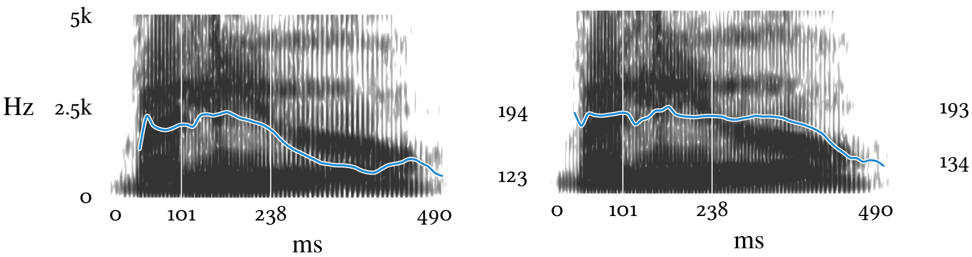
4 one for the time domain,

5 and one for the frequency domain.

A baseline F_0 contour was calculated per utterance and used to remove low frequency noise, improving the accuracy of subsequent measures. This was further improved by making use of internal functions for F_0 autotracking and refinement of voiced and unvoiced (VUV) segment detection. Once frequency data was extracted, the aperiodicity ratio of the sound was calculated and used for the calculation of the STRAIGHT spectrum. This series of steps can be seen in listing 5.1, which shows part of the code used to run the analysis in its Matlab implementation.



(a) Continuum A2 for Spanish



(b) Continuum B2 for Spanish

Figure 5.4: First and last steps for the continua between paroxytone and oxytone words. Sounds on the top row have the duration values of a paroxytone; while those on the bottom row have the duration of a oxytone. The displayed sounds are the female productions of the /balido/ set.

The resulting sound objects were then used to generate the corresponding *morphing substrates*, which are used by TANDEM-STRAIGHT for the creation of the continua. In TANDEM-STRAIGHT, a morphing substrate defines the specific output of the combination of two target sounds based on the sounds themselves (and the result of their STRAIGHT analysis); a set of time anchors to map corresponding parts of one sound; and a value for each of the five acoustic attributes described above, specifying whether that attribute should be set at that of either of the target sounds, or somewhere in between along a linear interpolation.

A set of custom scripts (which included the code in listing 5.1) read the time anchors from pre-existing annotations created manually with the software Praat, and handled the manipulation of the five attributes mentioned above. Within the continuum for each contrast, values for the duration attribute of all steps were set to either those of the initial or final targets, depending on whether the stimulus was from the A or B continua respec-

```
% Analyse waveform
% wave and rate are returned from Matlab's wavread
obj = exF0candidatesTSTRAIGHTGB(wave, rate);
wave = removeLF(wave, rate, obj.f0, obj.periodicityLevel);
obj = exF0candidatesTSTRAIGHTGB(wave, rate);
obj = autoF0Tracking(obj, wave);
obj.vuv = refineVoicingDecision(wave, obj);
% Aperiodicity extraction
apr  = aperiodicityRatioSigmoid(wave, obj, 1, 2, 0);
% STRAIGHT Spectrum
spec = exSpectrumTSTRAIGHTGB(wave, rate, apr);
o.waveform = wave;
o.samplingFrequency = rate;
o.refinedF0Structure.temporalPositions = obj.temporalPositions;
o.SpectrumStructure.spectrogramSTRAIGHT = spec.spectrogramSTRAIGHT;
o.refinedF0Structure.vuv = obj.vuv;
spec.spectrogramSTRAIGHT = unvoicedProcessing(o);
```

Listing 5.1: Code used with TANDEM-STRAIGHT to automatise the steps in the STRAIGHT analysis

tively. Values for F_0 were set for each step at one of 10 equidistant points between both target sounds. The remaining 3 attributes were fixed for all stimuli at a midpoint between both targets.

This explains why the continua between the two contrasts should be considered as separate, and not as two halves of a single continuum: in each case, the midpoint that was the reference value for the remaining attributes was different. In the case of the continua between the first and the second words the midpoint was between these words, while in the continua for the second contrast the midpoint was between the second and the third words.

5.3 The resulting materials

Speakers

The manipulation was done on all recordings of isolated minimal trios from all speakers for each language. The resulting samples were evaluated by native Spanish and Japanese speakers correspondingly to measure their naturalness, and by the researcher to monitor the number of artifacts introduced during the procedure. After this selection process, two speakers (one male and one female) were chosen from each language: the older male and female speakers in Spanish, and the younger male and female speakers in Japanese.

Keywords

Two minimal trios were chosen for each language, to bring the procedure behind the studies involving each language closer together. Words

Accent	Minimal sets	
	/numeɾo/	/balido/
Proparoxytone	Número number N	Válido valid N
Paroxytone	Numero number V-1SG-PRES	Valido validate V-1SG-PRES
Oxytone	Numeró number V-2SG-PAST	Validó validate V-2SG-PAST

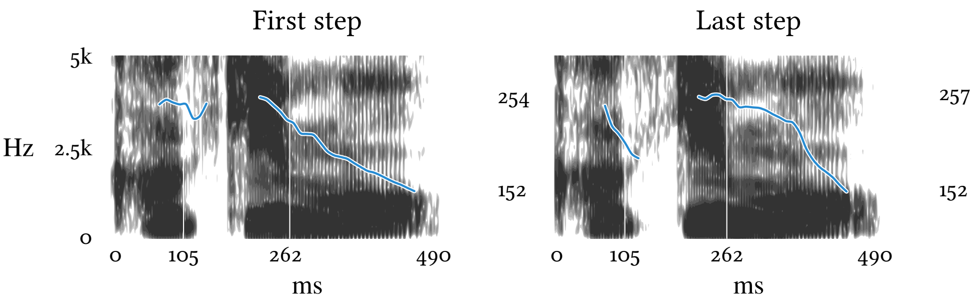
Table 5.1: Spanish target words used by this study. Compare with the Japanese keywords shown in table 5.2, on page 170.

from the /balido/ and the /numeɾo/ Spanish sets were used. These were chosen over the other two minimal sets since they contained only voiced segments, which again resulted in more natural sounding stimuli. The Japanese stimuli only had two trios, so both of them were used. The keywords used in this study are shown in table 5.1 for Spanish, and table 5.2 for Japanese.

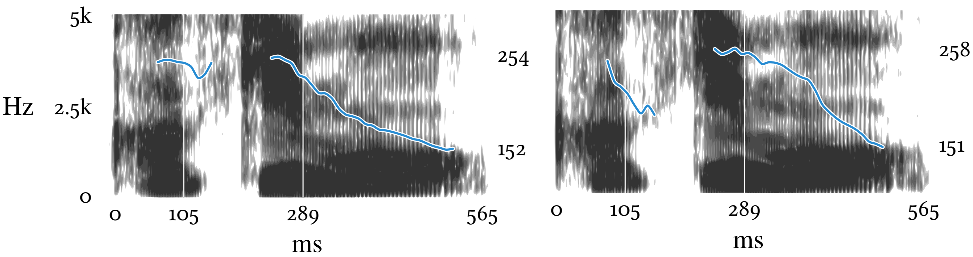
Examples of the resulting materials are shown in figures 5.3 to 5.6. Each figure shows the first and last step in both the A and B versions of the continuum between one pair of words. Figures 5.3 and 5.4 show the continua

Accent	Minimal sets	
	/kaki/	/haɕi/
First $\mu^{\downarrow}\mu$	牡蠣 oyster N	箸 chopsticks N
Last $\mu^{\uparrow}\mu^{\downarrow}$	垣 fence N	橋 bridge N
Unaccented $\mu^{\uparrow}\mu$	柿 persimmon N	端 end N

Table 5.2: Japanese target words used by this study. A $^{\downarrow}$ indicates a fall in pitch while a $^{\uparrow}$ indicates a pitch rise. μ indicates a mora. Compare with the Spanish keywords shown in table 5.1, on page 170.



(a) Continuum A_1 for Japanese



(b) Continuum B_1 for Japanese

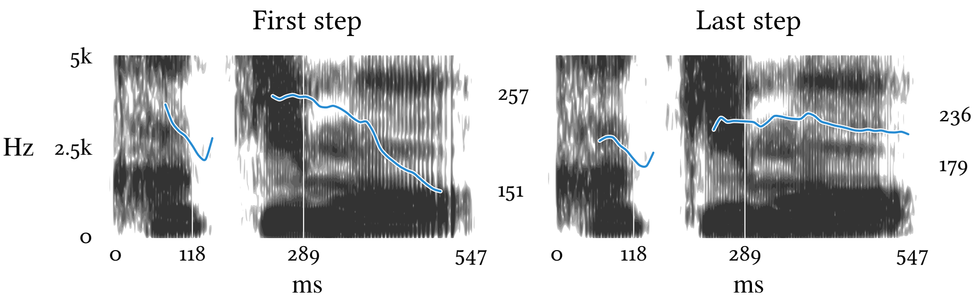
Figure 5.5: First and last steps for the continua between initial- and final-accented words. Sounds on the top row have the duration values of an initial-accented word; while those on the bottom row have the duration of a final-accented word. The displayed sounds are the female productions of the /kaki/ set.

for the first and second contrast for Spanish, while figures 5.5 and 5.6 do so for Japanese.

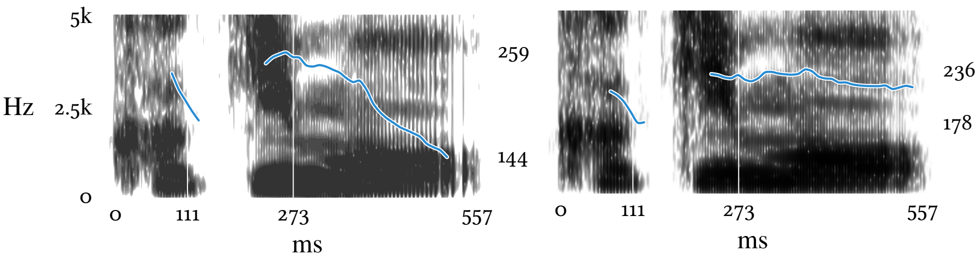
Although the figures show examples from a single word for a single speaker per language, they provide a reliable illustration of the effects of the manipulation. No negative effects of stretching were found, and since the start and end points for the manipulation of pitch for each contrast in continua A and B were the same, there was also no difference in F_0 range between these two.

5.4 Discussion

This chapter discussed the rationale for the choice of materials for the studies in chapters 6 and 7, and the procedure followed for their synthesis. An evaluation of the resulting stimuli by native speakers showed that the results of the manipulation were of very high quality in all cases, a fact that will be corroborated when examining the responses of native speakers when presented with the stimuli in chapter 6 for Spanish and chapter 7 for



(a) Continuum A2 for Japanese



(b) Continuum B2 for Japanese

Figure 5.6: First and last steps for the continua between final-accented and unaccented words. Sounds on the top row have the duration values of a final-accented word; while those on the bottom row have the duration of an unaccented word. The displayed sounds are the female productions of the /kaki/ set.

Japanese. The ease of programmatically defining the steps in each continua also made TANDEM-STRAIGHT a very good tool for the job.

This method was suitable in this case since the parameters that were to be manipulated were syllable duration and F_0, and both of these are directly represented as editable attributes for controlling the calculation of the STRAIGHT spectrum needed for synthesis. In tasks in which the objective is the careful manipulation of the formant structure of speech sounds, for example, a tool such as TANDEM-STRAIGHT might not prove as convenient to use, even if such manipulations are possible. In the case of this study, the fact that TANDEM-STRAIGHT did not offer a convenient interface for intensity manipulation had a definite role to play in the decision to prioritise the study of the effects of duration. In choosing a tool such as TANDEM-STRAIGHT, a decision was made to prioritise the naturalness of the results over the minute control of the aspects that were changing in each step that could be achieved with other methods like a Klatt synthesizer (Klatt, 1987;

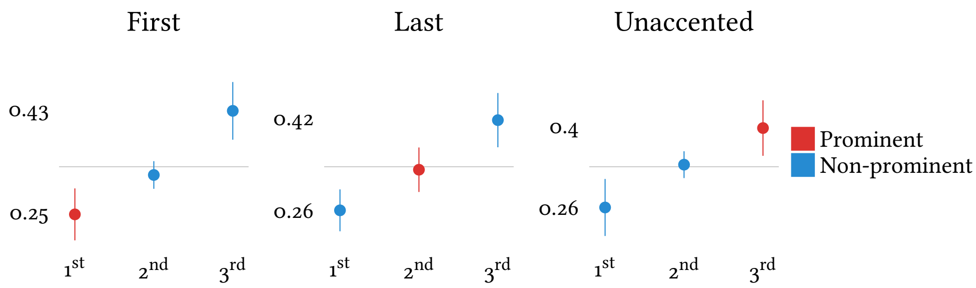


Figure 5.7: Distribution of the duration of Japanese isolated words per syllable for initial- and final-accented words, and for unaccented words. The gray ● line marks the point at which all syllables would have the same length. ($\bar{x} \pm \sigma$).

Klatt and Klatt, 1990). Whether this decision was worth the costs is up to the reader for consideration.

5.4.1 The Spanish stimuli

Differences in the syllable duration of the original stimuli, described in detail in section 2.3.2 and shown in figure 5.8, resulted in significant differences between the A and B continua, and therefore in suitable stimuli to highlight the sensitivity to duration as a secondary cue for the perception of L2 prominence. Results from chapter 3 showed that non-native participants are very capable of correctly perceiving the position of the prominence in Spanish words in isolation, and that they are furthermore sensitive to duration as a cue. Based on this, J1S2 are predicted to show response curves that differ between the A and the B continua, and to show similar categorical boundaries to those exhibited by native speakers. A more detailed break down of the proportion of the word that belonged to each syllable, for each of the original tokens used as the source for the manipulations, is shown in tables E.1a and E.1b, in the appendix.

This is the case particularly with the first contrast (continua A_1 and B_1). The lower performance shown by J1S2 participants with oxytone words even in isolated contexts (shown in figure 3.12) mean that the continua for the second contrast will likely elicit fewer correct responses, particularly towards the oxytone end of the scale. This relatively low performance with prominent syllables towards the end of the word was explained in section 3.3.1 with reference to the effects of duration, and in particular to those

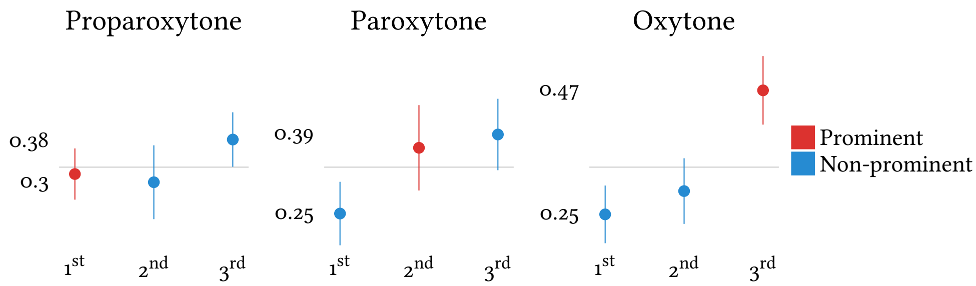


Figure 5.8: Distribution of the duration of Spanish isolated words per syllable for each of the three positions of the prominence. The gray ● line marks the point at which all syllables would have the same length. ($\bar{x}\pm\sigma$).

of the older male speaker. If this is indeed the case, then it should be clearly reflected on the responses to these stimuli, since half of them come from the speaker in question.

In a design like the one used here, there is also a clear possibility that stimuli at the extremes of the continua will suffer more from the conflict that exists between cues from different words. In this case, this conflict will be greatest for stimuli at the beginning of the B continua and the end of the A continua. Some degree of additional difficulty might become apparent for groups that are particularly sensitive to changes in duration.

5.4.2 The Japanese stimuli

In the case of the Japanese stimuli, the little duration variation in the original stimuli described in section 2.3.2, and shown in figure 5.7, resulted in a great similarity between the stimuli from continuum A and B. However, chapter 4 suggested an effect of secondary cues even under these conditions, an effect that was explained as an over-sensitivity to random variations in intensity and duration outside of the control of the native speakers. If that is the case, then what matters is not so much the duration variation in the production of native speakers, but the heightened sensitivity to duration changes in the non-native listeners; and duration is known to be a highly significant cue for the perception of Spanish prominence (Ortega-Llebaria, 2006).

A difference in the responses for both continua, in the face of the small differences apparent in the original stimuli, would strongly support the conclusions of chapter 4 as they relate to duration; while the lack of such a difference would suggest that the effect is either too small for this design to reveal it, or lie more on intensity differences than in those of duration.

In the case of the Japanese study, a more pressing question relates to the degree of categorical development in S1J2 participants. Results from chapter 4 showed that non-native performance for Japanese unaccented words was particularly low, with participants reaching chance level responses in both minimal trios and pairs. The task in that occasion was one of identification, for which some knowledge of the categories to be identified is a requirement. In this case, the low performance was explained as the lack of that knowledge, and in particular the poor development of an appropriate unaccented category.

If this is the case, S1J2 participants would be expected to show differences in the category slopes for the first and the second contrast, with category slopes between accented categories being much greater than those between accented and unaccented words.

Chapter 6

Effects of duration on the perception of Spanish as L2

Chapter 5 explained the motivation of a set of studies on the sensitivity to non-native secondary acoustic cues and the development of L2 accentual categories, and provided details about the preparation of the speech materials to be used in those studies. This chapter reports the results of the application of this methodology to the case of Japanese students of Spanish (J1S2).

6.1 Design

As explained in the previous chapter, the studies in part 2 focus on two main research questions:

1 Does duration have an effect on the perception of lexical prominence by non-native learners?

2 Do non-native participants show effects of their L1 in the weighing of secondary L2 acoustic cues?

Research variables

Two groups of synthetic continua were generated to be used as the stimuli for this study. The groups had the syllable durations of the words at opposing ends of the continuum, such that continua labelled "A" always had syllables with the duration of those in the first term, and those labelled "B" always had the syllable duration of the second. Based on the test paradigm

Digital materials related to this study can be found in the study's repository at http://www.pinguinorodriguez.cl/research/phd/secondary-cues/spanish.

described in section 5.1.1, the sensitivity to changes in duration would be operationalised the magnitude of the shift in the response curves for both continua. The magnitude of the category slopes, on the other hand, would provide information on the definition of those categories.

Linguistic questionnaire

To facilitate the comparison between the results of this experiment, and those of the study reported in chapter 3, participants were presented with a preliminary linguistic questionnaire which contained questions about their linguistic background, as well as their self-reported levels of L2 proficiency and use. More details about the preparation and suitability of this questionnaire are covered in section 1.4.2.

6.1.1 Participants[1]

Japanese participants: J1S2

A total of 24 (21 female and 3 male) Japanese participants between the ages of 18 and 32 ($\bar{x}$= 21.7, σ= 3.3) were recruited from the student body enrolled in the 4-year Spanish programs at Sophia University, Tōkyō University of Foreign Studies, and Waseda Universities, all of them in Tōkyō.

Figure 6.1 shows the distribution of participants according to the three self-reported measures of second language proficiency: length of studies, overall proficiency, and frequency of use. Based on their accounts, most participants were in a beginner-to-intermediate level ($\bar{x}$= 2.25 in a 5-point scale; σ= 0.86), and used their L2 on average close to once per week ($\bar{x}$= 2.73, σ= 0.84, with 3 meaning once a week).

Chilean participants

Chilean controls were recruited from the city of Concepción, in the south of the country, and tested at the local Universidad de Concepción. A total of 35 (18 female and 17 male) participants from Chile between 17 and 32 years old ($\bar{x}$= 21.6, σ= 3.2) participated in the test.

6.1.2 The test procedure

The test was separated into three stages:

1 The linguistic questionnaire mentioned above.

2 A practice stage, during which participants were shown examples of the

1 Special thanks to Hernán Acuña in Concepción; and Misae Kunitake and Kazuaki Yasuki in Tōkyō for assistance with participant recruitment and data collection.

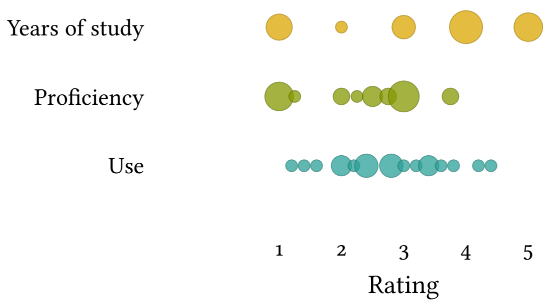


Figure 6.1: Self-reported measures of L2 proficiency of S1J2 participants. The horizontal position of each circle marks the mean score reported by each participant to the three scales. In the case of proficiency and use, participants were given a 5-point scale, while they were free to enter how many years they had studied for. The area of each circle reflects the number of participants for that particular score.

questions and given feedback on their answers.

3 A testing stage, with 6 different testing blocks separated by breaks.

Participants wore headphones and took the test in a quiet room under experimenter supervision. Each testing session (including all 3 stages and the breaks) lasted ~45 minutes in total.

Local testing

In an attempt to increase the number of participants, the studies reported in part 1 used an open registration process in which participants were allowed to take the test unsupervised, and using any computer connected to the internet. However, this did not result in the large number of participants that were anticipated.

As a result, and to prevent some of the methodological issues entailed by that approach (and discussed in section 3.1.2), participants in the current study were asked to register prior to the test and attend one of many supervised testing sessions hosted locally. Despite these changes, the same testing platform used for part 1, which allowed for both local and remote testing, was used here as well.

Each session took place in a quiet room with up to 15 participants at a time. All participants were provided with headphones which they were asked to wear during the test. They were allowed to set the loudness of the test to a comfortable level before starting.

Presentation

The 40 ($2 \times 2 \times 10$) tokens resulting from the synthesis explained in chapter 5 were compressed as 192kB/s MP3 files using `lame`[2] to limit the transmission rate from the test server without sacrificing too much of the original quality of the recordings. All tokens were presented once per testing block for a total of 240 stimuli.

The presentation of the stimuli was identical to that used in the study reported in chapters 3 and 4, and described in section 3.1.2: after listening to each token, participants were presented with three buttons labelled with the individual items in the trio to which the word belonged. From these, they were asked to choose the one corresponding to the word they thought they had heard. The only difference was that in this case the stimuli were all in isolation and had been synthetically manipulated. The recorded sound played automatically each time the question loaded, but participants were allowed to replay it up to 2 times (see section 3.1.2).

As in the previous study, all instructions and button labels was presented using the standard orthography of the target language (in this case, Spanish).

The practice stage

Before going into the testing state, participants were shown sets of six natural training items from two speakers in a random order. Items used for this stage came from the same speakers used for the real stimuli, but were selected from the unused recordings to ensure that they were not part of the testing stage to prevent effects of habituation. The presentation of the practice items was identical to the one described above, with the exception of additional feedback presented to the participants telling them whether their response had been correct or not.

After going through all six items once, participants who had incorrectly identified one or more items would be told to go through the practice session again. They could do this up to four times. During this time, they were encouraged to ask questions should any part of the test procedure be unclear. Only participants who successfully identified all six natural items in any of the attempts were allowed to continue.

2 Version 3.99.5, available at `http://lame.sourceforge.net/`. Encodings were made using its `-h` flag, which ensures the highest quality available.

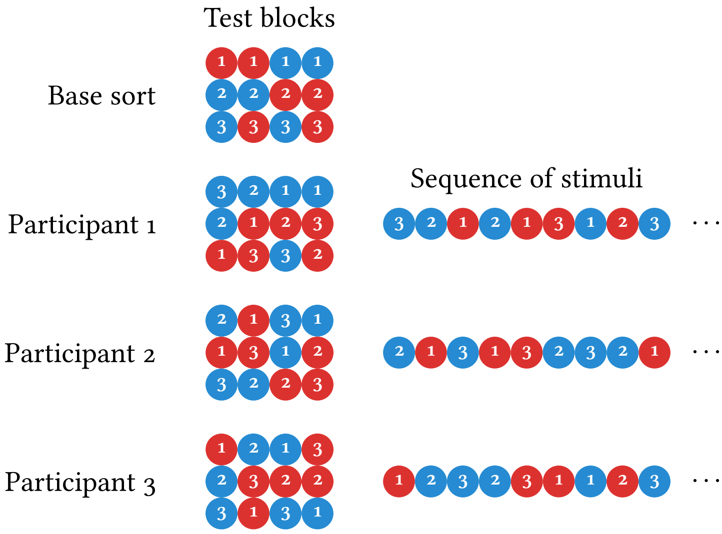


Figure 6.2: The stimuli randomisation scheme. A sequence of speakers is first chosen for each stimulus across testing blocks, and the order of stimuli within in each block is later randomised per participant. The resulting sequence is shown on the right. Numbers represent individual stimuli; colours represent speakers.

Stimuli within each testing block were presented in a pseudo-random order that was generated anew for each participant. The randomisation of the stimuli was done in two steps, to accommodate for limitations in the software that was chosen to run the tests. Randomisation

1 During the first step, speakers were randomly assigned to each of the 6 repetitions of each token across testing blocks, so that half of the repetitions were produced by each speaker. This assignment remained constant for all participants

2 The second step, which took place per participant, determined the order in which each stimulus would appear within each testing group.

All participants heard each stimulus in the same sequence of speakers across all testing groups; but the position of each stimulus in each testing block was different per participant. A graphical representation of this scheme is shown in figure 6.2.

6.2 Results

Responses were analysed using probit analysis (Nittrouer and Studdert-Kennedy, 1987; Mayo and Turk, 2004, 2005) on the responses of each participant for each of the two contrasts: first between proparoxytone and paroxytone words, and then between paroxytone and oxytone.

The fitted probit models were then used to predict the position of the category boundary and slope for each participant. When a participant's responses for a given category did not cross the 50% mark, they were marked as having no boundaries. Category slopes were derived from the model coefficients, indicating the rate of change of the identification function. The higher the value of the slope the steeper the response curve is. A response curve with slopes close to zero results from a participant not identifying the terms in the contrast as members of different categories.

Values for category boundaries and slopes per participant became the dependent variables in separate 3-way ANOVAs, using the response category, the language group (native or non-native), and the step in the continua as the predictors. All tests were run using the R statistics package.

To examine the possible effects of L2 proficiency on the task, an aggregate measure was calculated for the participants' self-reported language use and proficiency by averaging across the multiple ratings given for each in the linguistic questionnaire (see section 1.4.2 for details). The ANOVAs were repeated excluding native participants (and therefore the language group predictor) and including these measures as predictors.

Overall results for both continua are shown in figures 6.3 and 6.5, for the first and the second contrasts respectively. The figures show the percentage of responses for each category at every step in the continua, for both native and non-native listeners. Results in the figure are averaged across participant responses for both speakers. In each figure, the difference between the plots on the left and those on the right reflects the effect of the manipulations in syllable duration.

6.2.1 First contrast: proparoxytone *v.* paroxytone

Figure 6.3 shows the responses for the first contrast. Stimuli that were manipulated to have the duration of paroxytone words (on the right side of the figure) show a distribution of responses that is close to the canonical dis-

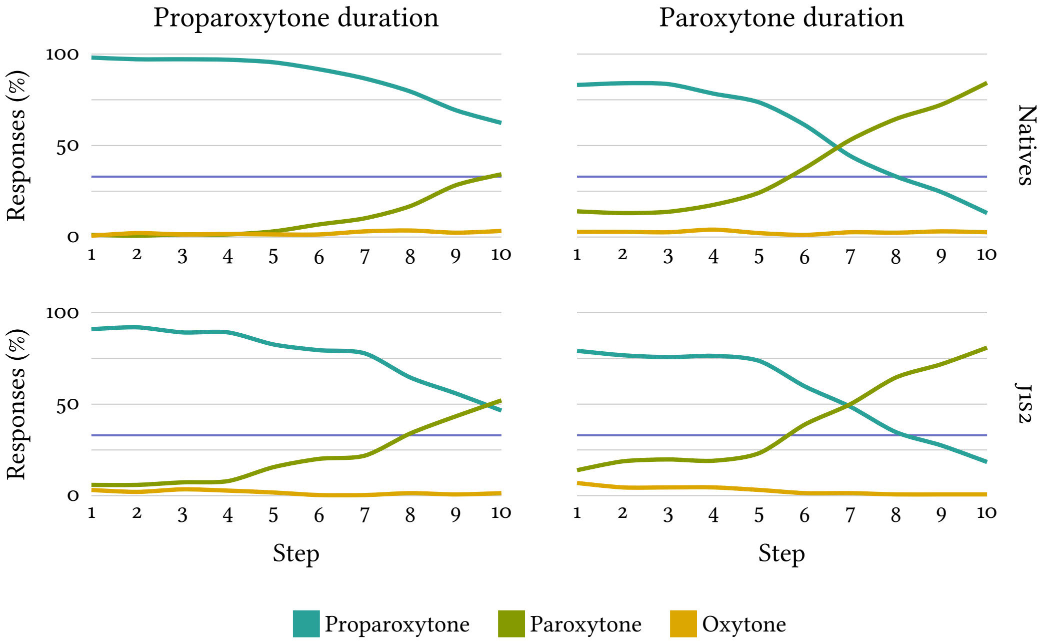


Figure 6.3: Percentage of responses per category at each step in the continua for the first contrast across all participants and speakers. Stimuli with paroxytone and oxytone responses are shown in the left and right columns respectively, while rows separate responses between language groups. The violet ● line marks the chance level. Compare with Japanese results in figure 7.2, on page 197.

tribution in studies of this design, with high percentages of contrasting responses in opposite extremes of the continuum. Stimuli that had the syllable durations of proparoxytone words, on the other hand, showed much shallower slopes and a much greater number of participants whose responses did not cross the 50% threshold. This was true for both the native and the non-native groups.

Category boundaries

Sensitivity to duration

Results showed a strong effect of syllable duration across groups, with the difference between the boundary positions being statistically significant ($F_{(1,156)} = 95.94$, $p < 0.001$). As expected, stimuli with the duration of proparoxytone words resulted in more proparoxytone responses and later boundary positions ($\bar{x} = 7.01$, $\sigma = 1.75$) than those of the stimuli with the paroxytone durations ($\bar{x} = 5.31$, $\sigma = 2.06$).

No other significant effects on boundary positions were found for this contrast. This was likely due to the number of cases in which no category boundaries could be determined: out of the 144 measures taken from J1S2 participants ($24 \times 2 \times 3$), only 75 (52.08%) had category boundaries; and similar results were obtained for the 35 native participants (89 out of 210 measures, or 42.38%). However, these numbers include responses for all categories, and proparoxytone responses were expected to be low because that category was not part of this contrast.

Category slopes

Main effect of accent category

As expected from figure 6.3, analysis of category slopes in this contrast revealed significant differences between accent categories ($F_{(2,342)} = 29.17$, $p< 0.001$), with mean values for the response functions for proparoxytone ($\bar{x}= -0.43$, $\sigma= 0.84$) and paroxytone ($\bar{x}= 0.49$, $\sigma= 0.94$) being significantly greater than those for oxytone words ($\bar{x}= 0.11$, $\sigma= 1.04$). This was expected, since in both contrasts one of the categories that the participants were presented with was not present: it was a 3AFC task with 2-way contrasts. Because of this, participants were expected to have low responses overall for this *third* category, and therefore flat response functions; and this is what was found in this case. Figure 6.4 shows the results for slopes per category.

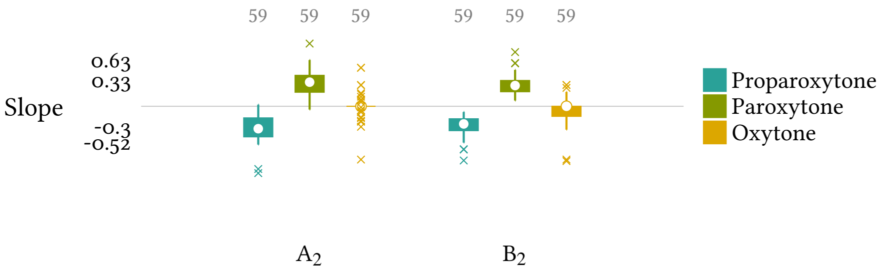
Interaction between category and continuum

A significant interaction between category and continuum was also found ($F_{(2,342)} = 4.54$, $p< 0.05$), with the difference in the slope of the response curves of both continua being greater for paroxytone words (0.39) than for proparoxytone words (0.32). On the other hand, the difference between the slopes of the two *active* categories (*i.e.* proparoxytone and paroxytone) was greater for native listeners (1.09) than for J1S2 listeners (0.67), but this difference was just outside significance ($F_{(2,342)} = 2.8$, $p= 0.062$). Figure 6.4 shows the results for both of these interactions per separate.

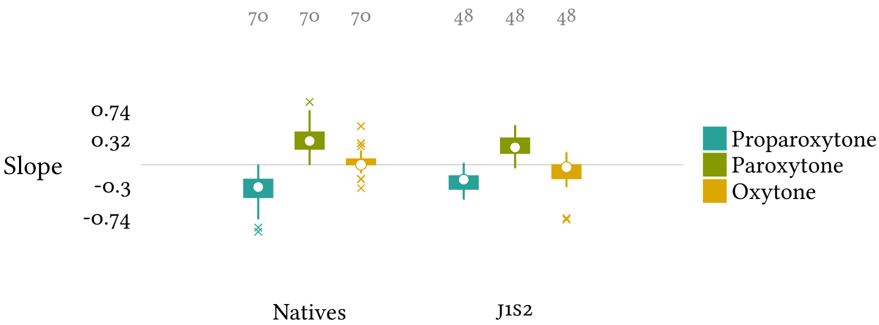
L2 proficiency

Only marginal interaction: category×proficiency

No significant main effects of either measure of L2 proficiency were found for category boundaries or slopes. In the case of slopes, an interaction between category and proficiency was found to be just outside the threshold of significance ($F_{(2,120)} = 2.49$, $p0.087$).



(a) By continuum and category, first contrast



(b) By language group and category, first contrast

Figure 6.4: Distribution of category slope values per category and language group (b) and by continuum (a) in the first contrast. The gray ● line indicates a flat response function. Numbers above each box indicate the corresponding *n*.

6.2.2 Second contrast: paroxytone *v.* oxytone

Category boundaries

Sensitivity to duration

Responses in the second contrast, shown in figure 6.3, reveal a smaller (but still highly significant; $F_{(1,226)} = 164.13$, $p < 0.001$) main effect of syllable duration. In this case, stimuli with the duration of paroxytones (in continuum A2) had later boundary positions ($\bar{x} = 6.24$, $\sigma = 1.48$) than those with the duration of oxytones (in continuum B2; $\bar{x} = 4.04$, $\sigma = 1.77$).

Effects of category and group

A main effect of category was also significant ($F_{(2,226)} = 38.09$, $p < 0.001$), with mean boundary positions for paroxytone words occurring earlier ($\bar{x} = 4.84$, $\sigma = 1.97$) than those for oxytone words ($\bar{x} = 5.88$, $\sigma = 1.64$). Meanwhile, a significant main effect of language group was also found ($F_{(1,226)} = 43.34$, $p < 0.001$), with category boundaries for non-native listeners ($\bar{x} = 5.71$, $\sigma = 2.06$) occurring closer to the end of the word than for native listeners ($\bar{x} = 4.66$, $\sigma = 1.8$) across continua.

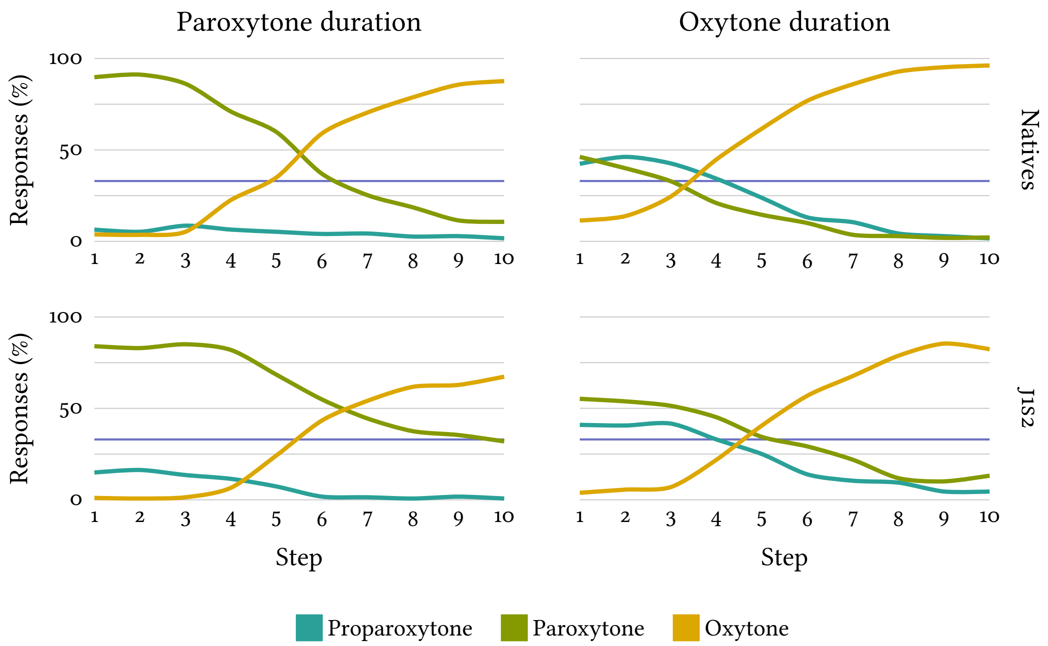


Figure 6.5: Percentage of responses per category at each step in the continua for the second Spanish contrast across all participants and speakers. Stimuli with paroxytone and oxytone responses are shown in the left and right columns respectively, while rows separate responses between language groups. The violet ● line marks the chance level. Compare with Japanese results in figure 7.6, on page 200.

These two main effects also had a significant interaction ($F_{(1,226)} = 3.13$, $p < 0.05$). Figure 6.6 shows the results for this effect, and illustrates that while boundary positions for paroxytones were similar for both language groups, J1S2 listeners placed the boundary of their oxytone category significantly later in the continua.

Effects of duration per category

This difference between the boundary positions of accentual categories was also susceptible to changes in syllable duration, as shown by the significant interaction between continuum and category ($F_{(1,226)} = 11.61$, $p < 0.001$). These results are shown in figure 6.7, which shows that the change in boundary position was greater for paroxytone (2.61) than for oxytone responses (1.46).

Category slopes

Different slopes per category

Slopes of the response functions of different accentual categories were found to be significantly different for this contrast ($F_{(2,342)} = 871.73$, $p < 0.001$), with the oxytone category having the steepest mean slope

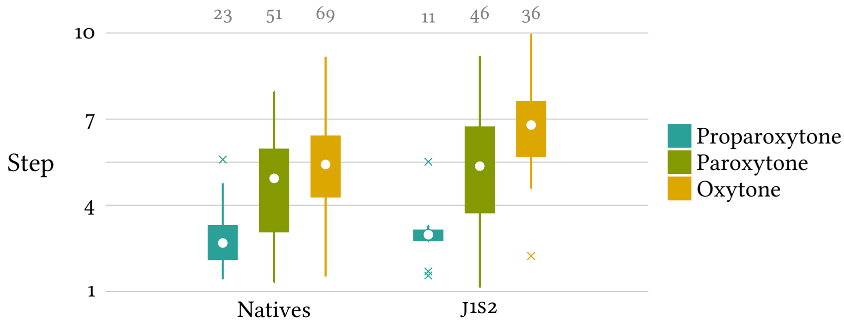


Figure 6.6: Distribution of category boundary positions per group and accent category for the second Spanish contrast. Numbers above each box indicate the corresponding *n*, after discounting participants for which boundary positions could not be estimated.

($\bar{x}$= 0.47, σ= 0.18), followed by that of paroxytones ($\bar{x}$= −0.32, σ= 0.18). The slope of proparoxytone responses was much smaller ($\bar{x}$= −0.19, σ= 0.15), but still greater than the value found for the first contrast, since in this case there were a considerable number of responses for the *absent* category, particularly in the first half of the B_2 continuum.

Effects of duration per accentual category

No main effect of the duration manipulation was found for category slopes, but the difference alluded to in the previous paragraph did result in a significant 2-way interaction between the continuum and the accentual category ($F_{(2,342)} = 12.42$, $p < 0.001$). Figure 6.8a shows that when syllable duration was that of paroxytone words, proparoxytone responses had slopes that were indistinguishable from zero ($\bar{x}$= −0.14, σ= 0.14), but a shift in syllable duration to that of oxytone words (in continuum B_2) caused the magnitude of the response function slope to increase significantly ($\bar{x}$= −0.26, σ= 0.12).

Differences between categories and language groups

Another significant 2-way interaction was found between accent category and language group ($F_{(2,342)} = 18.81$, $p < 0.001$). The driving factor in this case was the difference in paroxytone slopes for the native listeners ($\bar{x}$= −0.38, σ= 0.17), which were significantly greater than those of the J1S2 listeners ($\bar{x}$= −0.22, σ= 0.14). More detailed results for this interaction are shown in figure 6.8b.

Three-way interaction: category×group×cont.

Part of the variance of these two interactions is explained by the presence of a significant 3-way interaction between accentual category, lan-

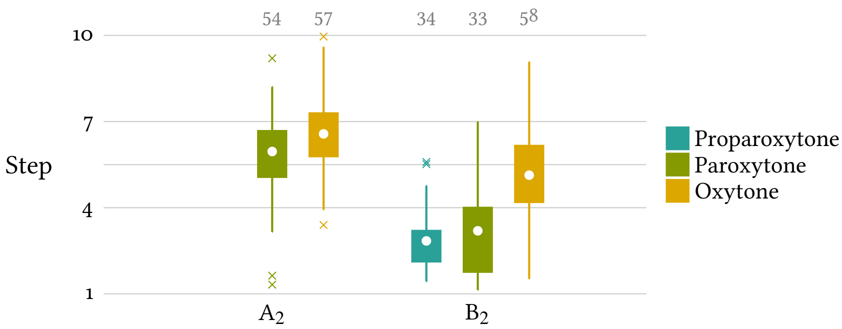
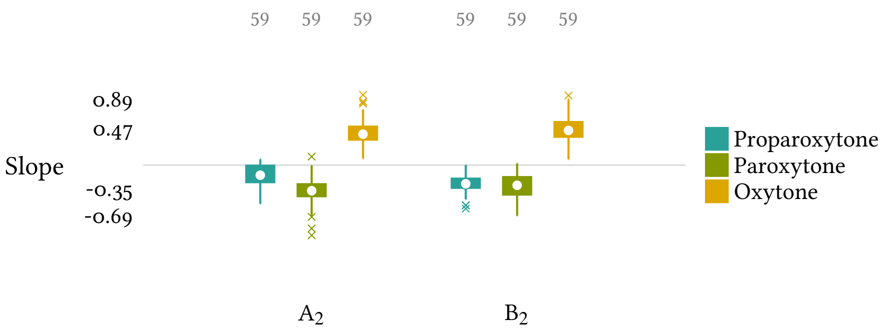
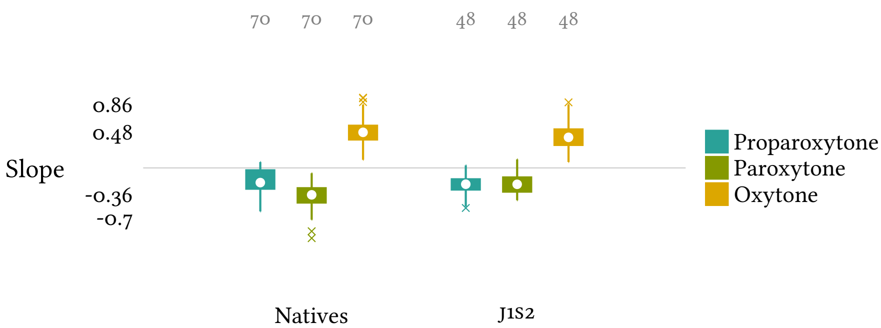


Figure 6.7: Differences in the effect of duration manipulations on boundary positions by accent category for the second Spanish contrast. Numbers above each box indicate the corresponding *n*, after discounting participants for which boundary positions could not be estimated.



(a) By continuum and category, second contrast



(b) By language group and category, second contrast

Figure 6.8: Distribution of category slope values per category and language group (b) and by continuum (a) in the second contrast. The gray ● line indicates a flat response function. Numbers above each box indicate the corresponding *n*.

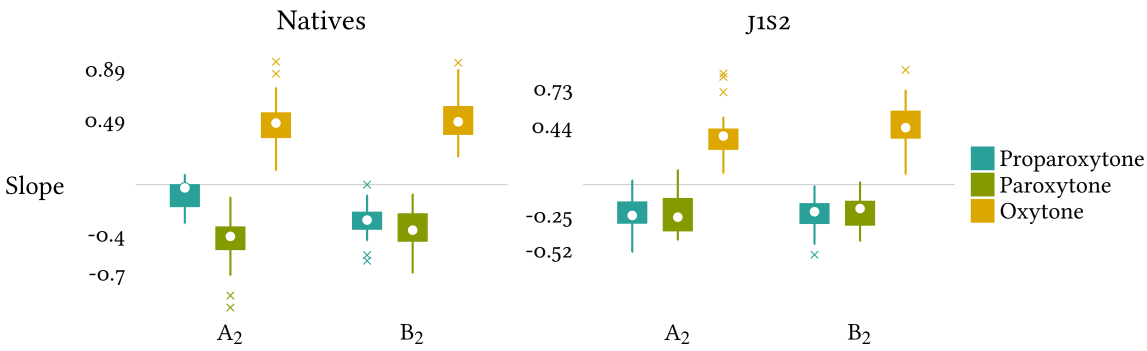


Figure 6.9: Change in category slopes for each category between continua for the between paroxytone and oxytone words. Higher values indicate the rate of change per step in the continuum. The gray ● line indicates a flat response function. (n =59/box).

guage group, and continuum ($F_{(2,342)} = 5.11$, $p < 0.01$), which is shown in figure 6.9. The main difference in this case is the large increase in the mean slopes for native proparoxytone responses between those for stimuli with the duration of paroxytones ($\bar{x} = -0.08$, $\sigma = 0.11$), and those for stimuli with the duration of oxytones ($\bar{x} = -0.28$, $\sigma = 0.12$).

L2 proficiency

Weak effects of proficiency and use

In this contrast, unlike the previous one, a weak main effect of proficiency on category boundaries was found ($F_{(2,73)} = 5.49$, $p < 0.05$), with participants who were more proficient showing a small tendency to have earlier boundaries. A 2-way interaction for category slopes between language use and category was found ($F_{(2,120)} = 5.19$, $p < 0.01$), with participants who reported using their L2 more frequently across contexts appearing to have steeper slopes for the oxytone category only, while no change was shown for the other two categories.

Marginally significant interactions

Other effects were found to be just outside the threshold of significance. These included a main effect of use ($F_{(2,73)} = 3.85$, $p = 0.054$) and an interaction between proficiency and category ($F_{(2,73)} = 2.84$, $p = 0.065$) for boundary positions; and a three-way interaction between category, language use and language proficiency ($F_{(2,120)} = 2.74$, $p = 0.069$) on category slopes.

6.3 Discussion

6.3.1 Sensitivity to duration

Similar sensitivity to duration for both language groups

As predicted in section 5.4.1, results from this study show that J1S2 participants are not only sensitive to duration changes for the perception of non-native lexical prominence, but that, at least in the first contrast, this sensitivity is similar to that of native speakers. When listening to stimuli from the first contrast, participants from both the native and the non-native groups showed significantly different boundary positions and slopes depending on whether the stimuli had the duration of a proparoxytone or a paroxytone word. But language group was not a significant predictor of responses either on its own or in combination with other factors. This degree of similarity is surprising even in the light of the native-like performance that J1S2 participants showed for isolated words in chapter 3, since these students had only ever studied Spanish in FLA environments.

The second contrast

Responses for the second contrast show a greater difference between native and non-native participants, with non-native speakers having different boundary positions than natives per category. But notably absent as a significant effect is the interaction between language group and continuum that would denote a difference in the sensitivity for duration as a cue.

Based on the relatively low performance exhibited by J1S2 participants with the prominence of oxytone words (even in isolation), responses for the second contrast were predicted to be lower in the second contrast, and lower in particular for oxytone words. This was the case, as is reflected in the lower maximum number of oxytone responses for both continua in figure 6.5, and the significantly lower slopes for non-native oxytone responses in continuum A2 in figure 6.9.

Effects of conflicting cues

The lower non-native performance with word-final prominence gets confounded with the effects of conflicting acoustic cues, which is likely what explains the effect shown in particular for continuum A2. As explained in section 5.4.1, the conflict between the information encoded in the duration and the F_0 of the stimuli would be strongest at the beginning of the B continua and the end of the A continua, which would predict that a population that was particularly sensitive to the interaction between those cues would show poorer performances in those stimuli. Interestingly, the results not only show precisely this effect (which is much stronger for the B2 continuum), but they also show that it affects both language groups in remarkably similar manners.

Still, responses for the second contrast in particular show some differences between language groups: S1J2 listeners tend to place their boundaries later, which is an indication that, while both groups are sensitive to duration, the evaluation that S1J2 learners make of that cue is subject to transfer effects from Japanese. But the effects are small, and they only became apparent for the second contrast. This makes it possible that there is a connection between these transfer effects and their additional difficulty with prominence that occurs late in the word. Further work is needed to fully explore this possible connection. Cue-weighting

6.3.2 Language proficiency

Unlike this study, the study reported in chapter 3 did not show any effects of language proficiency. This was originally explained as an effect of the poverty of the FLA learning environment and the lack of priority in prosodic training. Results from this study, however, question those conclusions.

Although participants for both studies come from different pools, they represent the same populations: they have comparable age ranges and distributions and they reported similar levels of language proficiency. There is a possibility that their learning environments were different, with most of the participants for this study coming from Sophia University, a university in which no recruitment was done for the previous study. But both groups of participants were FLA students, and in an exit interview that took place after the experiment was concluded, participants from this study did not report going through any special prosodic training.

It is much more likely that the specifics of this design allow for more granularity in the detection of the effects of L2 proficiency, and that the learning conditions that the participants were under made the effect too small to be detected before this. Although this dissertation did not perform any longitudinal analyses, the performance of participants with different overall measures of proficiency (including their reports of proficiency itself and language use) can be understood as snapshots of the performance at different stages in the development of language proficiency. Understood in this sense — and with the caveat that participants represented a relatively narrow scope of proficiency measures — results can be interpreted as a sign that improvement through instruction is possible. Considering the learning environments of these participants, this means that a more targeted training could have a much greater impact, which should serve as a considerable motivation for prioritising prosodic training in the future.

Chapter 7

Duration effects on the perception of Japanese as L2

The previous chapter reported the results of a study on the effects of duration as a secondary cue for the perception of Spanish lexical prominence. This chapter presents the results of the application of the same methodology (described in chapter 5) to the case of S1J2.

7.1 Design

Why duration in Japanese?

Like in the study reported in the previous chapter, the main motivation for this study was also to examine the effect that secondary acoustic cues had on the perception of non-native suprasegmental categories. However, while the procedure in both cases was kept to maintain maximum comparability between both studies (see section 1.6.1), the question that was of most interest in the case of Japanese as an L2 was different.

Both this study and that of chapter 6 focus on the effects of duration in particular. This decision was an obvious one for Spanish, since duration is the main secondary cue for the perception of stress (see section 1.3.1), but as explained in section 1.3.2, duration does not play any role whatsoever in the encoding or the perception of Japanese lexical prominence.

Indeed, figure 5.7 (on page 173) shows that, while different syllables will

Part of the results presented in this chapter were presented at the 18th meeting of the International Congress of Phonetic Sciences (ICPhS) in Glasgow, Scotland, in the summer of 2015. Digital materials related to this study can be found in the study's repository at `http://www.pinguinorodriguez.cl/research/phd/secondary-cues/japanese`.

have different durations in Japanese, there is no large difference in the distribution of syllable durations across words with different accent types. This contrasts strongly with the case of Spanish, illustrated in figure 5.8, which shows a strong interaction between the lengthening of final syllables and that of prominent syllables. However, results from chapter 4 suggested that S1J2 listeners might be sensitive to non-systematic variation in the cues that Japanese does not consider relevant, like duration and intensity, of which duration was particularly interesting (see section 5.1).

The chosen methodology makes it possible to examine the side effects of whatever small changes in duration may exist, while at the same time illustrating the degree of category development in L2 listeners, which in this case was a much more intriguing question. This because chapter 4 showed extremely low performances for S1J2 participants in the identification of a contrast that, from an auditory perspective, should not have been as difficult: that between accented and unaccented words.

Research variables

Like in chapter 6, the sensitivity to duration as a secondary cue was operationalised as the magnitude of the shift in the response functions across continua (and therefore of the categorical boundaries of those functions); while differences in the weighing would be made evident by differences in the shift across language groups. The degree of category development would be correlated with the steepness of those response functions, measured by the rate of change of that curve, or its slope. Additionally, participants in this study filled in the same linguistic questionnaire as in the previous chapter, providing information about their language proficiency. This would make it possible to measure the effects of language training on their perception.

Linguistic questionnaire

7.1.1 Participants[1]

Chilean participants: S1J2

A total of 34 (26 female and 8 male) Chilean participants between the ages of 14 and 33 ($\bar{x}$= 22.3, σ= 4.1) were recruited from the student body enrolled

1 Special thanks to Rosa Catalán, Imran Hossain, and Karina Cerda, and to professor Domingo Román in Santiago; and professors Arai Takayuki, Kimiyo Nishimura, and Yasuaki Shinohara in Tōkyō for assistance with participant recruitment and data collection.

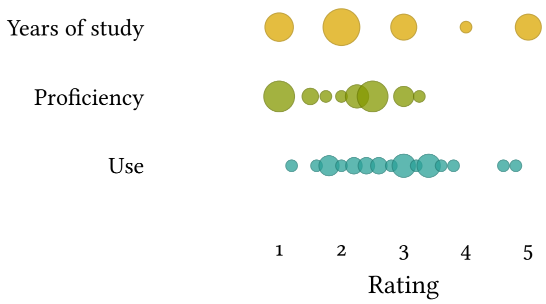


Figure 7.1: Self-reported measures of L2 proficiency of S1J2 participants. The horizontal position of each circle marks the mean score reported by each participant to the three scales. In the case of proficiency and use, participants were given a 5-point scale, while they were free to enter how many years they had studied for. The area of each circle reflects the number of participants for that particular score.

in the undergraduate Japanese translation program at Universidad de Santiago, in Santiago. Figure 7.1 shows the distribution of participants according to the three self-reported measures of second language proficiency: length of studies, overall proficiency, and frequency of use.

Japanese participants

Japanese controls were recruited in Tōkyō largely from the student body of the Tōkyō University of Foreign Studies and Sophia University, in Tōkyō. A total of 35 (18 female and 17 male) participants from the Kantō region between 18 and 49 years old ($\bar{x}$= 23.6, σ= 6.8) participated in the test.

Before the test, participants were asked to fill a linguistic background survey providing broad information about their L2 proficiency and use. Reports from S1J2 participants were very similar to those obtained from J1S2 participants in the previous test: the large majority of participants claimed to be in a beginner-to-intermediate level ($\bar{x}$= 2.06, σ= 0.77), and to use their L2 just under once per week ($\bar{x}$= 2.81, σ= 0.86).

7.1.2 The test procedure

The test followed exactly the same procedure as the one described in detail in section 6.1.2: participants took the test locally in Chile and Japan wearing headphones provided by the examiners. Testing session similarly lasted ~45 minutes, including regular breaks. Participants listened to each of the 40 tokens, once per test block, for a total of 240 stimuli (40×6), half of them with each speaker, and were tasked with identifying the word they had heard from a list.

Like in chapter 4, all instructions and button labels in the test were presented in Japanese, using the standard Japanese orthography in *kanji* and *hiragana* script. Since Japanese orthography does not mark the position of the accent, unlike the Spanish orthography, the accented syllable was marked in red. This included the items in the practice rounds, during which any questions they had on the procedure could be clarified by the examiners.

7.2 Results

Analysis of the responses followed the exact same procedure as used in chapter 6.

Figures 7.2 and 7.6 provide an overview of the obtained responses by showing the mean percentage of responses per category for each step in the continua, separated horizontally by continua (*i.e.* by the reference point used for syllable duration) and vertically by language group.

7.2.1 First contrast: initial- *v.* final-accented

Native responses for stimuli from the first contrast (shown at the top of figure 7.2) show the expected pattern of responses, with very high percentages of responses for the categories involved in the contrast (*i.e.* first- and final-accented) in the appropriate extremes of the continua, and extremely low numbers of responses throughout for the third category (*i.e.* unaccented). Steep slopes also show that the responses of native participants are categorical in nature, which is evidence that the possible artifacts introduced by the synthesis did not have a large impact on their responses, as suggested by the initial evaluation of the stimuli.

Non-native responses, in the same figure, show a similar pattern of categorical perception, albeit vertically compressed at the extremes. A side effect of this compression is an increase in the number of unaccented responses, but even though they reached an average of 25% of the total at the rightmost step of both continua, they still maintained a low number

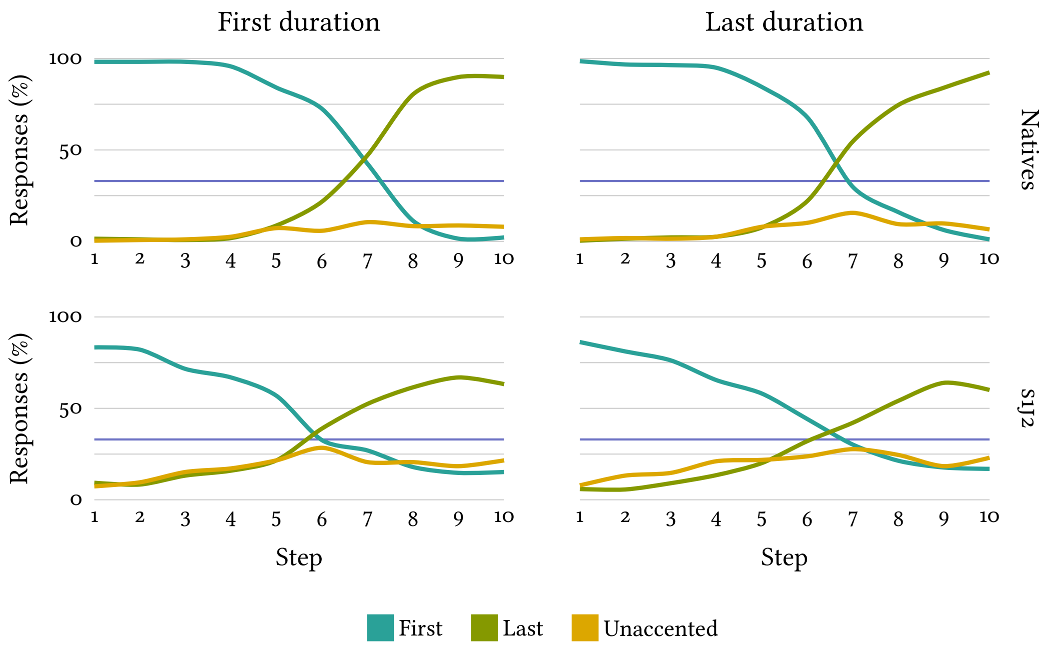


Figure 7.2: Percentage of responses per category at each step in the continua for the first contrast across all participants and speakers. Stimuli with initial- and final-accented durations are shown in the left and right columns respectively, while rows separate responses between language groups. The violet ● line marks the chance level. Compare with Spanish results in figure 6.3, on page 183.

throughout.

Category boundaries

A main effect of language group was found for the position of category boundaries ($F_{(1,201)} = 10.37$, $p < 0.01$), with those of native listeners ($\bar{x} = 6.84$, $\sigma = 0.83$) occurring slightly later than those of S1J2 participants ($\bar{x} = 6.34$, $\sigma = 1.61$). A difference between the boundary position of the categories in the contrast was also found ($F_{(1,201)} = 94.27$, $p < 0.001$). In a 2-way contrast, boundaries understood as the point at which responses cross the 50% mark should not differ between categories. But in this case, the presence of a significant number of unaccented responses made the boundaries differ, such that the boundary of first-accented words ($\bar{x} = 5.86$, $\sigma = 1.27$) occurred earlier than those of final-accented words ($\bar{x} = 7.29$, $\sigma = 0.94$).

This is not entirely attributable to the responses of S1J2 listeners, however, as attested by the presence of a significant 2-way interaction between

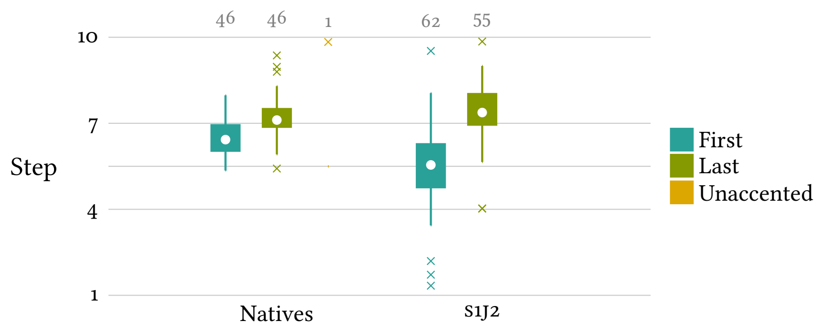


Figure 7.3: Distribution of category boundary positions per group and accent category for the first Japanese contrast. Numbers above each box indicate the corresponding *n*, after discounting participants for which boundary positions could not be estimated.

language group and accent category ($F_{(1,201)} = 57.1$, $p < 0.001$), shown in figure 7.3. As shown in the figure, while the difference between the position of the boundary for non-native listeners (1.99) is much greater than that of native listeners (0.69), this latter difference is also significant ($t_{(85.56)} = -4.83$, $p < 0.001$).

Category slopes

The fact that both groups had particularly low responses for the category that was absent from the contrast was reflected in a significant difference in the slopes of those categories ($F_{(1,330)} = 688.66$, $p < 0.001$), with that of the unaccented category being very close to zero ($\bar{x} = 0.09$, $\sigma = 0.08$), and the slopes of the other categories being at more or less equal distances from it ($\bar{x} = -0.53$, $\sigma = 0.36$; $\bar{x} = 0.48$, $\sigma = 0.33$, for initial- and final-accented words respectively).

This difference in slopes for different categories varied across language groups, as shown by a significant 2-way interaction between these two factors ($F_{(1,330)} = 155.19$, $p < 0.001$). This interaction, shown in figure 7.4, reveals that native speakers gave more categorical responses, with the difference between mean slopes for the initial- and final-accented categories being much greater for native listeners (1.6) than for S1J2 (0.62).

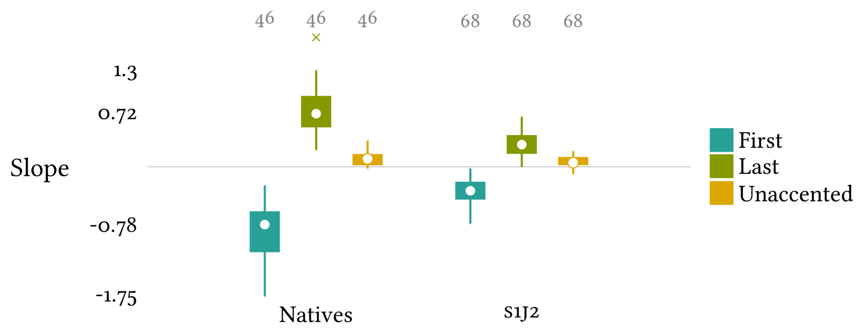


Figure 7.4: Distribution of category slopes per group and accent category for the first Japanese contrast. Numbers above each box indicate the corresponding n.

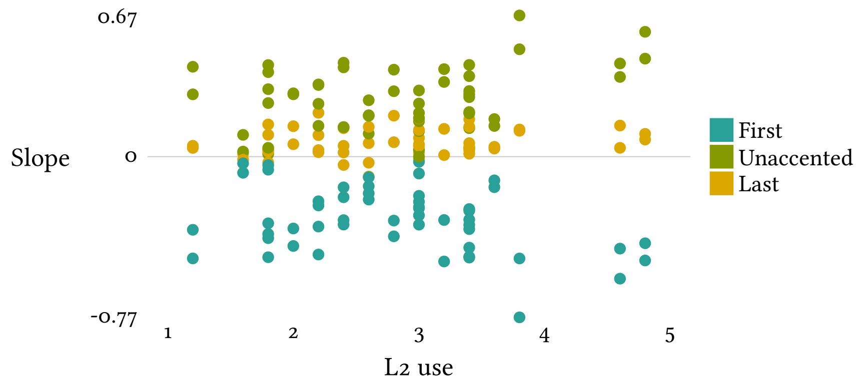


Figure 7.5: Interaction between slopes per category for the first contrast and self-reported language use of s1j2 participants.

L2 proficiency

No effects of L2 proficiency were found for the position of category boundaries, but there was a significant interaction between language use and accent category for category slopes ($F_{(2,132)} = 10.94$, $p < 0.001$) such that participants who reported using their L2 less often tended to have flatter response functions than those that reported relatively high degrees of language use. A significant 3-way interaction between these two factors and the aggregate measure of L2 proficiency ($F_{(2,132)} = 3.8$, $p < 0.05$) was also found.

7.2.2 Second contrast: final-accented *v.* unaccented

Figure 7.6, which shows the results for the second contrast, presents a strikingly different scenario to that shown in the previous section.

Responses for the native speakers show the same response patterns,

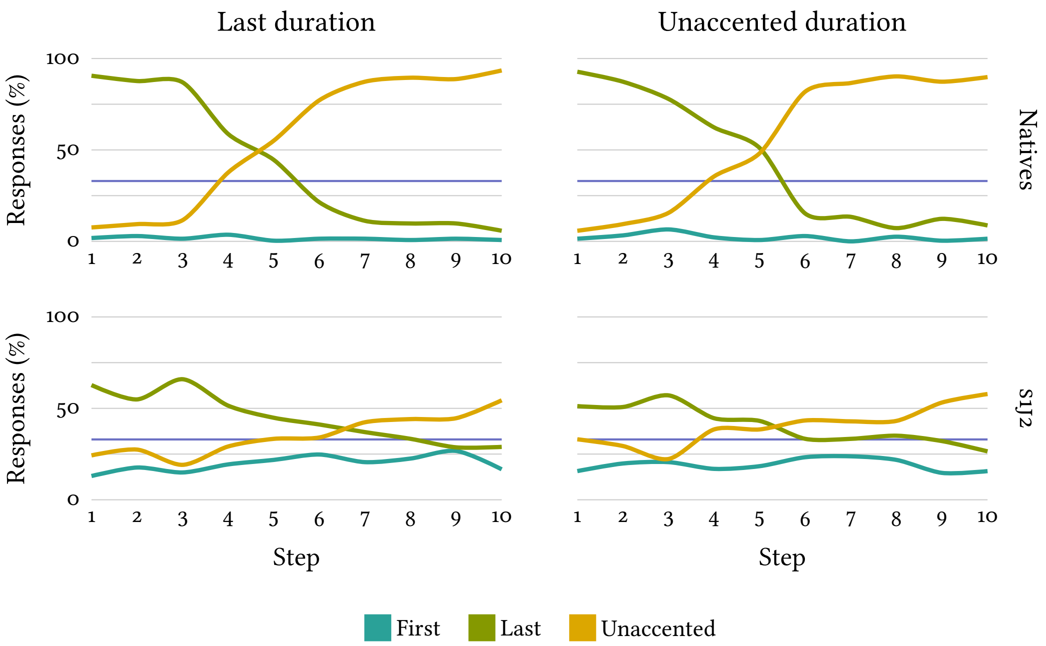


Figure 7.6: Percentage of responses per category at each step in the continua for the first contrast across all participants and speakers. Stimuli with final-accented and unaccented durations are shown in the left and right columns respectively, while rows separate responses between language groups. The violet ● line marks the chance level. Compare with Spanish results in figure 6.5, on page 186.

with if anything fewer responses for the absent category (in this case, that of initial-accented words), strong evidence that in this contrast too the synthesis did not introduce any undesired effects. But responses for the non-native listeners show a dramatic difference, with response functions that are almost entirely flat and hover around the level of chance responses. Interestingly, despite this extremely poor performance the averages of the response functions for the relevant categories do show a tendency in the appropriate direction, such that unaccented responses are marginally higher at the unaccented end of the continua, and the same for final-accented responses. As will be seen below, this difference was significant.

Category boundaries

Results shown in figure 7.6 show that non-native responses for this contrast were almost entirely flat. However, there was a great deal of individual variation, which caused the responses for a large number of participants to cross the 50% threshold. This point should be kept in mind while considering the

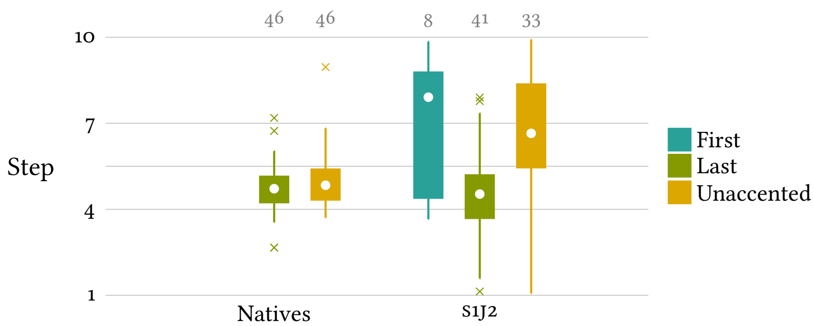


Figure 7.7: Distribution of category boundary positions per group and accent category for the second Japanese contrast. Numbers above each box indicate the corresponding *n*, after discounting participants for which boundary positions could not be estimated.

effects described in this section.

A main effect of language group was found for category boundaries ($F_{(1,164)} = 11.45$, $p < 0.001$), such that boundaries for native listeners ($\bar{x} = 4.84$, $\sigma = 0.91$) occurred earlier than those of S1J2 listeners ($\bar{x} = 5.59$, $\sigma = 2.14$). Likewise, a main effect of category was also found to be significant ($F_{(1,164)} = 16.47$, $p < 0.001$), but largely as a result of the variation of non-native responses. In both cases, the extremely large standard deviation for the non-native responses is explained by the issue described above.

More revealing is a significant 2-way interaction between language group and category ($F_{(1,164)} = 16.76$, $p < 0.001$), shown in figure 7.7, which illustrates the extent of the difference between native and non-native responses. The low number of native initial-accented responses (for which no boundaries were registered) mean that the average boundary positions of the remaining two categories show almost no difference (0.31), a fact that is certainly not the case for non-native listeners. Interestingly, the position of the detected boundaries for non-native final-accented responses aligns closely with that of the native listeners, and they show significantly less variance than the rest of the categories.

Category slopes

A main effect of accent category was also found to be significant for category slopes ($F_{(2,330)} = 36.43$, $p < 0.001$), with overall values for the slope of initial-accented responses being much closer to zero ($\bar{x} = -0.05$,

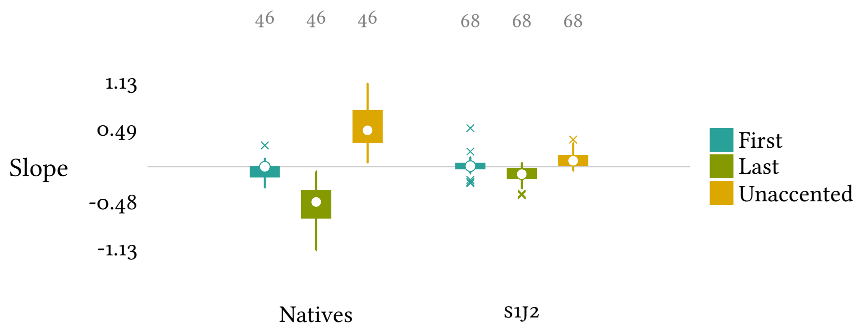


Figure 7.8: Distribution of category slopes per group and accent category for the second Japanese contrast. Numbers above each box indicate the corresponding *n*.

$\sigma = 0.79$) than those of the other two categories ($\bar{x} = -0.27$, $\sigma = 0.27$; $\bar{x} = 0.27$, $\sigma = 0.29$, for final-accented and unaccented respectively). A significant 2-way interaction between language group and accentual category ($F_{(2,330)} = 22.26$, $p < 0.001$), shown in figure 7.8, presents a better picture of these results by breaking up responses for native and non-native listeners. This shows that category slopes for S1J2 are all very close to zero, while those for native speakers present a much wider spread. As stated before, however, it bears pointing out that the slopes of the relevant categories for non-native speakers point in the right direction. A pairwise Bonferroni-corrected t-test between the non-native responses per category was used to check whether these slope values were significantly different from each other, and results showed this to be the case for all of them (even if the difference itself is small).

L2 proficiency

As for L2 proficiency, the only significant result was a high-level 3-way interaction between category, language use use and language proficiency which was found for category slopes ($F_{(2.132)} = 7.5$, $p < 0.001$).

7.3 Discussion

7.3.1 Sensitivity to duration

Results from chapter 4 suggested that the perception of S1J2 participants might have been impaired in part by an excessive sensitivity to random variation in acoustic cues like duration and intensity that were not being

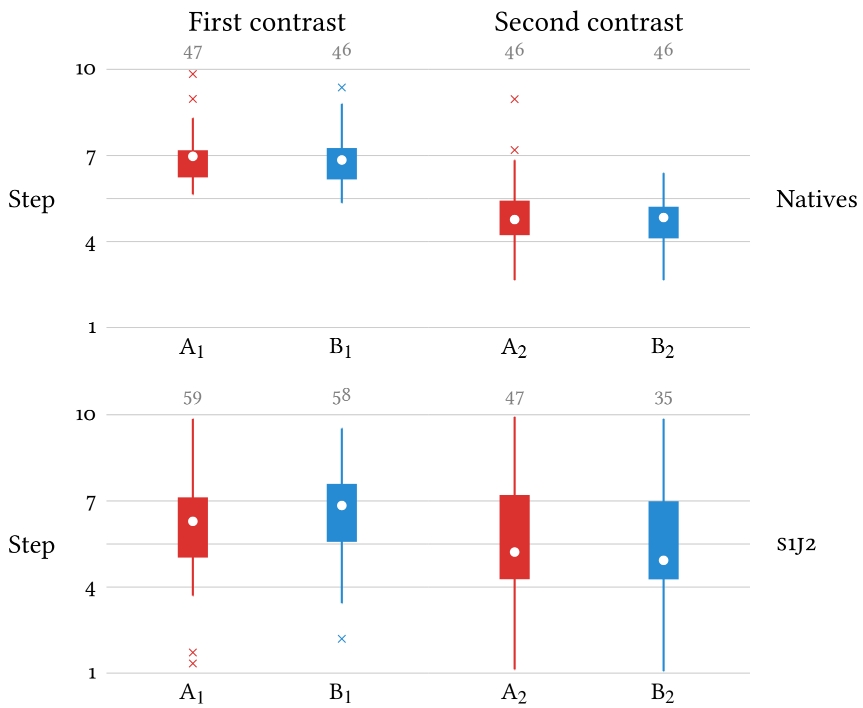


Figure 7.9: Effects of duration on the position of category boundaries per language group (vertically), and contrast (horizontally). Numbers above each box indicate the corresponding *n*, after discounting participants with no estimated boundary positions.

controlled for by native speakers of Japanese but were meaningful in Spanish.

The small differences in the distribution of syllable durations in the original stimuli meant that there was a real possibility that no difference in the responses of s1j2 would show, despite the fact that they are known to have a strong sensitivity to duration as a cue for suprasegmental categories. The results described above showed that this was indeed the case, with no difference in the responses across continua being significant for either slopes or boundaries, either on its own or in combination with other factors.

Marginal effects of duration

However, particularly considering how little difference there was in the stimuli to begin with, results for the relevant interaction between language group and continuum on the position of categorical boundaries were surprisingly close to significance ($F_{(1,201)} = 2.461$, $p = 0.118$). Figure 7.9 shows the data behind this interaction, and shows how similar the responses are for all groups. It also shows that the group that shows the most difference is precisely that of non-native responses for the first contrast, with boundaries in the A continuum ($\bar{x}$= 6.14, σ= 1.66) occurring ever so slightly *earlier* than those in the B continuum ($\bar{x}$= 6.55, σ= 1.55).

To be clear, this is most certainly not evidence of an effect, and it should be noted that the direction of the possible boundary shift goes in the opposite direction that would be expected. And any number of acoustic properties might be involved in such a small effect, a possibility which should be taken particularly seriously considering the stimuli were synthesised using a method (TANDEM-STRAIGHT) that prioritises sound quality over minute feature control (see section 5.2). But both groups of participants were presented with the same stimuli, and responses from native listeners strongly support the notion that, if any artifacts were introduced during manipulation, they were well below the perceptual threshold that is relevant for Japanese at least. And even if other sources of variation managed to go uncontrolled, duration should still be the one that varied most widely, since all other features in the synthesis were fixed to common values while duration was intentionally stretched (even if slightly).

7.3.2 Degree of category development

Results showed a clear difference between the responses for the first and the second contrast, with non-native participants showing a progressively compressed response space between both. However, within this space participants did show evidence of some degree of categorical development (in the first contrast) or at the very least a sensitivity to the relevant acoustic cues (in the second). In the first of these, S1J2 participants showed the characteristic pattern of categorical responses, with response functions that reached a plateau at the ends of the continua. And while the degree of the compression in the last contrast makes it hard to say this with certainty, there is evidence that even here participants show the potential to identify the relevant categories.

This is supported by the slope values measured for the non-native responses. Even in the second contrast, category slopes were statistically different for all category pairs, and described responses that increased as the stimuli approached the appropriate end of the continuum (*i.e.* the slopes had the correct mathematical sign).

Compression of response space

The compression in the response space of non-native participants accurately mirrors the responses reported in section 4.2, shown in more detail in figure 4.3b. Those results showed that the mean correct identification of the position of Japanese prominent syllables by s1j2 across sentence contexts was of approximately 75% for initial-accented words, 50% for final-accented words, and 33% for unaccented words, which are very close to the maximum numbers obtained at the extremes of the continua in this study. The match between the results of these two studies, which used different methodologies on different participant pools, lends additional support to those results.

But it also highlights one possible limitation of the design, in that this study also fundamentally relied in categorical identification, which had already been shown to be problematic for s1j2 participants. More concretely, the methodology's definition of the threshold for categorical boundaries and its dependence on that notion, particularly in light of the large number of s1j2 participants who never crossed the threshold for the second contrast, means that perhaps an altogether different method should be used to explore in more detail the behaviour of the non-native participants.

In particular, the results reported by chapter 4 for the /momo/ contrast between initial-accented and unaccented words (see figure 4.5) suggest that approaching the topic from the perspective of auditory discrimination might yield more information regarding the difficulties faced by s1j2 participants.

Despite those problems, the additional granularity of the approach used in this study (discussed in the previous chapters) revealed new information on the degree of categorical development of s1j2, and provided evidence that at the very least does not disprove the explanations given to the results of chapter 4.

7.3.3 Language proficiency

Like in the study reported in chapter 6, this study showed the effects of L2 proficiency where the study in chapter 4 was not able to do so. Results from this study suggested a correlation in particular between use of the L2 and the degree of development of accentual categories. The implications in this case, like in chapter 6 should serve as additional motivation for a renewed effort in prosodic training, and in the search for training methods that can provide lasting benefits in the perception of suprasegmentals as well.

Part 3

Closing

Chapter 8

General discussion

The present dissertation set out to investigate four main research questions.

1 Do the effects of perceptual interference in L2 learning shown at the segmental level occur as well at the suprasegmental level in the acquisition of L2 lexical stress?

2 Is the weighting of L2 acoustic cues of suprasegmental categories affected by that of L1 cues, as it happens for segmental categories?

3 Does L2 language proficiency have an effect on the perception of L2 lexical prominence in FLA learners?

4 Are the transfer effects affecting cross-linguistic perception of lexical prominence independent of the direction of learning?

Considering the strong effect that prosody has on comprehensibility and foreign accentedness, it is surprising that some of these questions have been as understudied as they are. And although there is evidence that training, age and transfer effects exist for the perception of L2 prosody (see sections 1.1.5 and 1.4.1), the relatively small body of work, and the limited scope of its focus, has meant that it is still not clear whether the existing models of speech perception are entirely applicable to the perception of a considerable portion of what constitutes speech. Recent efforts on this line do exist (*e.g.* So and Best, 2008, 2011), but more research is necessary. This dissertation aims to contribute to this conversation.

8.1 A bidirectional approach

One of the defining characteristics in this dissertation was the explicit effort of examining simultaneously the perception of suprasegmental traits from the perspective of Chilean students of Japanese and Japanese students of Spanish. The main reason to pursue this approach was to examine the specific perceptual difficulties affecting each population with a shared methodology, in a way that made it possible to identify what they had in common. The proposed methodology was a result of trying to answer specifically research question 4, and to explore what this might tell us about language perception in general, and about the perception of suprasegmental categories in particular. The findings relating this specific issue are presented below, in section 8.1.2.

But before moving onto that, and since this bidirectional approach is novel, some words must be said about its costs and benefits.

As the studies in this dissertation progressed, it became increasingly clear that the added complexity in the experimental design would be one of the most serious drawbacks of the approach. A shared methodology, particularly with aspects of two languages that are so fundamentally different at a phonological level, means that the aspects that will be interesting to examine in one case might not be the most appropriate in the other. The methodology that was chosen for the second experiments, reported in part 2 was a way to address this problem, by choosing an approach that would support an analysis at multiple levels and make it possible to focus on the results that were of main interest in each case: the effects of duration as a secondary cue for J1S2 learners, and the lack of appropriate non-native accentual categories for S1J2. But even then there was a cost to pay with regards to the loss of specificity in the design.

However, this common methodology also made it easier to realise that the differences in the perceptual problems affecting both language groups could be explained as fundamentally the manifestation of the same underlying principle: that the perception of suprasegmentals, just as that of segments, is governed primarily by phonetic similarities. This is indeed the notion that is at the core of the NLM model, but even the so-called "categorical" models (SLM and PAM-L2) emphasise the relevance of phonetic similarities in the perception of segments (see section 1.1).

8.1.1 Methodological limitations

It is perhaps surprising that a dissertation that places such a high emphasis on the contrast between two different populations of language learners, spends so little time making direct comparisons between the two. This is a direct results of the problems described above regarding the difficulty in the design of equivalent experiments.

In particular, the initial choice of minimal trios to study Spanish stress placement — a decision that was based on the experiences of a long tradition of similar studies in Spanish (*e.g.* Enríquez, Casado, and Santos, 1989; Llisterri et al., 2003, 2005; Alfano, Savy, and Llisterri, 2008; Schwab and Llisterri, 2011a,b; and specially Kimura et al., 2012) — was specially restrictive in Japanese because of the small number of minimal sets that fit those requirements (see section 2.1.2). This resulted in the forced choice of different sets of materials for both languages, and severely limited the possibility of making direct comparisons between both language groups. Particularly since a large part of the results for Japanese reported in chapter 4 resulted from minimal pairs.

In order to address this limitation, comparisons between language groups were done at a higher level, one step removed from the data itself, and based on the conclusions that could be derived from the results of the different studies. Further studies would be welcome that expand and improve upon this methodology, addressing this issue in a way that allows for the direct comparison between groups. This is a task that, while difficult, is likely to yield interesting and novel results.

A note for future studies

8.1.2 Phonetic similarities are direction-independent

The studies in part 1 provided an overview of the perceptual difficulties affecting the perception of lexical prominence for both language groups, gathering information that is completely novel up to this point. The results from those studies showed that while J1S2 participants were affected primarily by the effects of sentence intonation and not so much by prominence position; the effects shown by S1J2 participants were the exact opposite, with prominence position (and type) being responsible for the largest part of the difference, and the effects of sentence intonation being much smaller.

While the explanation for these results can be made based on what the source and target languages are, they can also be explained by focusing on the aspects that these contexts have in common.

In both languages, the contexts with the highest performances describe the conditions under which prominence is realised in the most similar way: focusing on the action of F_0, the only common cue, prominence is produced with a peak-like F_0 contour that is aligned either at the end of the prominent syllable or during the first half of the one following (see section 1.3.1 and section 1.3.2, and section 2.3.1 for data for stimuli in this dissertation). In these cases, prominence is always realised with an H*+L tone.

The problematic contexts, on the other hand, can be predicted by identifying the circumstances under which each language allows for variation in the realisation of prominence. In the case of Spanish, which has a multiplicity of cues, prominence can be realised in a number of different ways (see Fernández Planas and Martínez Celdrán, 2003; Estebas-Vilaplana and Prieto, 2010; Ortiz Lira, Fuentes, and Astruc, 2010, and the inventory in figure 1.1). But this variation does not depend on the position of the prominence within the word: it is almost entirely governed by the effects of sentence intonation, which is allowed to exert an influence on the intonation of the word because intonation is not needed to mark the position of the Spanish stress (Llisterri et al., 2005; Ortega-Llebaria and Prieto, 2007, among others). In the case of Japanese, on the other hand, the reliance on F_0 alone as the acoustic correlate of prominence has made it robust against changes in sentence intonation, but allows variation depending on the type and position of the prominence.

As detailed in section 1.1, both SLM and PAM-L2 can be said to be "multi-levelled", with speech perception being the result of an interaction between phonetic and phonological similarities. But results from part 1 seem to align better with the precepts of NLM, for which perception is pre-categorical, and therefore much lower-level and centred on the acoustics. The perceptual interference account proposed by Iverson, Kuhl, et al. (2003) and Iverson, Ekanayake, et al. (2008), which shares a close similarity with NLM, might provide a useful interpretation of the results, in particular those that relate to the sensitivity to duration, as will be discussed below.

Non-native sensitivity to secondary acoustic cues

The studies reported in part 2 were designed to simultaneously test the sensitivity to duration as an acoustic cue (based on the results presented in chapters 3 and 4), the effects of the participants' L1 on the weighting of those cues, and the degree of development of accentual categories for both populations (as measured by the steepness of their response curves).

Results from chapter 6 found a very clear effect of duration for J1S2, who not only were able to attend to changes in duration but also responded to them in ways that were extremely similar to those of native speakers. Boundary positions were very similar for both groups, and exactly the same for most conditions. But perhaps the strongest evidence of the degree of similarity between native and non-native responses was the effect caused by the conflicting cues in the synthetic stimuli, which was identical. And although no significant effects of duration were found in chapter 7, a trend shown by non-native learners of Japanese showed that maybe in their case there is also a sensitivity to duration, even though duration is not useful in Japanese (nor particularly significant in their stimuli).

Although similar cases of non-native learners relying on duration in particular for discrimination, even when it was not a native cue, have been previously reported in the literature (Bohn, 1995), these results could be explained by once again relying on the perceptual interference account.

Perceptual interference

The core precept behind the perceptual interference account is that perception is primarily phonetic in nature, and coloured by differences in the scaling of the perceptual space. It shares with NLM the idea that difficulties in the perception by non-native learners is caused by the perceptual space being literally warped, making certain sounds appear to be more or less distant from each other. However, unlike NLM, it makes fewer assumptions regarding what is the cause of this warping. While NLM attributes the warping to the "language magnet" effect of linguistic prototypes, around which phonological categories would develop, and which are the result of the statistical distribution of sounds in the input primarily during childhood, perceptual interference does not make direct claims as to what explains the distortion: it simply attempts to explain the way in which the distortion affects perception.

When examining the perceptual difficulties faced by Japanese learners of English in the acquisition of the /r/-/l/ contrast (Iverson, Kuhl, et al., 2003), the authors found that while the non-native learners showed a relatively high sensitivity to F2, this was a sensitivity that was useless for the discrimination between /r/ and /l/, which are primarily distinguished by native speakers by means of differences in F3. They posited that their difficulty in identification stemmed from the fact that the distortions they had on their perceptual space were not aligned with those that were useful for the relevant contrast, which were evident in both the native speakers and an additional group of German learners of English, who also performed highly.

This provides a good explanation of the results obtained from chapters 6 and 7. If the trend shown by S1J2 had turned out to be significant, it would have been the result of a high sensitivity of the Spanish speakers to duration changes for the perception of accentual contrasts, a sensitivity that, like that of the Japanese learners of English, made them attend to what in this case were the minute differences in the duration (or any other feature) of the stimuli. But perceptual interference can also explain the results for the J1S2, since it has already been shown that sensitivity to duration is easily acquired, but also because Japanese *does* have a sensitivity to duration, although it's exclusively used for the discrimination between short and long vowels.

On the other hand, this does not directly account for the low performance shown by J1S2 with unaccented Japanese words in non-final contexts, which, as was discussed in section 2.3.1, behave like accented words in those contexts. When this happens, they are realised like a word accented in the third syllable, which makes it possible that the low performance shown by the participants is related to the bias that was shown by non-native learners towards the perception of prominence earlier in the word. It is not entirely clear what mechanism is responsible for this, and indeed there is no guarantee that a single process can be held accountable for both of these effects.

This poses a number of interesting extensions to these studies, exploiting the fact that both populations have a perceptual sensitivity to duration, and that both make use of duration for contrasts in production (although in both cases with different ends). The interaction between the segmental and suprasegmental levels in this regard is also intriguing, since S1J2 learners often report having particular trouble perceiving the Japanese vowel length contrast even though they are clearly sensitive to duration. This might be the result of the segmental duration cue being interpreted as a secondary suprasegmental cue for prominence, where it would be masked by differences in pitch.

Segmental and suprasegmental categories

Although research on the perception of non-native contrasts has been extensive, it has been well established in this dissertation and elsewhere that the field of suprasegmental perception has been left wanting (see section 1.1.5). In particular, this has implications regarding the applicability of the perception models discussed in section 1.1.4, which up to this point has been a contested topic.

So and Best (2008, 2011) are among the few to have made a direct attempt at answering the question of whether the knowledge acquired from these models applies to the perception of suprasegmental categories, and if so how well. But even though their results are promising, by calling their categories *iCategories* they use a terminology that unfortunately implies the very thing against which their results seem to lead.

The studies presented in this dissertation did not set out to directly test the predictions made by the leading perception models discussed above. But they do present evidence in support of the idea that segmental and suprasegmental categories, insofar as they are linguistic categories, seem to behave similarly and be affected by similar phenomena in similar ways.

Further studies in these areas would be welcome.

8.2 Effects of language proficiency

Studies on both parts 1 and 2 examined the effects of language proficiency in the perception of lexical prominence, and while the first set of studies was not able to find any correlations, studies in part 2 seemed to have been much more sensitive, and found some significant correlations. Results on chapters 6 and 7 reported significant effects of the proficiency measures, and in particular of language use, such that participants who reported using their L2 more often behaved more like native speakers, having earlier category boundaries (JIS2 learners) and steeper category slopes (both groups).

These results are in line with other reports of improvement in the ability of L2 learners to perceive suprasegmental contrasts after training (*e.g.* Schwab and Llisterri, 2011a,b, 2012, 2014, for French learners of Spanish; and Alfano, Llisterri, and Savy, 2007; Alfano, Savy, and Llisterri, 2008; Alfano, Schwab, et al., 2010, for Italian learners of Spanish; as well as additional coverage in the introduction).

The results reported in this dissertation are particularly encouraging, because unlike most previous studies, participants considered here were FLA students who had never stayed for a considerable amount of time in an environment in which the L2 was dominant.

But one should be cautious, since it is still possible that the correlation that existed between L2 use and performance, while significant, might be due to other factors. This seems to have been the case in Saalfeld (2012), which reported the results of a training experiment in which English speakers were trained in the perception of Spanish lexical contrasts and tested in an indirect word recognition task, in which a target word could be interpreted in one of two different ways depending on the position of the prominence changing the meaning of the framing sentence, and participants were asked to identify the meaning of the sentence. In that study, although participants from the experimental group showed an improvement, an even larger improvement was shown by the control group, who received no special prosodic training whatsoever. As possible causes, they mention differences in participant motivation (the control group claimed to be more highly motivated) and the lack of distractors in the test (which meant that the test itself might have served as training), highlighting the need for further detailed studies controlling as much as possible the large number of confounding variables involved in these tasks.

Teacher's attitudes towards prosodic training

Just like the participants in Saalfeld (2012) might have been influenced by attitudinal differences in their approach to the learning process, language teachers are also affected by their own attitudes towards the teaching of pronunciation. Indeed, part of the motivation for examining the effects of language training was the realisation, through personal communication with language teachers active in Chile, Japan and the United States, that a large majority of them did not, in fact, spend any considerable amount of time specifically teaching students how to provide prominence to a given syllable, or how to best perceive where prominence had been assigned by other speakers. Similarly, conversations in particular with learners of Japanese showed that a large number of them were never made aware that the language had an accentual system, much less how to correctly interpret the acoustic cues that encoded it.

The reasons provided by language instructors to prefer not to teach intonation could be broadly grouped into two categories:

1 That they preferred to employ their limited resources in the training of other aspects of the language which were more important to ensure the students'

ability to express themselves and be understood in their L2.

2 That there was a general lack of training methods and tools to effectively improve the perception of these prosodic features among their students.

And these same attitudes regarding the difficulty and questionable relevance of prosodic training are widespread in the teaching community, at the very least among teachers of both English (*e.g.* Burgess and Spencer, 2000; Macdonald, 2002; Breitkreutz, Derwing, and Rossiter, 2009; Derwing and Munro, 2005, among others) and Spanish (*e.g.* Usó Viciedo, 2009; Orta Gracia, 2009, 2010, among others) as L2.

There are a number of reasons for this to be the case, including the reticence shown by teachers to change their established practices (Orta Gracia, 2010), and the lack of knowledge and confidence in relevant teaching methods (Macdonald, 2002). In particular the latter of these is strongly connected to the lack of impact of pronunciation research mentioned before, and to the gap that it generates between the knowledge reported in the literature and its application in the real world. Specifically, the first issue has been dealt with in detail in chapter 1, which presents some of the extensive evidence in support of the effect that that pronunciation — and in particular suprasegmental pronunciation — has on foreign accentedness, intelligibility, and comprehensibility.

As for the second issue, part of it stems from the very same gap that was mentioned above between the research and the practice of L2 instruction. This is particularly the case with the survival of outdated teaching methods of limited effectiveness and the lack of knowledge or instruction on more research-driven approaches (see Orta Gracia, 2010; Gilbert, 2014, for a review), and with the unsuitability of a large proportion of current teacher training programs, which do not properly prepare them to make use of existing methods and tools (see Murphy, 2014, for a review). And although teachers complain about the lack of tools, recently there has been considerable interest in the development of new technologies that can be used for language teaching in general (see Eskenazi, 2009, for a review). This is particularly the case with Japanese, with the recent development of

Development of teaching tools

tools aimed specifically at the improvement of intonation providing immediate feedback on its production, like the Online Japanese Accent Dictionary (OJAD, Minematsu, Suzuki, et al., 2012; Nakamura et al., 2013; Hirano et al., 2013; and see also Hirose, Gendrin, and Minematsu, 2003).

And although the efforts are considerably less well-concerted, this is true for Spanish as well, with several researchers reporting on the use of speech manipulation strategies to train students in the production of specific language features. This is the case with Lahoz, who used a Praat-based tool to automatically convert students' productions to so-called *sasasa* speech in order to train the perception of rhythm and stress placement (Lahoz, 2011, 2012).

A different kind of proposal is that made by Corominas Pérez (2014), who also focuses on the use of Praat in the classroom, but instead of automatic manipulation tools describes strategies with which the program can be used to illustrate specific mispronunciations or deviations in the students' production in a much more hands-on (and therefore resource-intensive). Although they offer guidelines and some instruction on the use of the program, the lack of ready-made tools will limit the applicability of these methods. But despite these limitations, there are numerous reports of language instructors using Praat for similar approaches (*e.g.* Gorjian, Hayati, and Pourkhoni, 2013, for Iranian students of English; Wilson, 2008, for Japanese students of English).

Results of this dissertation contribute to the discussion by providing clear evidence that, at least for isolated words, non-native learners studying under conditions that are far less than ideal — without any targeted training or an L2-predominant environment that can increase their exposure — show a link between their proficiency and their ability to perceive non-native suprasegmental contrasts. What remains is to examine in more detail the magnitude of that benefit, and the ways in which training can be improved so that its effects are longer lasting and have the broadest scope possible.

Bibliography

Alfano, Iolanda, Joaquim Llisterri, and Renata Savy (2007). "The perception of Italian and Spanish lexical stress: A first cross-linguistic study". In: *Proceedings of the 16th International Congress of Phonetic Sciences*, pp. 6–10. URL: http://liceu.uab.es/joaquim/publicacions/Alfano_Llisterri_Savy_07_Perception_Stress_Spanish_Italian.pdf (visited on 02/20/2013) (cit. on pp. 56, 58, 216).

Alfano, Iolanda, Renata Savy, and Joaquim Llisterri (2008). "Las características acústicas y perceptivas del acento léxico en español y en italiano: Los patrones acentuales paroxítonos [The acoustic and perceptive characteristics of Spanish and Italian lexical stress: paroxytone stress patterns]". In: *Language Design. Journal of Theoretical and Experimental Linguistics* (Special Issue 2: Experimental Prosody), pp. 23–30 (cit. on pp. 56, 58, 211, 216).

Alfano, Iolanda, Sandra Schwab, Renata Savy, and Joaquim Llisterri (2010). "Cross-language speech perception: Lexical stress in Spanish with Italian and Francophone subjects". In: *AISV 2009. La dimensione temporale del parlato. Atti del 5° Convegno Nazionale AISV - Associazione Italiana di Scienze della Voce*. Ed. by S. Schmid, M. Schwarzenbach, and D. Studer. Torriana: EDK Editore (cit. on pp. 56, 58, 64, 216).

Anderson-Hsieh, Janet, Ruth Johnson, and Kenneth Koehler (1992). "The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure". In: *Language Learning* 42.4, pp. 529–555. ISSN: 1467-9922. DOI: 10.1111/j.1467-1770.1992.tb01043.x. URL: http://onlinelibrary.wiley.

com/doi/10.1111/j.1467-1770.1992.tb01043.x/abstract (visited on 10/17/2015) (cit. on p. 30).

Aske, Jon (1990). "Disembodied rules versus patterns in the lexicon: Testing the psychological reality of Spanish stress rules". In: *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*. Ed. by Kira Hall, Jean-Pierre Koening, Michael Meachman, Sondra Reinman, and Laurel A. Sutton. Vol. 16. Ann Arbor, MI: Braun-Brumfield. URL: http://elanguage.net/journals/bls/article/download/2628/2620 (visited on 02/20/2013) (cit. on p. 41).

Bachman, Lyle F. (1990). "Constructing measures and measuring constructs". In: *The Development of Second Language Proficiency*. Cambridge Applied Linguistics. Cambridge University Press. ISBN: 9781139524568. URL: http://dx.doi.org/10.1017/CBO9781139524568.004 (cit. on p. 60).

Bansal, Ram Krishna (1966). "The intelligibility of Indian English: measurements of the intelligibility of connected speech, and sentence and word material, presented to listeners of different nationalities". PhD thesis. University College London. URL: http://eric.ed.gov/?id=ED177849 (visited on 10/17/2015) (cit. on p. 30).

Bates, Douglas, Martin Maechler, and Ben Bolker (2012). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999999-0. URL: http://cran.r-project.org/package=lme4 (cit. on pp. 114, 137).

Beckman, Mary E. and Janet B. Pierrehumbert (1986). "Intonational structure in Japanese and English". In: *Phonology* 3.1, pp. 255–309. URL: http://journals.cambridge.org/abstract_S095267570000066X (visited on 10/21/2015) (cit. on p. 38).

Beckman, Mary, Manuel Díaz-Campos, Julia Tevis McGory, and Terrell A. Morgan (2002). "Intonation across Spanish, in the Tones and Break Indices framework". In: *Probus* 14.1, pp. 9–36. URL: http://www.indiana.edu/~spanport/people/diaz-campos%20papersinpdf/SpanishIntonation.pdf (visited on 10/21/2015) (cit. on p. 38).

Bello, Andrés (1847). *Gramática de la lengua castellana destinada al uso de los americanos* [*Grammar of the Spanish language for the use of Americans*]. Santiago: Imprenta del Progreso. (Visited on 03/09/2016) (cit. on p. 43).

Benrabah, M. (1987). "The intelligibility of Algerian speakers of English: a phonetic/phonological study". PhD thesis. University College London. URL: http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.565111 (visited on 10/17/2015) (cit. on p. 30).

Benrabah, Mohamed (1997). "Word-stress — A source of unintelligibility in English". In: *IRAL, International Review of Applied Linguistics in Language Teaching* 35.3, p. 157. ISSN: 0019042X. URL: http://search.proquest.com.libproxy.ucl.ac.uk/docview/204937439/citation/3C72CAD982AE4130PQ/1?accountid=14511 (visited on 06/27/2015) (cit. on p. 30).

Best, Catherine T. (1995). "A direct realist view of cross-language speech perception". In: *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*. Ed. by Winifred Strange. Timonium, MD: York Press, pp. 171–204 (cit. on pp. 31, 132).

Best, Catherine T., Gerald W. McRoberts, and Elizabeth Goodell (2001). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system". In: *Journal of the Acoustical Society of America* 109, pp. 775–794 (cit. on p. 29).

Best, Catherine T. and Michael D. Tyler (2007). "Nonnative and second-language speech perception: Commonalities and complementarities". In: *Second language speech learning: The role of language experience in speech perception and production*. Ed. by Murray J. Munro and Ocke-Schwen Bohn. John Benjamins, pp. 13–34 (cit. on pp. 32, 60, 132).

Boersma, Paul and David Weenink (2013). *Praat: doing phonetics by computer*. URL: http://www.praat.org/ (cit. on pp. 82, 88, 93, 168, 219).

Bohn, Ocke-Schwen (1995). "Cross-language speech perception in adults: First language transfer doesn't tell it all". In: *Speech perception and linguistic experience: Theoretical and methodological issues in cross-

language speech research. Ed. by Winifred Strange. Timonium, MD: York Press, pp. 275–300 (cit. on pp. 162, 213).

Breitkreutz, Judy, Tracey M. Derwing, and Marian J. Rossiter (2009). "Pronunciation teaching practices in Canada". In: *TESL Canada Journal* 19.1, pp. 51–61. URL: http://teslcanadajournal.ca/index.php/tesl/article/view/919 (visited on 03/15/2016) (cit. on p. 218).

Broselow, Ellen, Richard R. Hurtig, and Catherine Ringen (1987). "The perception of second language prosody". In: *Interlanguage phonology: The acquisition of a second language sound system*, pp. 350–361 (cit. on p. 29).

Burgess, John and Sheila Spencer (2000). "Phonology and pronunciation in integrated language teaching and teacher education". In: *System* 28.2, pp. 191–215. URL: http://www.sciencedirect.com/science/article/pii/S0346251X00000075 (visited on 03/15/2016) (cit. on p. 218).

Cabezas, José María (2009). "Las creencias de los estudiantes japoneses sobre la pronunciación española: un análisis exploratorio [The beliefs of Japanese students on Spanish pronunciation: an exploratory analysis]". MSc. Sevilla: Universidad Pablo de Olavide. 35 pp. (cit. on p. 59).

Carranza, Mario (2012). "Errores y dificultades específicas en la adquisición de la pronunciación del español LE por hablantes de japonés y propuestas de corrección [Errors and difficulties specific to the acquisition of L2 Spanish by Japanese speakers and a proposal for their correction]". In: *Nuevos enfoques en la enseñanza del español en Japón*. Ed. by Concha Moreno and Grupo de Investigación de la Didáctica del Español. Tōkyō: Asahi, pp. 51–78 (cit. on p. 59).

Carranza, Mario, C. Cucchiarini, Joaquim Llisterri, María Jesús Machuca, and Antonio Ríos (2014). "A corpus-based study of Spanish L2 mispronunciations by Japanese speakers". In: *Edulearn14 Proceedings*. 6th International Conference on Education and New Learning Technologies. Barcelona: IATED Academy, pp. 3696–3705 (cit. on p. 59).

Chikamatsu, Nobuko (1996). "The effects of L1 orthography on L2 word recognition: a study of American and Chinese learners of Japanese".

In: *Studies in Second Language Acquisition* 18.4, pp. 403–432. ISSN: 1470-1545. DOI: 10.1017/S0272263100015369. URL: http://journals.cambridge.org/article_S0272263100015369 (visited on 04/22/2015) (cit. on p. 63).

Chun, Dorothy M (2002). “Teaching stress and rhythm”. In: *Discourse intonation in L2: From theory and research to practice*. 1st ed. Vol. 1. Philadelphia, PA: John Benjamins, pp. 147–197. ISBN: 1-58811-168-7 (cit. on pp. 30, 162).

Corominas Pérez, Sandra (2014). “¿Estás en la onda? Aplicación de Praat a la enseñanza de la entonación [Are you on the wave? Applying Praat to the teaching of intonation]”. Master. Girona: Universitat de Girona. URL: http://dugi-doc.udg.edu//handle/10256/10072 (visited on 03/14/2016) (cit. on p. 219).

Cuervo, Rufino José (1907). *Apuntaciones críticas sobre el lenguaje bogotano, con frecuente referencia al de los países de Hispano-América* [*Critical remarks on the language of Bogotá, with frequent reference to that of Latin American countries*]. 5th ed. Paris: A. & R. Roger y F. Chernoviz. 692 pp. (cit. on p. 42).

Cutler, Anne and Hsuan-Chih Chen (1997). “Lexical tone in Cantonese spoken-word processing”. In: *Perception & Psychophysics* 59.2, pp. 165–179. URL: http://link.springer.com/article/10.3758/BF03211886 (visited on 11/09/2015) (cit. on p. 37).

Cutler, Anne, Delphine Dahan, and Wilma van Donselaar (1997). “Prosody in the comprehension of spoken language: a literature review”. In: *Language and Speech* 40.2, pp. 141–201. URL: http://pubman.mpdl.mpg.de/pubman/item/escidoc:68854:6/component/escidoc:76882/Cutler_1997_Prosody%20in%20the.pdf (cit. on p. 37).

Cutler, Anne and Takashi Otake (1994). “The processing of word prosody in Japanese”. In: *Proceedings of the Sixth Australian International Conference on Speech Science and Technology*. Adelaide, pp. 599–604 (cit. on pp. 51, 55, 162).

De Looze, Céline and Daniel J. Hirst (2008). "Detecting changes in key and range for the automatic modelling and coding of intonation". In: *Speech Prosody 2008*. Ed. by Plínio A. Barbosa, Sandra Madureira, and Cesar Reis, pp. 135–138. URL: http://www.isca-speech.org/archive/sp2008/sp08_135.html (visited on 07/23/2015) (cit. on p. 82).

Delgado, Pedro, Gabriela Guerrero, Judith P. Goggin, and Barbara B. Ellis (1999). "Self-assessment of linguistic skills by bilingual Hispanics". In: *Hispanic Journal of Behavioral Sciences* 21.1, pp. 31–46. URL: http://hjb.sagepub.com/content/21/1/31.short (visited on 05/05/2015) (cit. on p. 62).

Derwing, Tracey M. and Murray J. Munro (2005). "Second language accent and pronunciation teaching: a research-based approach". In: *Tesol Quarterly* 39.3, pp. 379–397. ISSN: 0039-8322. DOI: 10.2307/3588486. URL: http://www.jstor.org/stable/3588486 (visited on 10/12/2015) (cit. on pp. 29, 218).

Dupoux, Emmanuel, Christophe Pallier, Nuria Sebastian, and Jacques Mehler (1997). "A destressing "deafness" in French?" In: *Journal of Memory and Language* 36, pp. 406–421. URL: http://cogprints.org/749 (visited on 10/21/2015) (cit. on pp. 36, 56).

Dupoux, Emmanuel, Núria Sebastián-Gallés, Eduardo Navarrete, and Sharon Peperkamp (2008). "Persistent stress 'deafness': The case of French learners of Spanish". In: *Cognition* 106.2, pp. 682–706. ISSN: 0010-0277. DOI: 10.1016/j.cognition.2007.04.001. URL: http://www.sciencedirect.com/science/article/pii/S0010027707001102 (visited on 02/20/2013) (cit. on pp. 36 sq., 56 sq.).

Eimas, Peter D., Einar R. Siqueland, Peter Jusczyk, and James Vigorito (1971). "Speech perception in infants". In: *Science* 171.3968, pp. 303–306. URL: http://www.sciencemag.org/content/171/3968/303.short (visited on 10/18/2015) (cit. on p. 34).

Enríquez, Emilia V., Celia Casado, and Andrés Santos (1989). "La percepción del acento en español [Stress perception in Spanish]". In: *LEA: Lingüística española actual* 11.2, pp. 241–270. URL: http://dialnet.

unirioja.es/servlet/articulo?codigo=55491 (visited on 03/01/2016) (cit. on pp. 42 sq., 45, 211).

Eskenazi, Maxine (2009). “An overview of spoken language technology for education”. In: *Speech Communication.* Spoken Language Technology for Education: Spoken Language 51.10, pp. 832–844. ISSN: 0167-6393. DOI: 10.1016/j.specom.2009.04.005. URL: http://www.sciencedirect.com/science/article/pii/S0167639309000673 (visited on 07/28/2015) (cit. on p. 218).

Estebas Vilaplana, Eva and Pilar Prieto Vives (2008). “La notación prosódica del español: una revisión del SP-TOBI [Prosodic annotation in Spanish: a revision of SP-TOBI]”. In: *Estudios de fonética experimental* 17, pp. 264–283. URL: http://www.raco.cat/index.php/EFE/article/viewArticle/140072/0 (visited on 10/21/2015) (cit. on p. 38).

Estebas-Vilaplana, Eva and Pilar Prieto (2010). “Castillian Spanish intonation”. In: *Transcription of Intonation of the Spanish Language.* Ed. by Pilar Prieto and Paolo Roseano. München: Lincom Europa, pp. 17–48. ISBN: 9783862901845 (cit. on pp. 48, 212).

Face, Timothy L. (2006). “Cognitive factors in the perception of Spanish stress placement: implications for a model of speech perception”. In: *Linguistics* 44.6, pp. 1237–1267. URL: http://www.degruyter.com/view/j/ling.2006.44.issue-6/ling.2006.040/ling.2006.040.xml (visited on 02/20/2013) (cit. on pp. 40 sqq., 46, 156).

Fernández Planas, Ana María and Eugenio Martínez Celdrán (2003). “El tono fundamental y la duración: dos aspectos de la taxonomía prosódica en dos modalidades de habla (enunciativa e interrogativa) del español [Fundamental frequency and duration: two aspects of prosodic taxonomy in two modes of speech (enunciative and interrogative)]”. In: *Estudios de Fonética Experimental* 12, pp. 165–200. URL: http://stel.ub.edu/labfon/sites/default/files/XII-10.pdf (cit. on pp. 46, 84, 212).

Flege, James E. (1995). “Second language speech learning: Theory, findings, and problems”. In: *Speech Perception and Linguistic Experience: Issues* *in Cross-Language Research*. Timonium, MD: York Press, pp. 233–277 (cit. on pp. 33, 60).

Flege, James E. (2003). "Assessing constraints on second-language segmental production and perception". In: *Phonetics and phonology in language comprehension and production, differences and similarities*. Ed. by A. Meyer and N. Schiller. Berlin: Mouton de Gruyter, pp. 319–355 (cit. on p. 29).

— (2007). "Language contact in bilingualism: phonetic system interactions". In: *Laboratory Phonology 9*. Ed. by J. Cole and José Hualde. Mouton de Gruyter, pp. 353–380 (cit. on p. 33).

Flege, James E. and Ian R.A. Mackay (2011). "What accounts for "age" effects on overall degree of foreign accent?" In: *Achievements and perspectives in SLA of Speech: new Sounds 2010*. Vol. 2. Poznań, Poland: Peter Lang, pp. 65–82. ISBN: 978-3-631-60723-7 (cit. on p. 33).

Flege, James E., Ian R.A. Mackay, and Diane Meador (1999). "Native Italian speakers' perception and production of English vowels". In: *The Journal of the Acoustical Society of America* 106.5, pp. 2973–87. ISSN: 0001-4966. URL: http://www.ncbi.nlm.nih.gov/pubmed/10573909 (cit. on p. 33).

Fort, Karën, Gilles Adda, and K. Bretonnel Cohen (2011). "Amazon mechanical turk: Gold mine or coal mine?" In: *Computational Linguistics* 37.2, pp. 413–420. URL: http://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00057 (visited on 03/09/2016) (cit. on p. 107).

Fowler, Carol A. (1986). "An event approach to the study of speech perception from a direct-realist perspective". In: *Journal of Phonetics* 14.1, pp. 3–28. URL: http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA173677#page=148 (visited on 10/17/2015) (cit. on p. 31).

Fox, John and Sanford Weisberg (2011). *An R companion to applied regression*. Second. Thousand Oaks CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion (cit. on p. 116).

Fujisaki, Hiroya and Keikichi Hirose (1984). "Analysis of voice fundamental frequency contours for declarative sentences of Japanese". In: *Jour-*

nal of the Acoustical Society of Japan 5.4, pp. 233–242. URL: http://jlc.jst.go.jp/DN/JALC/00150256333?from=Google (visited on 10/21/2015) (cit. on p. 38).

Fulop, S.A., Ethelbert Kari, and Peter Ladefoged (1998). "An acoustic study of the tongue root contrast in Degema vowels". In: *Phonetica* 55.1, pp. 80–98. ISSN: 1423-0321, 0031-8388. DOI: 10.1159/000028425. URL: http://www.karger.com/Article/FullText/28425 (visited on 02/08/2013) (cit. on p. 44).

Gandour, Jack, Donald Wong, Li Hsieh, Bret Weinzapfel, Diana Van Lancker, and Gary D. Hutchins (2000). "A crosslinguistic PET study of tone perception". In: *Journal of Cognitive Neuroscience* 12.1, pp. 207–222 (cit. on p. 37).

Garcia, Miguel (2011). "The Intonational Patterns of the Peruvian Amazonian Spanish (PAS)". PhD thesis. State University of New York at Stony Brook. URL: http://gradworks.umi.com/15/01/1501331.html (visited on 09/01/2015) (cit. on p. 48).

García, Miguel (2014). "Peruvian Amazonian Spanish: Status of tritonal pitch accents". In: 1st Symposium on Intonation and Tone in Spanish. Amherst, MA. URL: http://intospan2014.weebly.com/slides.html (visited on 04/23/2015) (cit. on p. 48).

Garrido, Juan M., Joaquim Llisterri, Carme de la Mota, and Antonio Ríos (1993). "Prosodic differences in reading style: isolated vs. contextualized sentences". In: *Third European Conference on Speech Communication and Technology* (cit. on pp. 45, 49).

Gibson, Eleanor J (1991). *An odyssey in learning and perception*. MIT Press. ISBN: 9780262571036 (cit. on p. 31).

Gilbert, Judy (2014). "Intonation is hard to teach". In: *Pronunciation myths: applying second language research to classroom teaching*. Ed. by Linda Grant. Ann Arbor, MI: University of Michigan Press, pp. 107–136. ISBN: 978-0-472-03516-8 (cit. on p. 218).

Goldsmith, John A. (1976). *Autosegmental phonology*. Citeseer. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.170.2501&rep=rep1&type=pdf (visited on 10/21/2015) (cit. on p. 38).

Gorjian, Bahman, Abdolmajid Hayati, and Parisa Pourkhoni (2013). "Using Praat software in teaching prosodic features to EFL learners". In: *Procedia-Social and Behavioral Sciences* 84, pp. 34–40. URL: http://www.sciencedirect.com/science/article/pii/S1877042813015723 (visited on 03/14/2016) (cit. on p. 219).

Grant, Linda (2014). "Pronunciation myths: applying second language research to classroom teaching". In: ed. by Linda Grant. Ann Arbor, MI: University of Michigan Press, pp. 1–33. ISBN: 978-0-472-03516-8 (cit. on p. 29).

Haraguchi, Shōsuke (1977). *The tone pattern of Japanese: An autosegmental theory of tonology*. Kaitakusha (cit. on p. 52).

Hidalgo Navarro, Antonio and Mercedes Quilis Merín (2004). *Fonética y fonología españolas* [*Spanish phonetics and phonology*]. Valencia: Tirant lo blanch (cit. on p. 41).

Hirano, Hiroko, Ibuki Nakamura, Nobuaki Minematsu, Masayuki Suzuki, Chieko Nakagawa, Noriko Nakamura, Yukinori Tagawa, Keikichi Hirose, and Hiroya Hashimoto (2013). "A free online accent and intonation dictionary for teachers and learners of Japanese". In: *INTERSPEECH 2013*. 14th Annual Conference of the International Speech Communication Association. Ed. by F. Bimbot, C. Cerisara, C. Fougeron, G. Gravier, L. Lamel, F. Pellegrino, and P. Perrier. Lyon, pp. 1875–1876. URL: http://isca-speech.org/archive/interspeech_2013/i13_1875.html (visited on 07/28/2015) (cit. on p. 219).

Hirose, Keikichi, Frédéric Gendrin, and Nobuaki Minematsu (2003). "A Pronunciation Training System for Japanese Lexical Accents with Corrective Feedback in Learner's Voice". In: *8th European Conference on Speech Communication and Technology*. EUROSPEECH 2003 - INTERSPEECH 2003. Ed. by Jean-Cédric Chappelier. Geneva, pp. 141–144. URL:

http://isca-speech.org/archive/eurospeech_2003/e03_0141.html (cit. on p. 219).

Hirst, Daniel J. (2007). "A Praat plugin for Momel and INTSINT with improved algorithms for modelling and coding intonation". In: *Proceedings of the XVIth International Congress of Phonetic Sciences*. Ed. by Jurgen Trouvain and William J. Barry. Saarbrücken, pp. 1233–1236. URL: http://www.icphs2007.de/conference/Papers/1443/1443.pdf (visited on 05/12/2015) (cit. on p. 82).

— (2012). "The analysis by synthesis of speech melody: From data to models". In: *Journal of Speech Sciences* 1.1, pp. 55–83. ISSN: 2236-9740. URL: http://www.journalofspeechsciences.org/index.php/journalofspeechsciences/article/view/21 (visited on 03/05/2014) (cit. on pp. 82 sq.).

Hyman, Larry M. (2007). "Universals of tone rules: 30 years later". In: *Typological Studies in Word and Sentence Prosody*. Ed. by Tomas Riad and Carlos Gussenhoven. Walter de Gruyter, pp. 1–34. ISBN: 9783110207569 (cit. on p. 52).

Iverson, Paul, Dulika Ekanayake, Silke Hamann, Anke Sennema, and Bronwen G. Evans (2008). "Category and perceptual interference in second-language phoneme learning: an examination of English /w/-/v/ learning by Sinhala, German, and Dutch speakers". In: *Journal of experimental psychology. Human perception and performance* 34.5, pp. 1305–16. ISSN: 0096-1523. DOI: 10.1037/0096-1523.34.5.1305. URL: http://www.ncbi.nlm.nih.gov/pubmed/18823213 (cit. on p. 213).

Iverson, Paul, Valerie Hazan, and Kerry Bannister (2005). "Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults". In: *The Journal of the Acoustical Society of America* 118.5, p. 3267. ISSN: 00014966. DOI: 10.1121/1.2062307. URL: http://link.aip.org/link/JASMAN/v118/i5/p3267/s1&Agg=doi (cit. on p. 60).

Iverson, Paul, Patricia K. Kuhl, Reiko Akahane-Yamada, and Eugen Diesch (2003). "A perceptual interference account of acquisition difficulties for non-native phonemes”. In: *Cognition* 87, pp. 47–57 (cit. on pp. 163, 213 sq.).

Iverson, Paul, Melanie Pinet, and Bronwen G. Evans (2011). “Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels”. In: *Applied Psycholinguistics*, pp. 1–16. ISSN: 0142-7164. DOI: 10.1017/S0142716411000300. URL: http://www.journals.cambridge.org/abstract_S0142716411000300 (cit. on p. 60).

Katada, Fusa (1990). “On the representation of moras: Evidence from a language game”. In: *Linguistic Inquiry* 21.4, pp. 641–646. URL: http://www.jstor.org/stable/4178699 (cit. on p. 50).

Kawahara, Hideki, Ikuyo Masuda-Katsuse, and Alain de Cheveigné (1999). “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F_0 extraction: possible role of a repetitive structure in sounds”. In: *Speech Communication* 27.3, pp. 187–207. ISSN: 0167-6393. DOI: 10.1016/S0167-6393(98)00085-5. URL: http://www.sciencedirect.com/science/article/pii/S0167639398000855 (visited on 10/14/2015) (cit. on p. 166).

Kawahara, Hideki and Masanori Morise (2011). “Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework”. In: *Sādhanā* 36.5, pp. 713–727. ISSN: 0256-2499, 0973-7677. DOI: 10.1007/s12046-011-0043-3. URL: http://link.springer.com/article/10.1007/s12046-011-0043-3 (visited on 10/13/2015) (cit. on p. 166).

Kawahara, Hideki, Masanori Morise, T. Takahashi, R. Nisimura, T. Irino, and H. Banno (2008). “TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F_0, and aperiodicity estimation”. In: *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*, pp. 3933–3936. DOI: 10.1109/ICASSP.2008.4518514 (cit. on p. 166).

Kawahara, Hideki, R. Nisimura, T. Irino, Masanori Morise, T. Takahashi, and H. Banno (2009). "Temporally variable multi-aspect auditory morphing enabling extrapolation without objective and perceptual breakdown". In: *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009*. IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009, pp. 3905–3908. DOI: 10.1109/ICASSP.2009.4960481 (cit. on p. 166).

Kim, Ji-Young (2015). "Perception and production of Spanish lexical stress by Spanish heritage speakers and English L2 learners of Spanish". In: ed. by Erik W. Willis, Pedro Martín Butragueño, and Esther Herrera Zendejas, pp. 106–128. URL: http://www.lingref.com/cpp/larp/6/paper3196.pdf (visited on 07/15/2015) (cit. on pp. 45, 162).

Kimura, Takuya, Hirotaka Sensui, Miyuki Takasawa, Atsuko Toyomaru, and José Joaquín Atria (2012). "Influencia de la entonación oracional sobre la percepción del acento español por estudiantes japoneses [Influence of sentence intonation on the perception of Spanish stress by Japanese learners]". In: *Estudios de fonética experimental* 21, pp. 11–42 (cit. on pp. 59, 61, 64–69, 73 sqq., 78, 86, 99, 105 sq., 112, 114, 117, 126, 130 sqq., 152, 155, 211, 258 sqq.).

— (2015). "Realización fonética del acento español en entonación ascendente [Phonetic realisation of the Spanish stress in a rising intonation]". In: *Perspectivas actuales en el análisis fónico del habla: tradición y avances en la fonética experimental*. Ed. by Adrián Cabedo Nebot. Vol. 7. Anejos de Normas. Valencia: Universitat de València, pp. 91–98. ISBN: 978-84-606-8969-0. (Visited on 08/25/2015) (cit. on p. 130).

Kindaichi, Haruhiko (1981). *Meikai nihongo akusento jiten* [*Clear Japanese accent dictionary*]. 2nd ed. Tōkyō: Sanseido (cit. on p. 77).

Kingston, John (2003). "Learning foreign vowels". In: *Language and Speech* 46.2, pp. 295–348. URL: http://las.sagepub.com/content/46/2-3/295.short (visited on 07/26/2015) (cit. on p. 35).

Klatt, Dennis H. (1987). "Review of text-to-speech conversion for English". In: *The Journal of the Acoustical Society of America* 82.3, pp. 737–793.

ISSN: 0001-4966. DOI: 10.1121/1.395275. URL: http://scitation.aip.org/content/asa/journal/jasa/82/3/10.1121/1.395275 (visited on 10/14/2015) (cit. on p. 172).

Klatt, Dennis H. and Laura C. Klatt (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers". In: *The Journal of the Acoustical Society of America* 87.2, pp. 820–857. ISSN: 0001-4966 (cit. on p. 172).

Koster, Cor J. and Ton Koet (1993). "The evaluation of accent in the english of Dutchmen". In: *Language Learning* 43.1, pp. 69–92. ISSN: 1467-9922. DOI: 10.1111/j.1467-1770.1993.tb00173.x. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-1770.1993.tb00173.x/abstract (visited on 10/17/2015) (cit. on p. 30).

Kubozono, Haruo (1989). "The mora and syllable structure in Japanese: Evidence from speech errors". In: *Language and Speech* 32.3, pp. 249–278 (cit. on p. 50).

— (2006). "Where does loanword prosody come from?: A case study of Japanese loanword accent". In: *Lingua*. Loanword Phonology: Current Issues 116.7, pp. 1140–1170. ISSN: 0024-3841. DOI: 10.1016/j.lingua.2005.06.010. URL: http://www.sciencedirect.com/science/article/pii/S0024384105001099 (visited on 10/20/2015) (cit. on p. 51).

Kuhl, Patricia K. (1991). "Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not". In: *Attention, Perception, & Psychophysics* 50.2, pp. 93–107. URL: http://www.springerlink.com/index/R232234580036375.pdf (visited on 02/25/2013) (cit. on p. 35).

— (2000). "A new view of language acquisition". In: *Proceedings of the National Academy of Sciences* 97.22, pp. 11850–11857. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.97.22.11850. URL: http://www.pnas.org/content/97/22/11850 (visited on 10/16/2015) (cit. on p. 34).

Kuhl, Patricia K., Barbara T. Conboy, Sharon Coffey-Corina, Denise Padden, Maritza Rivera-Gaxiola, and Tobey Nelson (2008). "Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e)". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 363.1493, pp. 979–1000. URL: http://rstb.royalsocietypublishing.org/content/363/1493/979.abstract (visited on 02/25/2013) (cit. on pp. 29, 34).

Kuhl, Patricia K. and James D. Miller (1975). "Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants". In: *Science* 190.4209, pp. 69–72. URL: http://www.sciencemag.org/content/190/4209/69.short (visited on 10/18/2015) (cit. on p. 35).

— (1978). "Speech perception by the chinchilla: identification functions for synthetic VOT stimuli". In: *The Journal of the Acoustical Society of America* 63.3, pp. 905–917. URL: http://scitation.aip.org/content/asa/journal/jasa/63/3/10.1121/1.381770 (visited on 10/18/2015) (cit. on p. 35).

Kuhl, Patricia K., Karen A. Williams, Francisco Lacerda, Ken N. Stevens, and Björn Lindblom (1992). "Linguistic experience alters phonetic perception in infants by 6 months of age". In: *Science* 255.5044, pp. 606–608. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1736364. URL: http://www.sciencemag.org/content/255/5044/606 (visited on 10/19/2015) (cit. on p. 34).

Labrune, Laurence (2012). "Accent". In: *The phonology of Japanese*. New York: Oxford University Press. ISBN: 978-0-19-954583-4 (cit. on pp. 50 sqq., 54, 101).

Lahoz, José María (2011). "Manipulación de claves acústicas para la corrección del acento léxico en la enseñanza de lenguas extranjeras". In: *Laboratorio de idiomas y la enseñanza-aprendizaje de lenguas*. Editorial Complutense, S.A. ISBN: 978-84-9938-095-7 (cit. on p. 219).

— (2012). "Syllable, accent, rhythm: typological and methodological considerations for teaching Spanish as a foreign language". In: *Revista Internacional de Lenguas Extranjeras* 1, pp. 129–150. URL: http://

revistes.publicacionsurv.cat/index.php/rile/article/view/13 (visited on 03/01/2016) (cit. on p. 219).

Lee, Kwing Lok Albert (2015). "The dynamics of Japanese prosody". PhD thesis. University College London (cit. on pp. 39, 51).

Lenneberg, Eric H. (1967). *Biological foundations of language*. Oxford, England: Wiley (cit. on p. 33).

Levis, John M. (1999). "Intonation in Theory and Practice, Revisited". In: *Tesol Quarterly* 33.1, pp. 37–63. ISSN: 0039-8322. DOI: 10.2307/3588190. URL: http://www.jstor.org/stable/3588190 (visited on 10/12/2015) (cit. on p. 30).

Lewis, M. Paul, Gary F. Simons, and Charles D. Fenning, eds. (2015). *Ethnologue: Languages of the World*. 18th ed. Dallas, TX: SIL International. URL: www.ethnologue.com (cit. on pp. 47, 49).

Li, Ping, Sara Sepanski, and Xiaowei Zhao (2006). "Language history questionnaire: A web-based interface for bilingual research". In: *Behavior research methods* 38.2, pp. 202–210. URL: http://link.springer.com/article/10.3758/BF03192770 (visited on 03/14/2016) (cit. on p. 62).

Li, Ping, Fan Zhang, Erlfang Tsai, and Brendan Puls (2014). "Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool". In: *Bilingualism: Language and Cognition* 17.3, pp. 673–680. ISSN: 1469-1841. DOI: 10.1017/S1366728913000606. URL: http://journals.cambridge.org/article_S1366728913000606 (visited on 03/14/2016) (cit. on p. 62).

LimeSurvey Project Team/Carsten Smith (2012). *LimeSurvey: An open source survey tool*. Comp. software. Hamburg. URL: http://www.limesurvey.org (cit. on pp. 107, 135, 253).

Llisterri, Joaquim, María Jesús Machuca, Carme de la Mota, Montserrat Riera, and Antonio Ríos (2003). "Algunas cuestiones en torno al desplazamiento acentual en español [Some issues on accent displacement in Spanish]". In: *La tonía: dimensiones fonéticas y fonológicas*. Ed. by Esther Herrera Z. and Pedro Martín Butragueño. México: El Colegio de

México, pp. 163–185. ISBN: 968-12- 1123-5. URL: http://liceu.uab.es/%20joaquim/publicacions/Llisterri_et_al_2002.pdf (cit. on pp. 43, 45, 64, 73 sq., 78, 211).

— (2005). "La percepción del acento léxico en español [The perception of lexical stress in Spanish]". In: *Filología y lingüística. Estudios ofrecidos a Antonio Quilis*. Vol. 1. Madrid, pp. 271–297. ISBN: 84-00-08401-2 (cit. on pp. 43, 45, 64, 73 sq., 78, 211 sq.).

Llisterri, Joaquim, Rafael Marín, Carme de la Mota, and Antonio Ríos (1995). "Factors affecting F_0 peak displacement in Spanish". In: *Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Technology*. Vol. 3. Madrid: ISCA, pp. 2061–2064. URL: http://liceu.uab.cat/%20joaquim/publicacions/Llisterri_Marin_Mota_Rios_95_Peak_Displacement.pdf (cit. on pp. 45, 49).

Lucero, Jorge C., Kevin G. Munhall, Vincent L. Gracco, and James O. Ramsay (1997). "On the registration of time and the patterning of speech movements". In: *Journal of Speech, Language, and Hearing Research* 40.5, pp. 1111–1117. URL: http://jslhr.pubs.asha.org/article.aspx?articleid=1781909 (visited on 03/14/2016) (cit. on p. 83).

Macdonald, Shem (2002). "Pronunciation-views and practices of reluctant teachers". In: URL: http://www.researchonline.mq.edu.au/vital/access/services/Download/mq:35597/DS01 (visited on 03/15/2016) (cit. on p. 218).

MacIntyre, Peter D., Kimberly A. Noels, and Richard Clément (1997). "Biases in self-ratings of second language proficiency: The role of language anxiety". In: *Language learning* 47.2, pp. 265–287. URL: http://www.psych.ualberta.ca/~knoels/personal/Kim%27s%20publications/MacIntyreNoelsClement1997.pdf (visited on 05/05/2015) (cit. on p. 62).

Marian, Viorica, Henrike K. Blumenfeld, and Margarita Kaushanskaya (2007). "The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals". In: *Journal of Speech, Language, and Hearing Research* 50.4, pp. 940–967.

URL: http://jslhr.pubs.asha.org/article.aspx?articleid=1765522 (visited on 05/05/2015) (cit. on p. 62).

Martínez, I. (1995). "Aproximación de las dificultades de aprendizaje en fonética y fonología españolas por parte de alumnos japoneses [On the learning difficulties of Japanese learners of Spanish phonetics and phonology]". In: *Lingüística Hispánica* 18, pp. 35–60 (cit. on p. 59).

Mayo, Catherine and Alice Turk (2004). "Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions". In: *The Journal of the Acoustical Society of America* 115, pp. 3184–3194. URL: http://link.aip.org/link/?JASMAN/115/3184/1 (visited on 05/14/2013) (cit. on pp. 162, 181).

— (2005). "The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception". In: *The Journal of the Acoustical Society of America* 118.3, pp. 1730–1741. ISSN: 0001-4966. DOI: 10.1121/1.1979451. URL: http://link.aip.org/link/?JASMAN/118/1730/1 (visited on 05/14/2013) (cit. on pp. 162, 181).

McCawley, James D. (1968). *The phonological component of a grammar of Japanese*. 2. Mouton (cit. on p. 51).

— (1978). "What is a tone language?" In: *Tone: a linguistic survey*. Ed. by Victoria Fromkin. New York: Academic Press, Inc., pp. 113–132. ISBN: 0-12-267350-6 (cit. on p. 49).

Minematsu, N., M. Suzuki, H. Hirano, C. Nakagawa, N. Nakamura, Y. Tagawa, and K. Hirose (2012). "Onsei shutsuryoku kinō wo yūshita onrain akusento jisho no kōchiku [Development of an online Japanese accent dictionary with speech output facility]". In: *Proceedings of the International Conference on Japanese Language Education (ICJLE)*, p. 94 (cit. on p. 219).

Minematsu, Nobuaki and Keikichi Hirose (1995). "Role of prosodic spoken features words in the human process of perceiving spoken words and sentences in Japanese". In: *Journal of the Acoustical Society of Japan*

16.5, pp. 311–320. URL: https://www.jstage.jst.go.jp/article/ast1980/16/5/16_5_311/_pdf (cit. on p. 55).

Muñoz Garcia, Margarita (2010). *La perception et la production de l'accent lexical de l'espagnol par des francophones: aspects phonétiques et psycholinguistes* [*The perception and production of Spanish lexical stress by French speakers: phonetic and psycholinguistic aspects*]. Toulouse 2. URL: http://www.theses.fr/2010TOU20058 (visited on 03/01/2016) (cit. on p. 56).

Muñoz, N. Panissal, M. Billières, and L. Baqué (2009). "¿La metáfora de la criba fonológica se puede aplicar a la percepción del acento léxico español? Estudio experimental con estudiantes francófonos [Is the metaphor of the phonological sieve applicable to the perception of Spanish lexical stress? An experimental study with rench speakers]". In: pp. 489–500 (cit. on pp. 56 sq.).

Munro, Murray J. (1995). "Nonsegmental factors in foreign accent: ratings of filtered speech". In: *Studies in Second Language Acquisition* 17.1, p. 17. ISSN: 0272-2631, 1470-1545. DOI: 10.1017/S0272263100013735. URL: http://www.journals.cambridge.org/abstract_S0272263100013735 (visited on 10/17/2015) (cit. on p. 30).

Munro, Murray J. and Tracey M. Derwing (1999). "Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners". In: *Language Learning* 49, pp. 285–310. ISSN: 1467-9922. DOI: 10.1111/0023-8333.49.s1.8. URL: http://onlinelibrary.wiley.com/doi/10.1111/0023-8333.49.s1.8/abstract (visited on 10/12/2015) (cit. on p. 30).

Murphy, John (2014). "Teacher training programs provide adequate preparation in how to teach pronunciation". In: *Pronunciation myths: Applying second language research to classroom teaching*. Ed. by Linda Grant. Ann Arbor, MI: University of Michigan Press. ISBN: 978-0-472-03516-8 (cit. on p. 218).

Nakamura, Ibuki, Nobuaki Minematsu, Masayuki Suzuki, Hiroko Hirano, Chieko Nakagawa, Noriko Nakamura, Yukinori Tagawa, Keikichi Hirose, and Hiroya Hashimoto (2013). "Development of a web framework for teaching and learning Japanese prosody: OJAD (Online Japanese Accent Dictionary)". In: *INTERSPEECH*, pp. 2554–2558. URL: http://pallas.gavo.t.u-tokyo.ac.jp/~mine/paper/PDF/2013/INTERSPEECH_p2554-2558_t2013-8.PDF (visited on 07/28/2015) (cit. on p. 219).

Navarro Tomás, Tomás (1918). *Manual de entonación española* [*Handbook of Spanish intonation*]. Madrid: Imprenta de los Sucesores de Hernando (cit. on p. 42).

Nishinuma, Yukihiro, Masako Arai, and Takako Ayusawa (1996). "Perception of tonal accent by Americans learning Japanese". In: *Proceedings of the 4th International Conference on Spoken Language Processing*. Philadelphia, pp. 646–649 (cit. on pp. 58 sq., 67 sq.).

Nittrouer, Susan and Michael Studdert-Kennedy (1987). "The role of coarticulatory effects in the perception of fricatives by children and adults". In: *Journal of Speech, Language and Hearing Research* 30.3, pp. 73–93. URL: http://jslhr.asha.org/cgi/content/abstract/30/3/319 (visited on 09/23/2013) (cit. on pp. 162, 181).

Orta Gracia, Antonio (2009). "Creencias de los profesores acerca de la pronunciación y sus repercusiones en el aula [Teachers beliefs on pronunciation and their repercussions in the classroom]". In: *PHONICA* 5. ISSN: 1699-8774. URL: http://revistes.ub.edu/index.php/phonica/article/view/5595 (visited on 03/01/2016) (cit. on p. 218).

— (2010). "La enseñanza explícita de la pronunciación: creencias de los profesores y sus repercusiones en el aula de E-LE [Explicit pronunciation teaching: teacher beliefs and their repercussions in the E-LE classroom]". Masters. Barcelona: Universitat de Barcelona Virtual. 122 pp. URL: http://redined.mecd.gob.es/xmlui/handle/11162/77003 (visited on 03/01/2016) (cit. on p. 218).

Ortega-Llebaria, Marta (2006). "Phonetic cues to stress and accent in Spanish". In: *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*. Ed. by Manuel Díaz-

Campos. Cascadilla Proceedings Project. Somerville, MA: Cascadilla, pp. 104–118 (cit. on pp. 44, 162, 174).

Ortega-Llebaria, Marta and Pilar Prieto (2007). "Disentangling stress from accent in Spanish: production patterns of the stress contrast in deaccented syllables". In: *Segmental and prosodic issues in Romance Phonology*. Ed. by Pilar Prieto, J. Mascaró, and Maria-Josep Solé. Current Issues in Linguistic Theory. John Benjamins. URL: http://filcat.uab.cat/clt/publicacions/reports/pdf/GGT-05-11.pdf (cit. on pp. 44, 149, 162, 212).

— (2009). "Perception of word stress in Castilian Spanish". In: *Phonetics and phonology: interactions and interrelations*. Ed. by Marina Vigário, Sonia Frota, and M. João Freitas. Vol. 306. Amsterdam studies in the theory and history of linguistic science IV. Amsterdam, PA: John Benjamins, pp. 35–50. ISBN: 978-90-272-4822-0. (Visited on 07/15/2015) (cit. on pp. 44, 162).

— (2011). "Acoustic correlates of stress in central Catalan and Castillian Spanish". In: vol. 54. 1, pp. 73–97. DOI: 10.1177/0023830910388014. URL: http://las.sagepub.com/content/54/1/73 (cit. on p. 44).

Ortega-Llebaria, Marta, Pilar Prieto, and María del Mar Vanrell (2008). "Perceptual evidence for direct acoustic correlates of stress in Spanish". In: *Proceedings of the XVIth International Conference on Phonetic Sciences*, pp. 1121–1124 (cit. on pp. 44, 162).

Ortiz Lira, Héctor (2003). "Los acentos tonales en un corpus de español de Santiago de Chile: su distribución y realización [Pitch accents in a corpus of Spanish from Santiago de Chile: their distribution and production]". In: *La tonía: dimensiones fonéticas y fonológicas*, pp. 303–316 (cit. on pp. 41, 49).

Ortiz Lira, Héctor, Marcela Fuentes, and Lluïsa Astruc (2010). "Chilean Spanish intonation". In: *Transcription of intonation of the Spanish language*. Ed. by Pilar Prieto and Paolo Roseano. München: Lincom Europa, pp. 255–283. ISBN: 9783862901845 (cit. on pp. 48, 212).

Otake, Takashi and Anne Cutler (1999). “Perception of suprasegmental structure in a non-native dialect”. In: *Journal of Phonetics* 27.3, pp. 229–253. ISSN: 00954470. DOI: 10.1006/jpho.1999.0095. URL: http://linkinghub.elsevier.com/retrieve/pii/S0095447099900955 (cit. on p. 55).

Otake, Takashi, Giyoo Hatano, Anne Cutler, and Jacques Mehler (1993). “Mora or syllable? Speech segmentation in Japanese”. In: *Journal of Memory and Language* 32, pp. 258–278 (cit. on p. 56).

Paolacci, Gabriele and Jesse Chandler (2014). “Inside the Turk Understanding Mechanical Turk as a Participant Pool”. In: *Current Directions in Psychological Science* 23.3, pp. 184–188. ISSN: 0963-7214, 1467-8721. DOI: 10.1177/0963721414531598. URL: http://cdp.sagepub.com/content/23/3/184 (visited on 03/08/2016) (cit. on pp. 107 sq.).

Pierrehumbert, Janet B. (1980). *The phonology and phonetics of English intonation*. Cambridge, MA: Massachusetts Institute of Technology (cit. on pp. 38, 49 sq.).

Pierrehumbert, Janet B. and Mary E. Beckman (1988). *Japanese tone structure*. MIT Press. ISBN: 978-0-262-16109-1 (cit. on p. 38).

Price, Patti J., Mari Ostendorf, Stefanie Shattuck-Hufnagel, and Cynthia Fong (1991). “The use of prosody in syntactic disambiguation”. In: *the Journal of the Acoustical Society of America* 90.6, pp. 2956–2970. URL: http://scitation.aip.org/content/asa/journal/jasa/90/6/10.1121/1.401770 (visited on 10/21/2015) (cit. on p. 38).

Prieto, Pilar, Jan Van Santen, and Julia Hirschberg (1995). “Tonal alignment patterns in Spanish”. In: *Journal of Phonetics* 23.4, pp. 429–451. URL: http://www.sciencedirect.com/science/article/pii/S0095447085700320 (visited on 03/01/2016) (cit. on p. 45).

Purcell, Edward T. and Richard W. Suter (1980). “Predictors of pronunciation accuracy: a reexamination”. In: *Language Learning* 30.2, pp. 271–287. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-1770.1980.tb00319.x/abstract (visited on 10/26/2015) (cit. on p. 60).

Quilis, Antonio (1983). “Frecuencia de los esquemas acentuales en español [The frequency of accentual patterns in Spanish]”. In: *Estudios ofrecidos a Emilio Alarcos Llorach*. Vol. 5. Universidad de Oviedo, pp. 113–126. ISBN: 84-600-3197-7. URL: http://dialnet.unirioja.es/servlet/articulo?codigo=1311553 (visited on 03/01/2016) (cit. on pp. 40 sq.).

R. Core Team (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN: 3-900051-07-0. URL: http://www.R-project.org/ (cit. on pp. 83, 114, 137, 182).

Ramus, Franck, Marc D. Hauser, Cory Miller, Dylan Morris, and Jacques Mehler (2000). “Language discrimination by human newborns and by cotton-top tamarin monkeys”. In: *Science* 288.5464, pp. 349–351. URL: http://www.sciencemag.org/content/288/5464/349.short (visited on 10/18/2015) (cit. on p. 35).

Reed, Edward and Rebecca Jones (1982). *Reasons for realism: Selected essays of James J. Gibson*. Hillsdale, NJ: Erlbaum (cit. on p. 31).

Ruiz Mella, Magaly and Yasna Pereira (2010). “Acento léxico: tendencias de los correlatos acústicos [Lexical stress: tendencies of its acoustic correlates]”. In: *Onomázein* 22.2, pp. 43–58. URL: http://www.onomazein.net/Articulos/22/2_Ruiz.pdf (cit. on p. 45).

Saalfeld, Anita K. (2012). “Teaching L2 Spanish Stress”. In: *Foreign Language Annals* 45.2, pp. 283–303. ISSN: 1944-9720. DOI: 10.1111/j.1944-9720.2012.01191.x. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1944-9720.2012.01191.x/abstract (visited on 03/01/2016) (cit. on p. 217).

Sadowsky, Scott and Ricardo Martínez Gamboa (2012). *Lista de frecuencias de palabras del castellano de Chile (Lifcach)*. Electronic database. URL: http://sadowsky.cl/lifcach.html (visited on 09/01/2015) (cit. on p. 74).

Sanders, Lisa D., Helen J. Neville, and Marty G. Woldorff (2002). “Speech segmentation by native and non-native speakers: The use of lexical, syntactic, and stress-pattern cues". In: *Journal of Speech, Language, and Hearing Research* 45.3, pp. 519–530 (cit. on p. 30).

Schwab, Sandra and Joaquim Llisterri (2011a). "La perception de l'accent lexical espagnol par des apprenants francophones [The perception of Spanish lexical stress by French-speaking learners]". In: *La langue et l'être communiquant. Hommage à Julio Murillo.* Ed. by L. Baqué and M. Estrada, pp. 311–328 (cit. on pp. 56 sq., 211, 216).

— (2011b). "The perception of Spanish lexical stress by French speakers: stress identification and time cost". In: *Achievements and perspectives in SLA of Speech: New Sounds 2010.* Ed. by Katarzyna Dziubalska-Kołaczyk, Magdalena Wrembel, and Małgorzata Kul. Vol. 1. Poznań: Peter Lang, pp. 229–242. ISBN: 978-83-928167-9-9. (Visited on 02/19/2013) (cit. on pp. 56 sq., 211, 216).

— (2012). "The role of acoustic correlates of stress in the perception of Spanish accentual contrasts by French speakers". In: 6th International Conference on Speech Prosody. Ed. by Q. Ma, H. Ding, and Daniel J. Hirst. Vol. 1. Shanghai: Tongji University Press, pp. 350–353. URL: http:%20//liceu.uab.cat/~joaquim/publicacions/Schwab_%20Llisterri_Perception_Lexical_Stress_12.pdf (cit. on pp. 56 sqq., 216).

— (2014). "Does training make French speakers more able to identify lexical stress?" In: *Concordia Working Papers in Applied Linguistics* 5, pp. 624–636 (cit. on pp. 56 sq., 60, 216).

Sharoff, Serge (2006). "Creating general-purpose corpora using automated search engine queries". In: *WaCky! Working papers on the web as corpus.* Ed. by Marco Baroni and Silvia Bernardini. Bologna: Gedit, pp. 63–98. ISBN: 88-6027-004-9. URL: http://wackybook.sslmit.unibo.it/pdfs/sharoff.pdf (visited on 09/04/2015) (cit. on pp. 78 sq.).

— (2007). "Classifying Web corpora into domain and genre using automatic feature identification". In: *Proc. of Web as Corpus Workshop.* Louvain-la-Neuve (cit. on pp. 78 sq.).

Shipley-Brown, Frances, William O. Dingwall, Charles I. Berlin, Grace Yeni-Komshian, and Sandra Gordon-Salant (1988). "Hemispheric processing of affective and linguistic intonation contours in normal subjects". In: *Brain and language* 33.1, pp. 16–26. URL: http://www.sciencedirect.com/science/article/pii/0093934X8890051X (visited on 11/10/2015) (cit. on p. 37).

Sibata, Takeshi (1994). "Gairaigo ni okeru akusentokaku no ichi [Position of the accent kernel in Japanese loan words]". In: *Kokugo ronkyū 4: gendaigo hōgen no kenkyū*. Ed. by Kiyoji Satō. Tōkyō: Meiji Shoin, 418–388(1–31). ISBN: 4-625-52064-9 (cit. on p. 51).

Sibata, Takeshi and Ritei Shibata (1990). "Akusento wa dōongo wo dono teidō benbetsu shiuru ka [Is word-accent significant in differentiating homonyms in Japanese, English and Chinese?]" In: *Keiryō kokugogaku* 17.7, pp. 317–327 (cit. on pp. 54 sq., 76 sq.).

Silverman, Kim E.A., Mary E. Beckman, John F. Pitrelli, Mari Ostendorf, Colin W. Wightman, Patti Price, Janet B. Pierrehumbert, and Julia Hirschberg (1992). "ToBI: a standard for labeling English prosody". In: *The Second International Conference on Spoken Language Processing, ICSLP 1992, Banff, Alberta, Canada, October 13-16, 1992* (cit. on pp. 38, 84).

So, Connie K. and Catherine T. Best (2008). "Do English speakers assimilate Mandarin tones to English prosodic categories?" In: *INTERSPEECH 2008*. 9th Annual Conference of the International Speech Communication Association. Ed. by Janet Fletcher, Deborah Loakes, Roland Gocke, Denis Burnham, and Michael Wagn. Brisbane, p. 1120. (Visited on 10/18/2015) (cit. on pp. 36 sq., 209, 215).

— (2011). "Categorizing Mandarin tones into listeners' native prosodic categories: the role of phonetic properties". In: *Poznań Studies in Contemporary Linguistics PSiCL* 47, p. 133. URL: http://www.degruyter.com/view/j/psicl.2011.47.issue-1/psicl-2011-0011/psicl-2011-0011.xml (visited on 10/18/2015) (cit. on pp. 36 sq., 209, 215).

Sosa, Juan Manuel (2003). "La notación tonal del español en el modelo SP-TOBI [Tonal annotation in SP-TOBI]". In: *Teorías de la entonación.*

Ariel, pp. 185–208. URL: http://dialnet.unirioja.es/servlet/articulo?codigo=3903853 (visited on 10/21/2015) (cit. on p. 38).

Sprouse, Jon (2011). “A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory”. In: *Behavior research methods* 43.1, pp. 155–167. URL: http://link.springer.com/article/10.3758/s13428-010-0039-7 (visited on 03/08/2016) (cit. on pp. 107 sq.).

Sugitō, Miyoko (1982). “Tōkyō akusento ni okeru *hana* to *hana* no sēsē to chikaku [The production and perception of *flower* and *nose* with Tōkyō accent]”. In: *Nihongo akusento no kenkyū*. Tōkyō: Sanseidō, pp. 182–201 (cit. on pp. 51 sq.).

Thomson, R. I. and T. M. Derwing (2014). “The effectiveness of L2 pronunciation instruction: a narrative review”. In: *Applied Linguistics* 36.3, pp. 326–344. ISSN: 0142-6001, 1477-450X. DOI: 10.1093/applin/amu076. URL: http://applij.oxfordjournals.org/cgi/doi/10.1093/applin/amu076 (visited on 10/17/2015) (cit. on p. 29).

Tiffen, Brian Witney (1974). “The intelligibility of Nigerian English.” PhD thesis. University of London. URL: http://discovery.ucl.ac.uk/1349554/ (cit. on p. 30).

Toledo, Guillermo and Lluïsa Astruc (2008). “Los acentos tonales ascendentes en el español de Santiago de Chile [Rising pitch accents in the Spanish of Santiago de Chile]”. In: *Language Design* (Special Issue 2), pp. 65–72 (cit. on p. 49).

Tremblay, Annie (2011). “Proficiency assessment standards in second language acquisition research”. In: *Studies in Second Language Acquisition* 33.3, pp. 339–372. URL: http://journals.cambridge.org/abstract_S0272263111000015 (visited on 04/22/2015) (cit. on pp. 60 sq., 63).

Tremblay, Annie and Meryl D. Garrison (2010). “Cloze tests: A tool for proficiency assessment in research on L2 French”. In: *Selected proceedings of the Second Lamnguage Research Forum 2008*. Second Language Research Forum 2008. Ed. by M.T. Prior, Y. Watanabe, and S-K. Lee.

Somerville, MA: Cascadilla Press, pp. 73–88. (Visited on 04/22/2015) (cit. on p. 61).

Trofimovich, Pavel and Wendy Baker (2006). "Learning second language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of L2 speech". In: *Studies in Second Language Acquisition* 28.1, pp. 1–30. ISSN: 1470-1545. DOI: 10.1017/S0272263106060013. URL: http://journals.cambridge.org/article_S0272263106060013 (visited on 10/12/2015) (cit. on pp. 30, 37).

Tsujimura, Natsuko, ed. (2007). *The Handbook of Japanese Linguistics*. 2nd ed. Blackwell (cit. on p. 49).

Ueda, Hiroto (1977). "Estudio contrastivo de los sonidos españoles y japoneses (1): vocales y semivocales [A contrastive study of the sounds of Spanish and Japanese (1): vowels and semivowels]". In: *Lexicon* 6, pp. 29–46 (cit. on p. 59).

— (1978). "Estudio contrastivo de los sonidos españoles y japoneses (2): consonantes [A contrastive study of the sounds of Spanish and Japanese (2): consonants]". In: *Lexicon* 7, pp. 16–37 (cit. on p. 59).

Urrutia Cardenas, Hernán (2007). "La naturaleza del acento en español: nuevos datos y perspectivas [The nature of Spanish stress: new data and perspectives]". In: *Revista de lingüística teórica y aplicada* 45.2, pp. 135–142. ISSN: 0718-4883. DOI: 10.4067/S0718-48832007000200010. URL: http://dx.doi.org/10.4067/S0718-48832007000200010 (visited on 03/09/2016) (cit. on pp. 42, 45).

Usó Viciedo, Lidia (2009). "Creencias de los profesores en formación sobre la enseñanza de la pronunciación del español [Beliefs of teachers in training on the teaching of Spanish pronunciation]". In: *Marco ELE* 8. ISSN: 1885-2211. URL: http://marcoele.com/creencias-de-los-profesores-en-formacion-sobre-la-pronunciacion-del-espanol/ (cit. on p. 218).

Vance, Timothy J. (1995). "Final accent vs. no accent: utterance-final neutralization in Tōkyō Japanese". In: *Journal of Phonetics* 23, pp. 487–499 (cit. on pp. 51 sq., 79, 88, 100).

Vance, Timothy J. (2008). *The sounds of Japanese*. New York: Cambridge University Press. ISBN: 978-0-521-61754-3 (cit. on pp. 50, 52–55, 151).

Velásquez, Francisco, Alexander Gelbukh, and Grigori Sidorov (2002). "AGME: un sistema de análisis y generación de la morfología del español [AGME: a system for the analysis and generation of Spanish morphology]". In: *Proc. of Workshop on Multilingual information access & natural language processing of IBERAMIA 2002*, pp. 1–6. URL: http://micron.gelbukh.com/CV/Publications/2002/MLIA-2002-Morph.pdf (visited on 02/26/2013) (cit. on p. 41).

Venditti, Jennifer J. (1995). "Japanese TOBI labelling guidelines". In: *Working Papers In Linguistics-Ohio State University Department Of Linguistics*, pp. 127–162. URL: http://www1.cs.columbia.edu/~jjv/pubs/jtobi_oct.ps (visited on 10/21/2015) (cit. on p. 38).

— (2005). "The J_TOBI model of Japanese intonation". In: *Prosodic typology: the phonology of intonation and phrasing*. Ed. by Sun-Ah Jun. Oxford; New York: Oxford University Press, pp. 172–200. ISBN: 978-0-19-924963-3 (cit. on p. 38).

Wagner, Petra, Antonio Origlia, Cinzia Avesani, George Christodoulides, Francesco Cutugno, Mariapaola D'Imperio, David Mancebo Escudero, Barbara Gili Fivela, Anne Lacheret, Bogdan Ludusan, Helena Moniz, Ailbhe Ní Chasaide, Oliver Niebuhr, Lucie Rousier-Vercruyssen, Anne-Catherine Simon, Juraj Šimko, Fabio Tesser, and Martti Vainio (2015). "Different parts of the same elephant: a roadmap to disentangle and connect different perspectives on prosodic prominence". In: *Proceedings of the 18th International Congress of Phonetic Sciences*. International Congress of Phonetic Sciences. Ed. by The Scottish Consortium for ICPhS 2015. Paper number 202. Glasgow: University of Glasgow, pp. 1–5. ISBN: 978-0-85261-941-4. URL: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0202.pdf (visited on 09/01/2015) (cit. on p. 39).

Wang, Yue and Patricia K. Kuhl (2003). "Evaluating the "critical period" hypothesis: perceptual learning of Mandarin tones in American adults and American children at 6, 10 and 14 years of age". In: *Proceedings of the 15th International Congress of Phonetic Sciences*. International Congress of Phonetic Sciences. Ed. by Maria-Josep Solé, Daniel Recasens, and Joaquín Romero. Barcelona, pp. 1537–1540. URL: http://www.internationalphoneticassociation.org/icphs/icphs2003 (visited on 10/27/2015) (cit. on p. 60).

Wang, Yue, Michelle M. Spence, Allard Jongman, and Joan A. Sereno (1999). "Training American listeners to perceive Mandarin tones". In: *The Journal of the Acoustical Society of America* 106.6, pp. 3649–58. ISSN: 0001-4966. URL: http://www.ncbi.nlm.nih.gov/pubmed/10615703 (cit. on p. 60).

Wei, Peipei, Lucy Gubbins, and Kaori Idemaru (2015). "Strong influence of prosody on the perception of foreign accent". In: *Proceedings of the 18th International Congress of Phonetic Sciences*. International Congress of Phonetic Sciences. Ed. by The Scottish Consortium for ICPhS 2015. Paper number 185. Glasgow: University of Glasgow, pp. 1–4. ISBN: 978-0-85261-941-4. URL: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0185.pdf (visited on 09/01/2015) (cit. on p. 29).

Wennerstrom, Ann (2000). "The role of intonation in second language fluency". In: *Perspectives on fluency*. Ed. by Heidi Riggenbach. Ann Arbor, MI: University of Michigan Press, pp. 102–127. ISBN: 978-0-472-08604-7 (cit. on p. 30).

Werker, Janet F. and Richard C. Tees (1984). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life". In: *Infant behavior and development* 7.1, pp. 49–63. URL: http://www.sciencedirect.com/science/article/pii/S0163638384800223 (cit. on p. 34).

Wilson, Ian (2008). “Using Praat and Moodle for teaching segmental and suprasegmental pronunciation”. In: *Proceedings of the 3rd international WorldCALL Conference: Using Technologies for Language Learning (WorldCALL 2008)*, pp. 112–115. URL: http://www.u-aizu.ac.jp/~wilson/Wilson2008WorldCALLProcLowRez.pdf (visited on 03/14/2016) (cit. on p. 219).

Xu, Yi (2004). “Transmitting tone and intonation simultaneously — The parallel encoding and target approximation (PENTA) model”. In: *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*. URL: http://www.isca-speech.org/archive_open/tal2004/tal4_215.html (visited on 10/21/2015) (cit. on p. 38).

Yasutomi, Yuhei (1994). “Sobre la pronunciación de la [k] española: estudio acústico de la consonante oclusiva velar sorda entre el español y el japonés [On the pronunciation of Spanish [k]: an acoustic study of the voiceless velar stop in Spanish and Japanese]”. In: *Hispanica* 1994.38, pp. 137–148. URL: http://jlc.jst.go.jp/JST.Journalarchive/hispanica1965/1994.137?from=Google (visited on 10/21/2015) (cit. on p. 59).

Zen, Heiga, Tomoki Toda, Masaru Nakamura, and Keiichi Tokuda (2007). “Details of the Nitech HMM-based speech synthesis system for the Blizzard Challenge 2005”. In: *IEICE transactions on information and systems* 90.1, pp. 325–333. URL: http://search.ieice.org/bin/summary.php?id=e90-d_1_325 (visited on 10/14/2015) (cit. on p. 166).

Appendices

Appendix A

Linguistic questionnaire

To illustrate the language background questionnaire that was presented to the participants, the questions and format are reproduced below in English. However, it must be noted that like the rest of the tests, the actual questionnaire was implemented using LimeSurvey, and was shown to the participants in their native language. Also, since the original questionnaire was dynamically generated, the questions that were shown to each participant depended on their previous answers, and so did the labels for some of the questions themselves.

What is shown below should therefore not be taken to reflect the exact questionnaire that was used for data collection, but to illustrate the list of questions, their number and arrangement, and the type of answers that were collected.

1 What is your gender?

☐ Male ☐ Female

2 In what region of (Japan, Chile) were you raised?

____

3 For how many years have you lived in (Santiago, the Kantō region)?

____

4 How old are you?

____

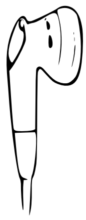
5 Are you currently enrolled as a student?

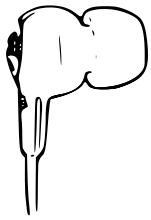
☐ Yes ☐ No

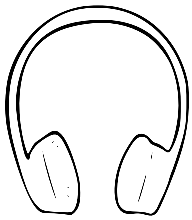
6 Are you taking this test…

6.1	…as part of a group?	☐ Yes	☐ No
6.2	…wearing headphones?	☐ Yes	☐ No
6.3	…in a quiet room?	☐ Yes	☐ No

7 If you are wearing headphones, what kind of headphones are you wearing?


☐ Earbud


☐ In-ear


☐ On-ear


☐ Over-ear

8 If you are wearing headphones, how good would you say they are?

Below average	…	Average	…	Above average
☐	☐	☐	☐	☐

9 How long have you been studying (Spanish, Japanese)?

____

10 If you've studied (Spanish, Japanese): For each of the following linguistic skills in Spanish, how proficient do you consider yourself as being?

	Beginner	…	Intermediate	…	Advanced
Listening	☐	☐	☐	☐	☐
Speaking	☐	☐	☐	☐	☐
Reading	☐	☐	☐	☐	☐
Writing	☐	☐	☐	☐	☐

11 If you've studied (Spanish, Japanese): For each of the following, how much would you say you use it?

	Daily	Weekly	< Weekly	Monthly	< Monthly
At home	☐	☐	☐	☐	☐
With friends	☐	☐	☐	☐	☐
For fun	☐	☐	☐	☐	☐
At work	☐	☐	☐	☐	☐
By yourself	☐	☐	☐	☐	☐

12 Have you ever lived in a (Spanish, Japanese)-speaking country? If so, for how long?

Never	< 6 months	< 1 year	>= 1 year
☐	☐	☐	☐

13 If you have, please write the name of the country or countries where you have lived.

____

14 Can you speak any other languages?

☐ Yes ☐ No

15 If so, please list the languages you know, starting from the one you consider to know best. You can leave your native language out.

____

Appendix B

Additional material from Kimura et al. (2012)

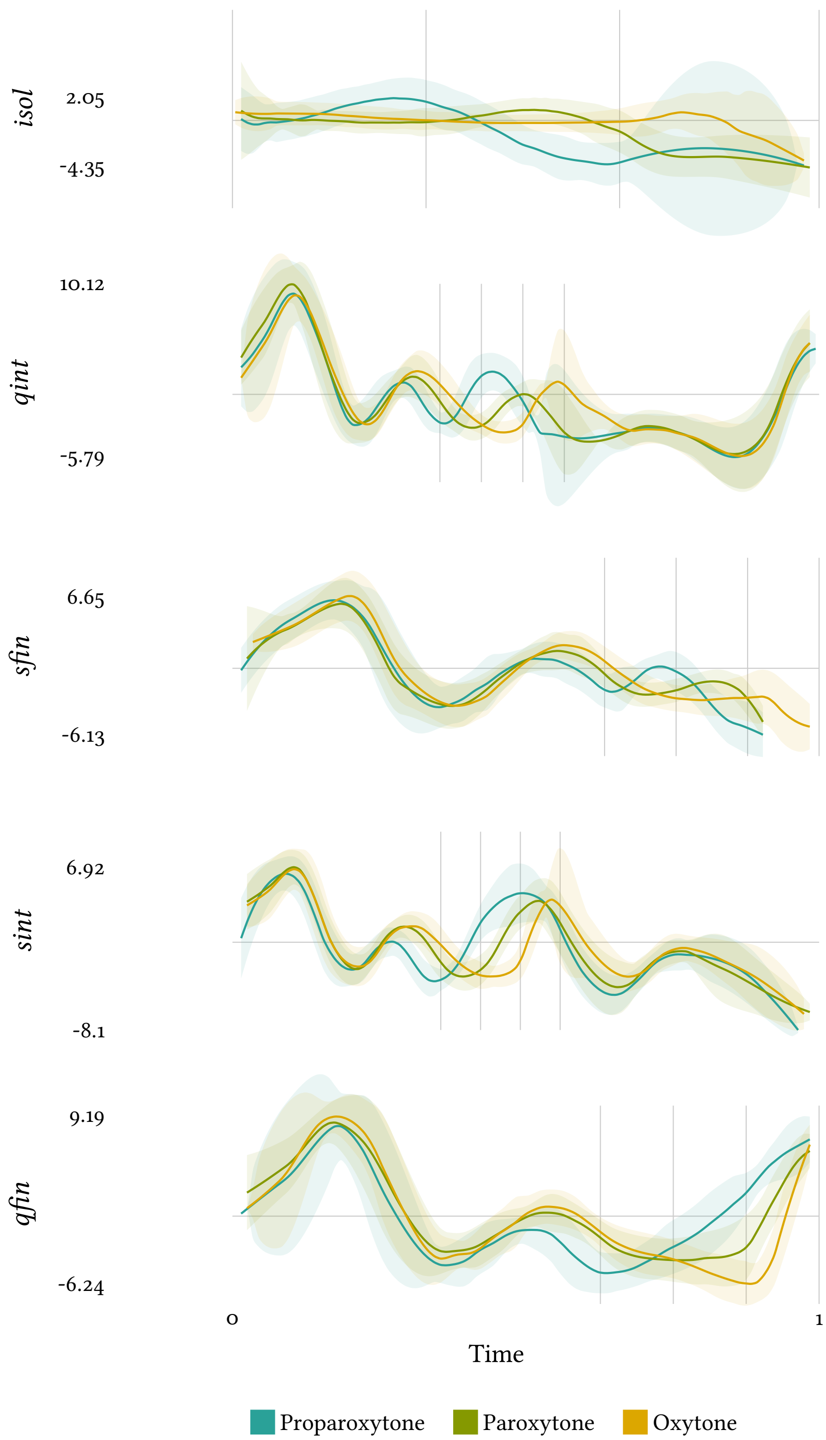
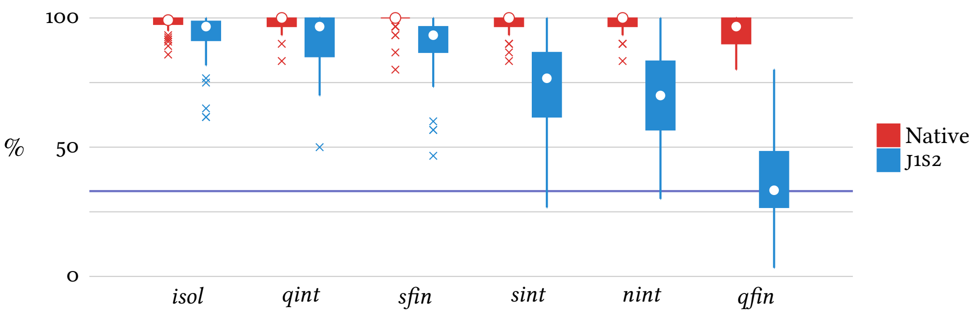
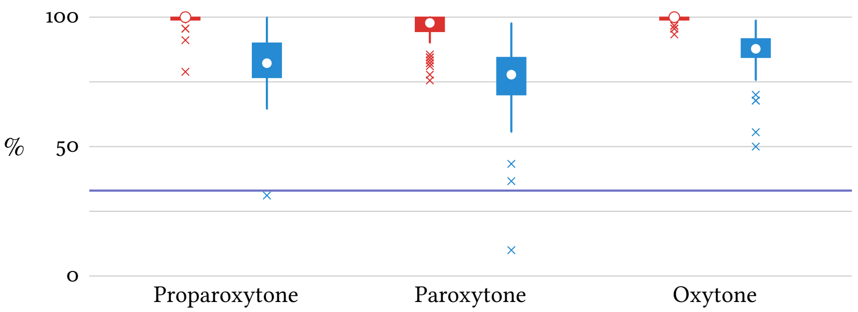


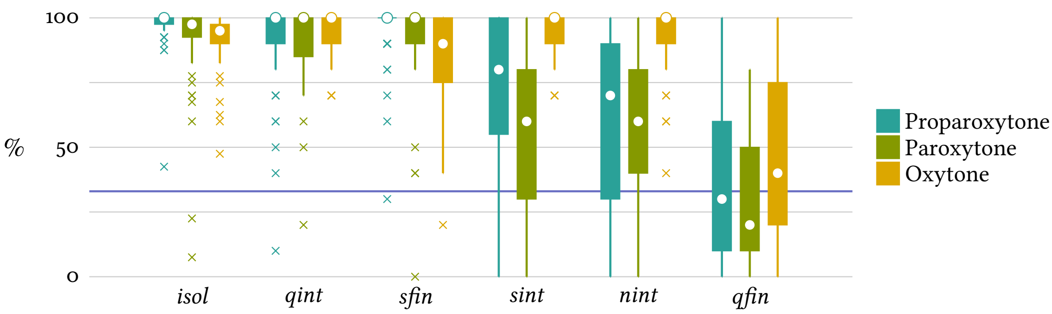
Figure B.1: normalised F0 tracks and prediction intervals of stimuli used in Kimura et al. (2012) per sentence context. F0 values have been converted to semitones in reference to the utterance mean (marked by the horizontal line) and averaged in time over the multiple repetitions. Syllables in the keyword, whose boundaries are shown in the vertical lines, has been equalised so comparisons across words with different accent types are possible. Compare with the new Spanish stimuli in figure 2.1, on page 85.



(a) Results per sentence



(b) Results per accent



(c) Interaction between sentence and accent

Figure B.2: Overall results from the data reported on in Kimura et al. (2012). Figures contain box plots of responses per sentence type (top) and accent position (centre). The bottom plot shows the interaction between the two for non-native speakers. The violet ● line marks the chance level. Compare with more recent results for Spanish in figure 3.3, on page 116.

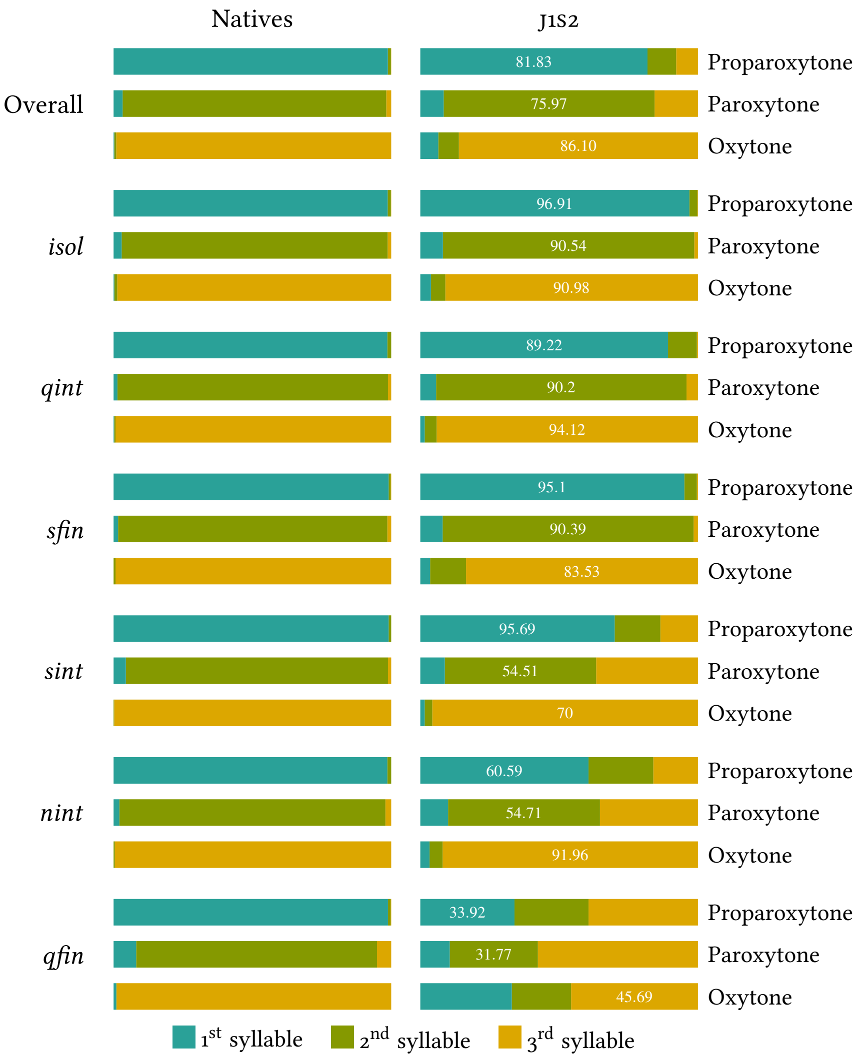


Figure B.3: results from Kimura et al. (2012). Syllables perceived as accented by both groups of participants across all sentence contexts, and in each individually.

Appendix C

Normalised F_0 tracks per speaker

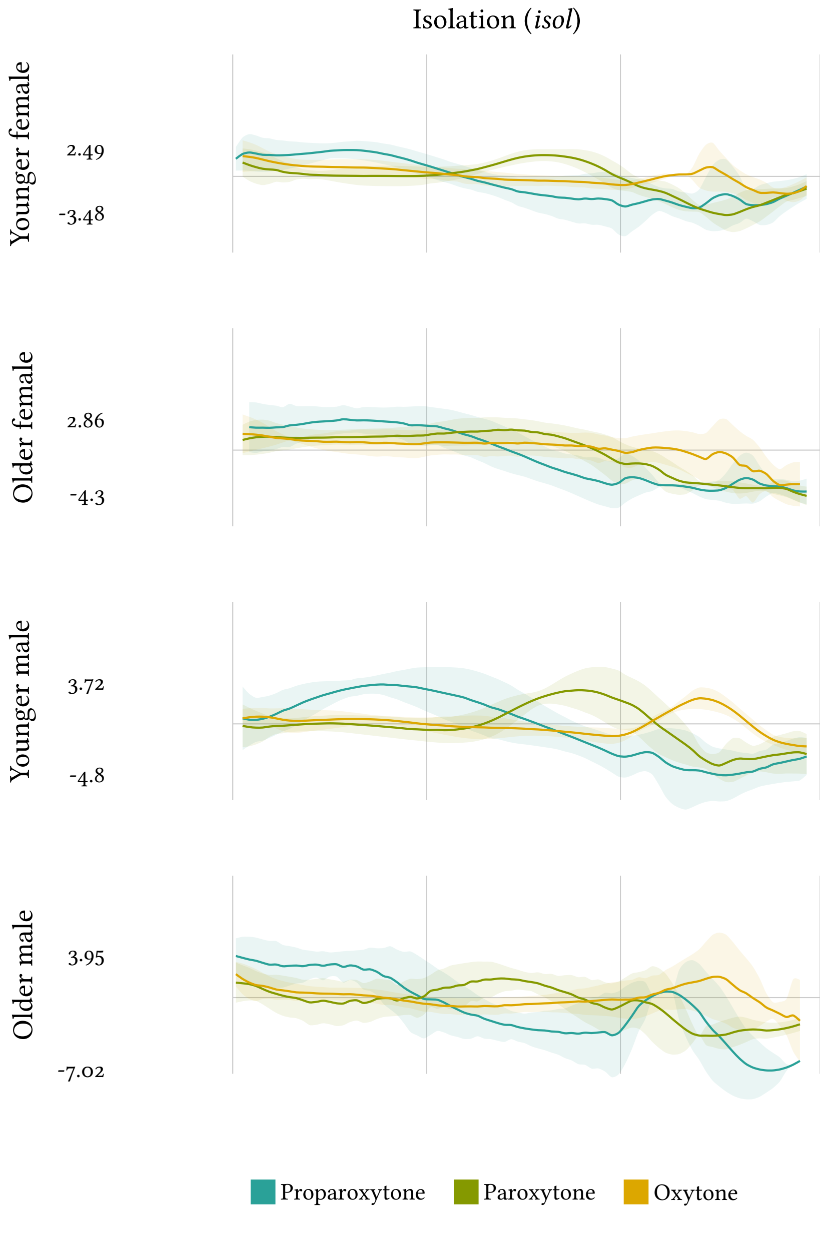


Figure C.1: Normalised F0 tracks for Spanish utterances with words in word in isolation, for each speaker. Compare with figure C.6 in page 267.

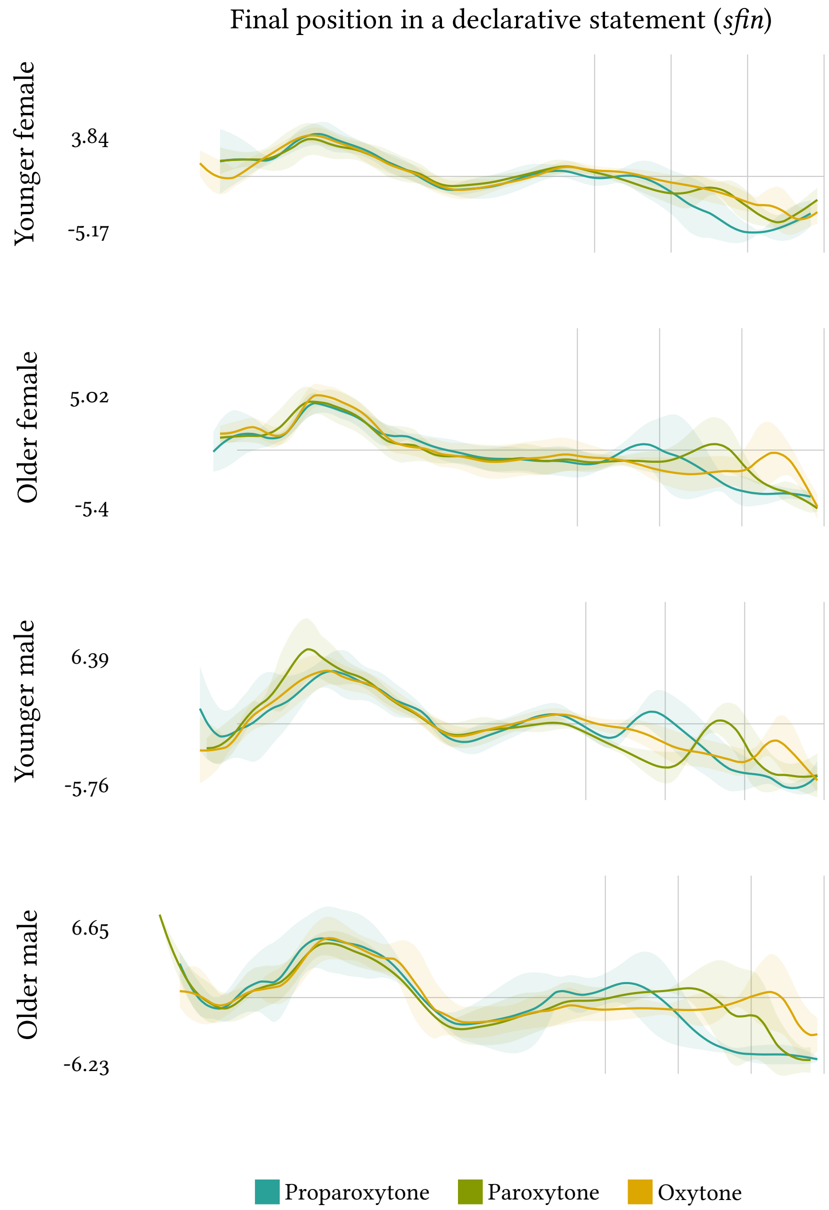


Figure C.2: Normalised F0 tracks for Spanish utterances with words in final position in a declarative statement, for each speaker. Compare with figure C.7 in page 268.

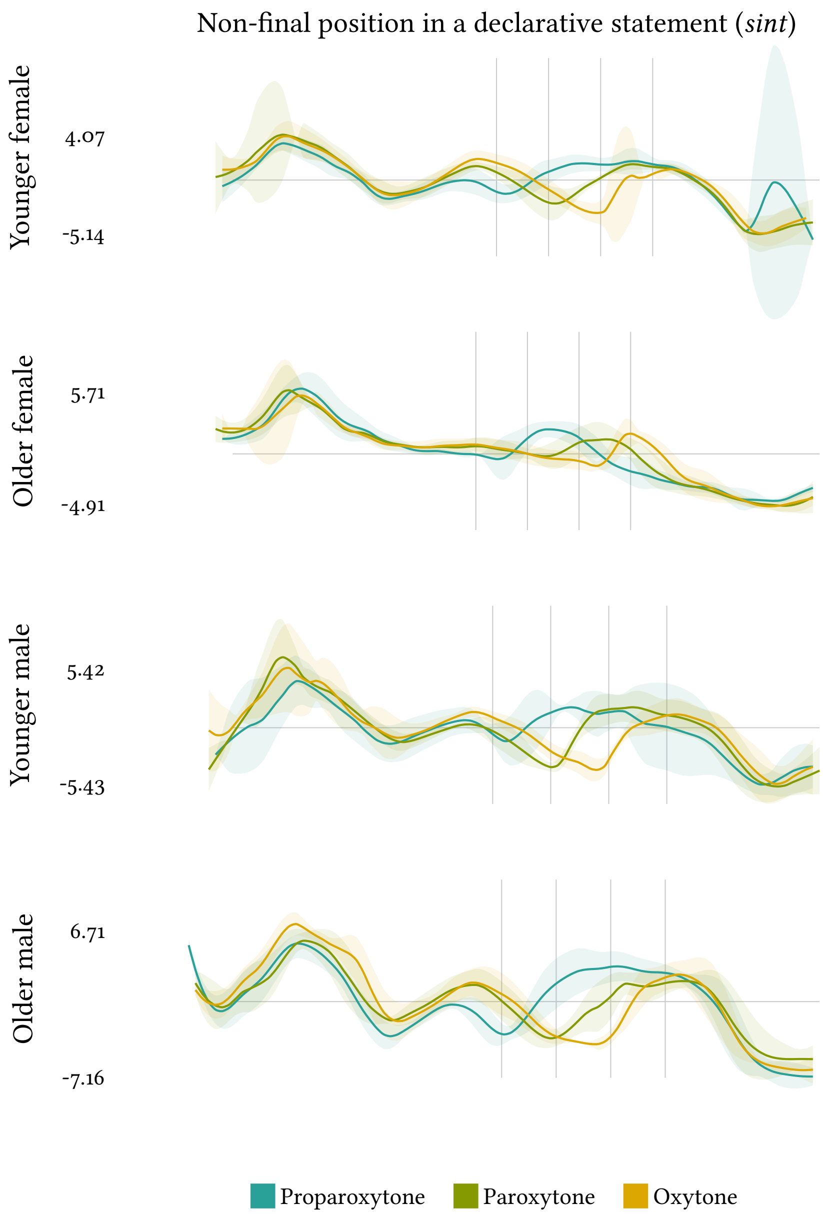


Figure C.3: Normalised F0 tracks for Spanish utterances with words in non-final position in a declarative statement, for each speaker. Compare with figure C.8 in page 269.

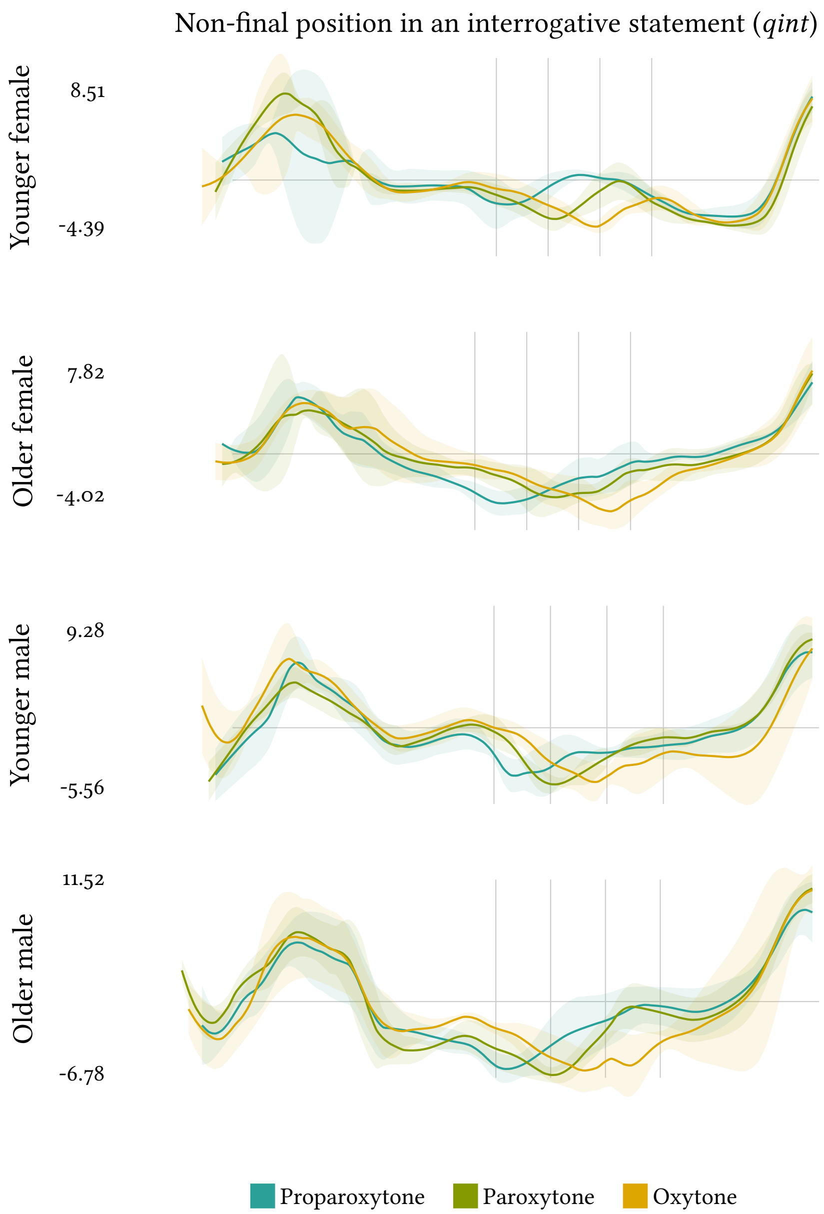


Figure C.4: Normalised F0 tracks for Spanish utterances with words in non-final position in an interrogative statement, for each speaker. Compare with figure C.9 in page 270.

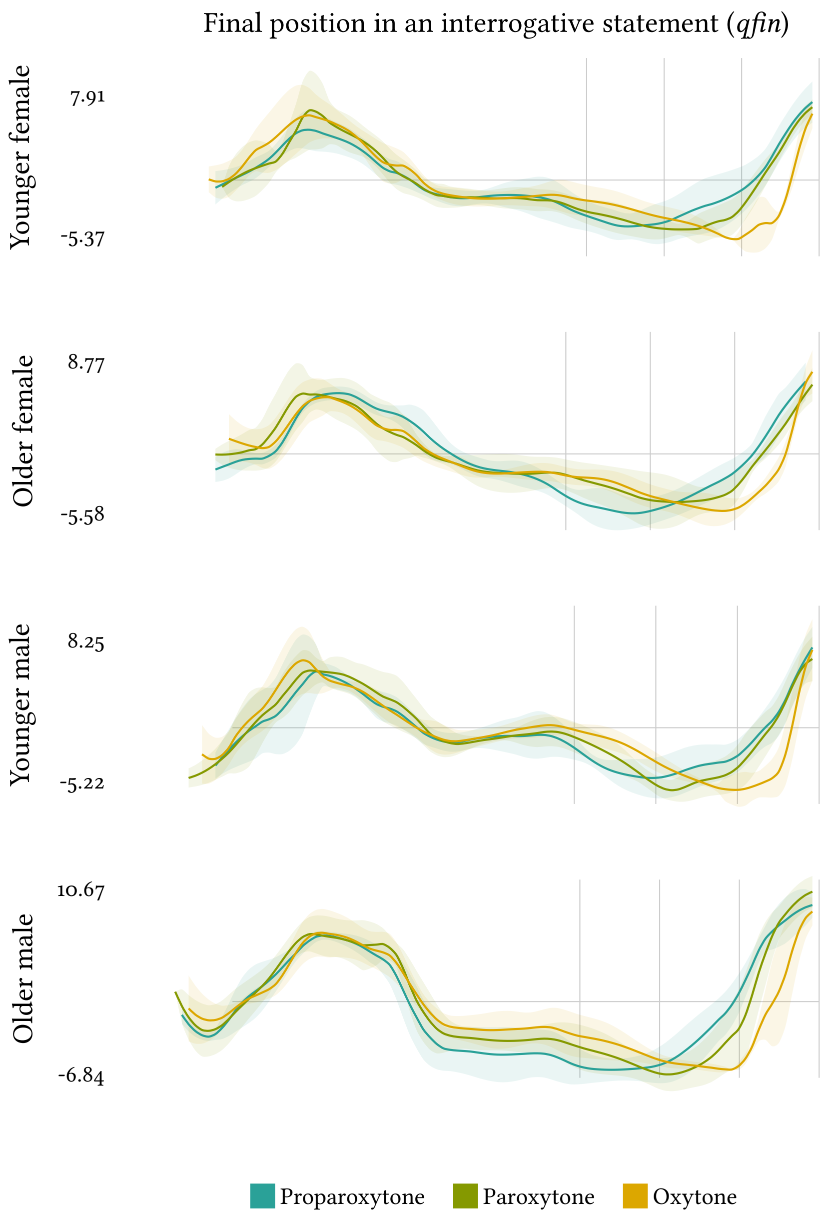


Figure C.5: Normalised F_0 tracks for Spanish utterances with words in final position in an interrogative statement, for each speaker. Compare with figure C.10 in page 271.

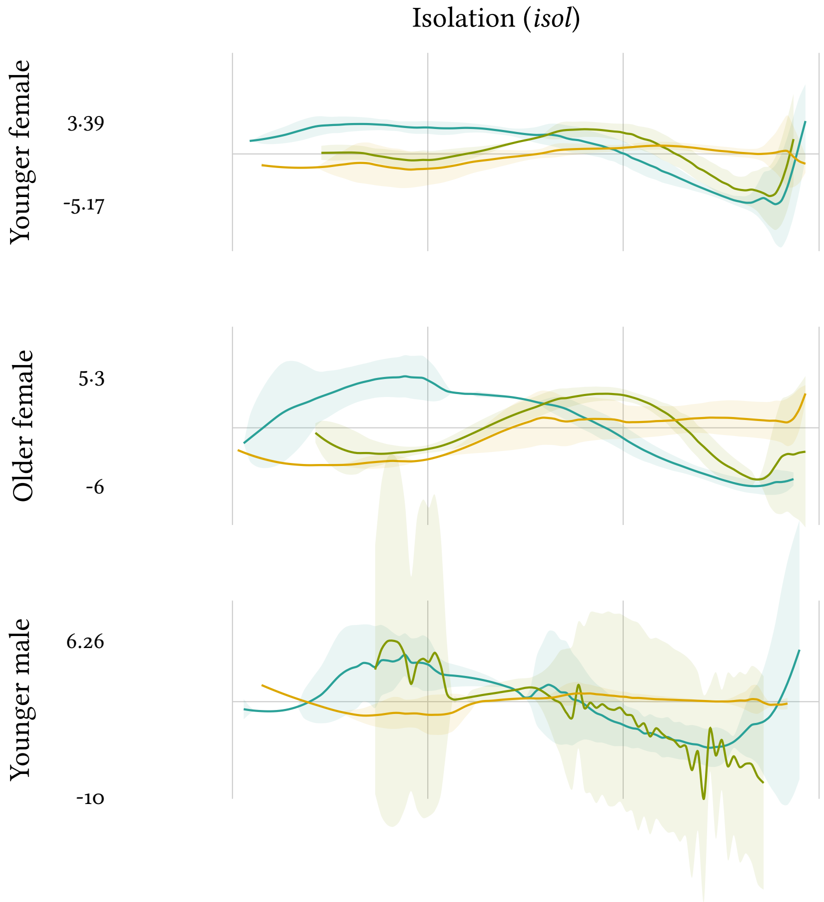


Figure C.6: Normalised F0 tracks for Japanese utterances with words in word in isolation, for each speaker. Compare with figure C.1 in page 262.

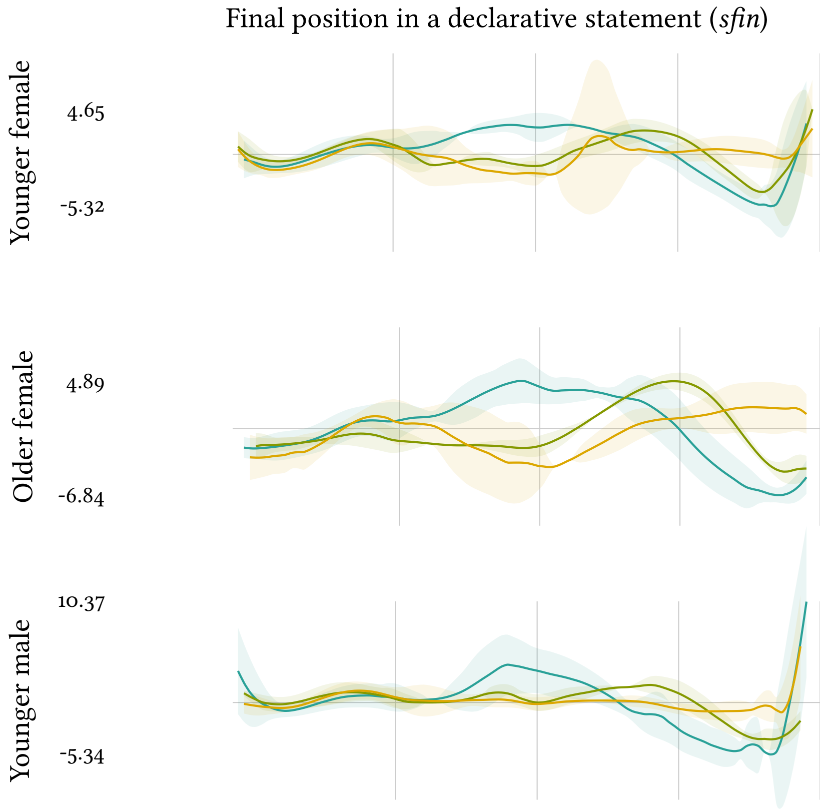


Figure C.7: Normalised F_0 tracks for Japanese utterances with words in final position in a declarative statement, for each speaker. Compare with figure C.2 in page 263.

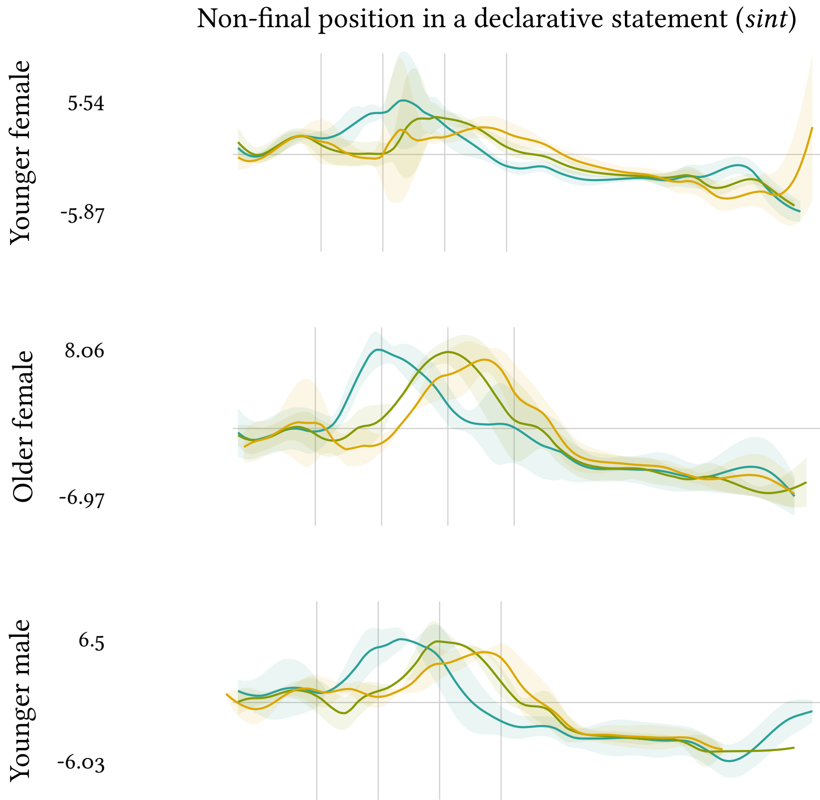


Figure C.8: Normalised F0 tracks for Japanese utterances with words in non-final position in a declarative statement, for each speaker. Compare with figure C.3 in page 264.

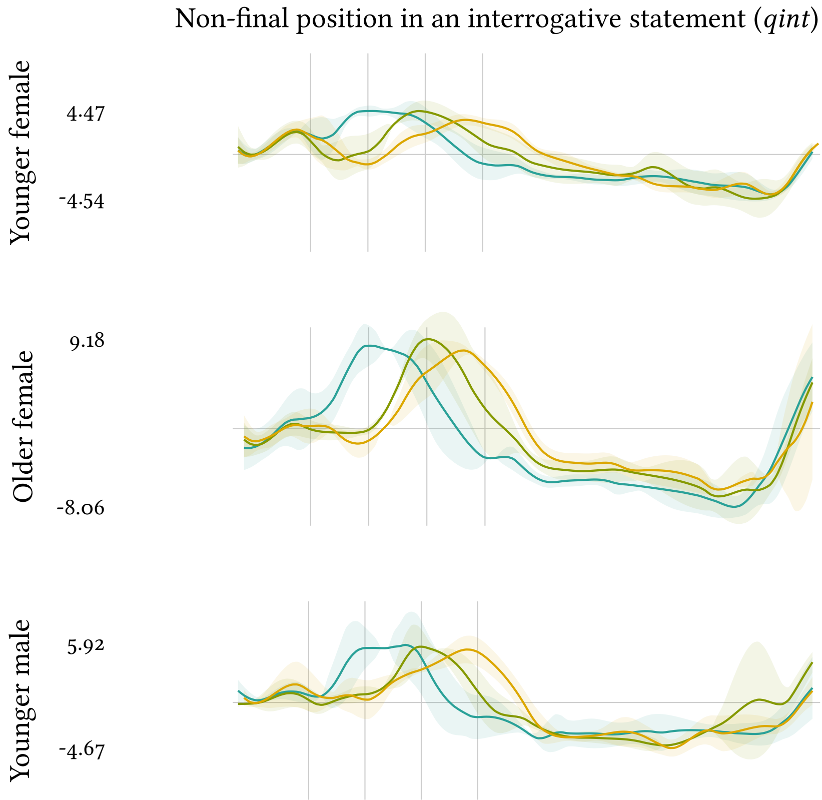


Figure C.9: Normalised F0 tracks for Japanese utterances with words in non-final position in an interrogative statement, for each speaker. Compare with figure C.4 in page 265.

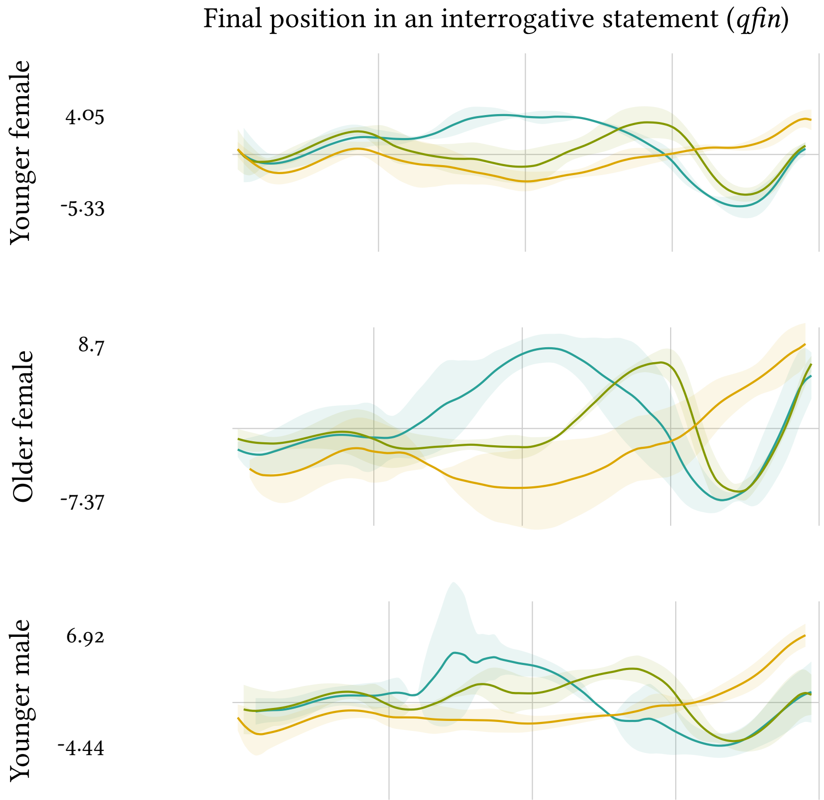


Figure C.10: Normalised F0 tracks for Japanese utterances with words in final position in an interrogative statement, for each speaker. Compare with figure C.5 in page 266.

Appendix D

Duration and intensity ratio plots

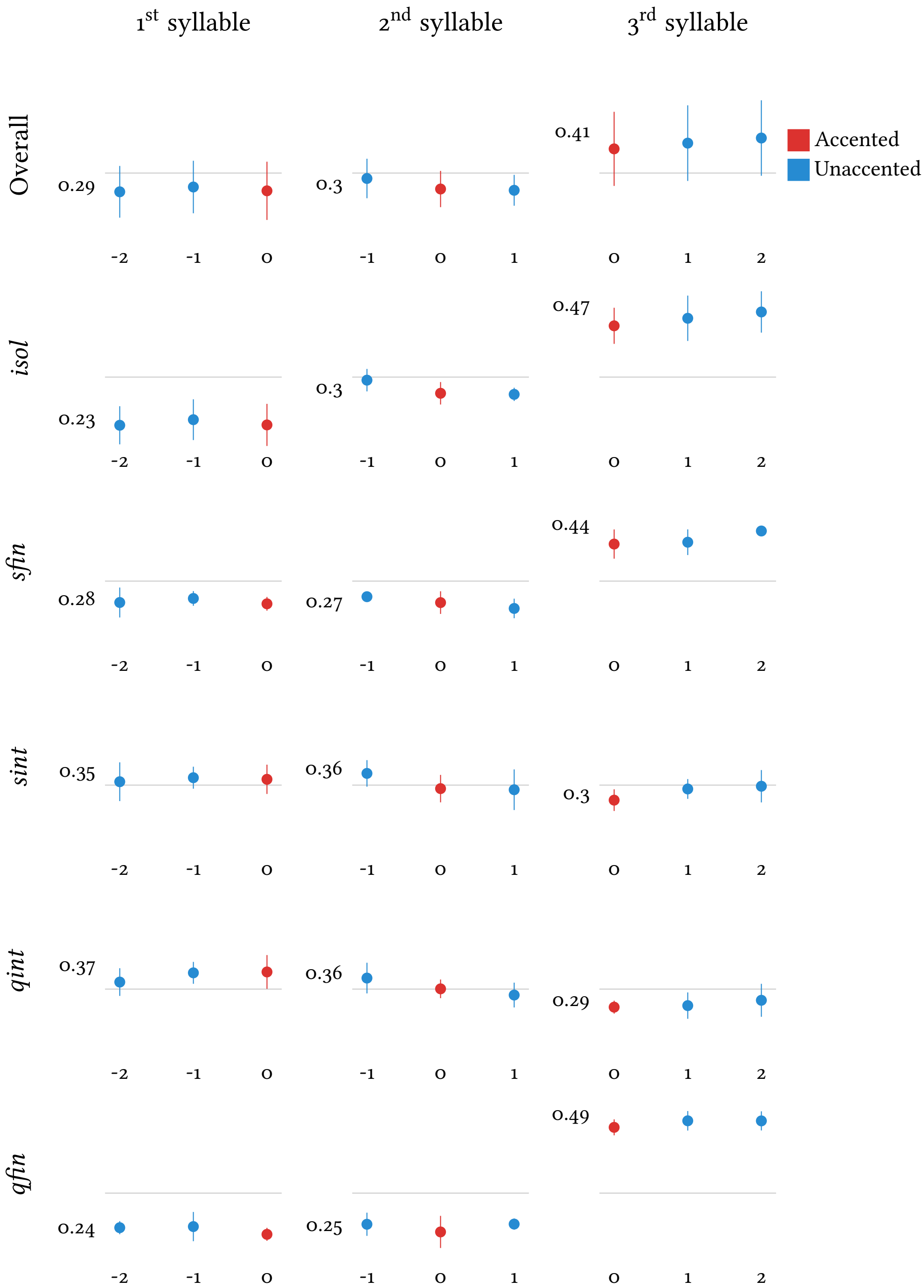


Figure D.1: Changes in the distribution of duration values per syllable for the Japanese younger female speaker.

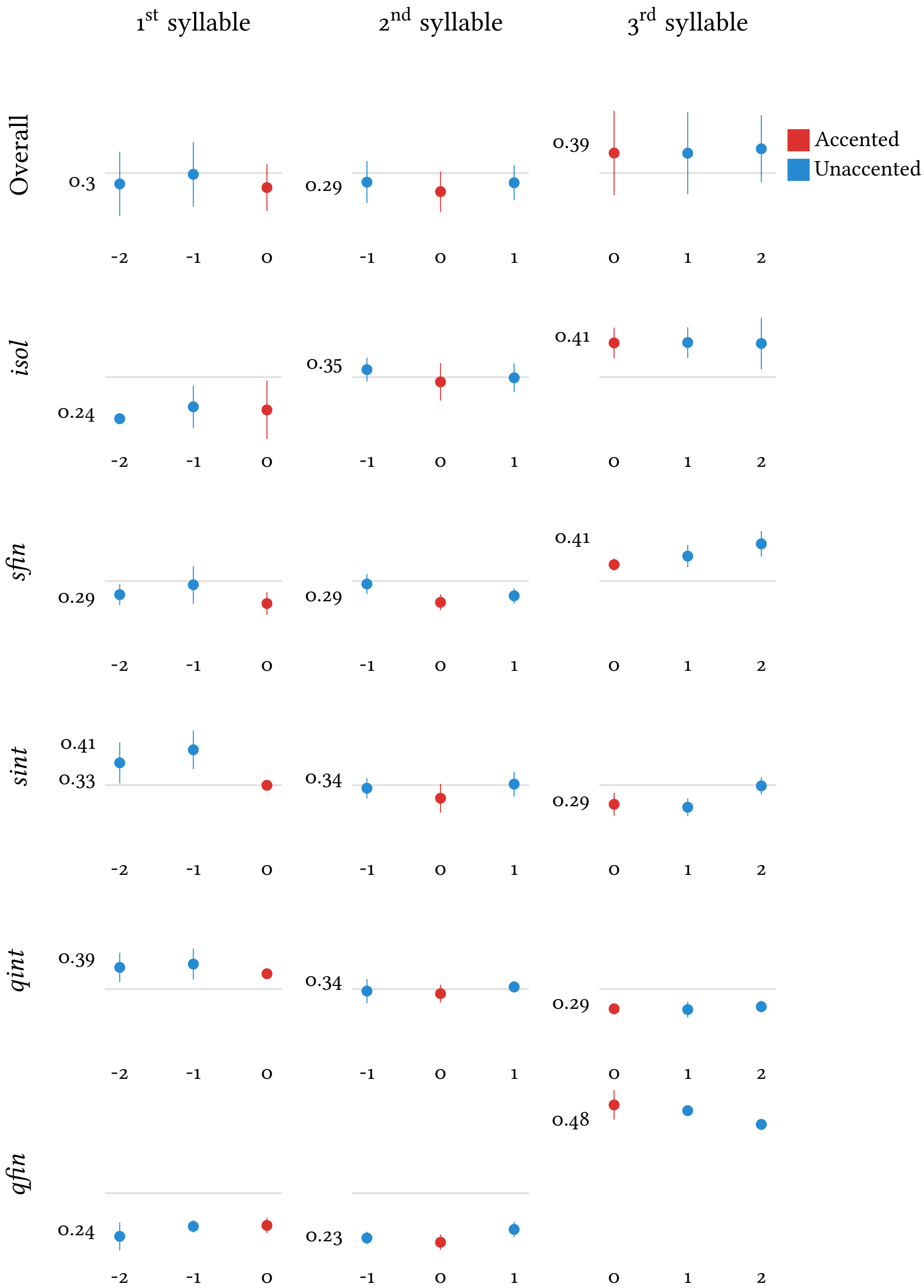


Figure D.2: Changes in the distribution of duration values per syllable for the Japanese older female speaker.

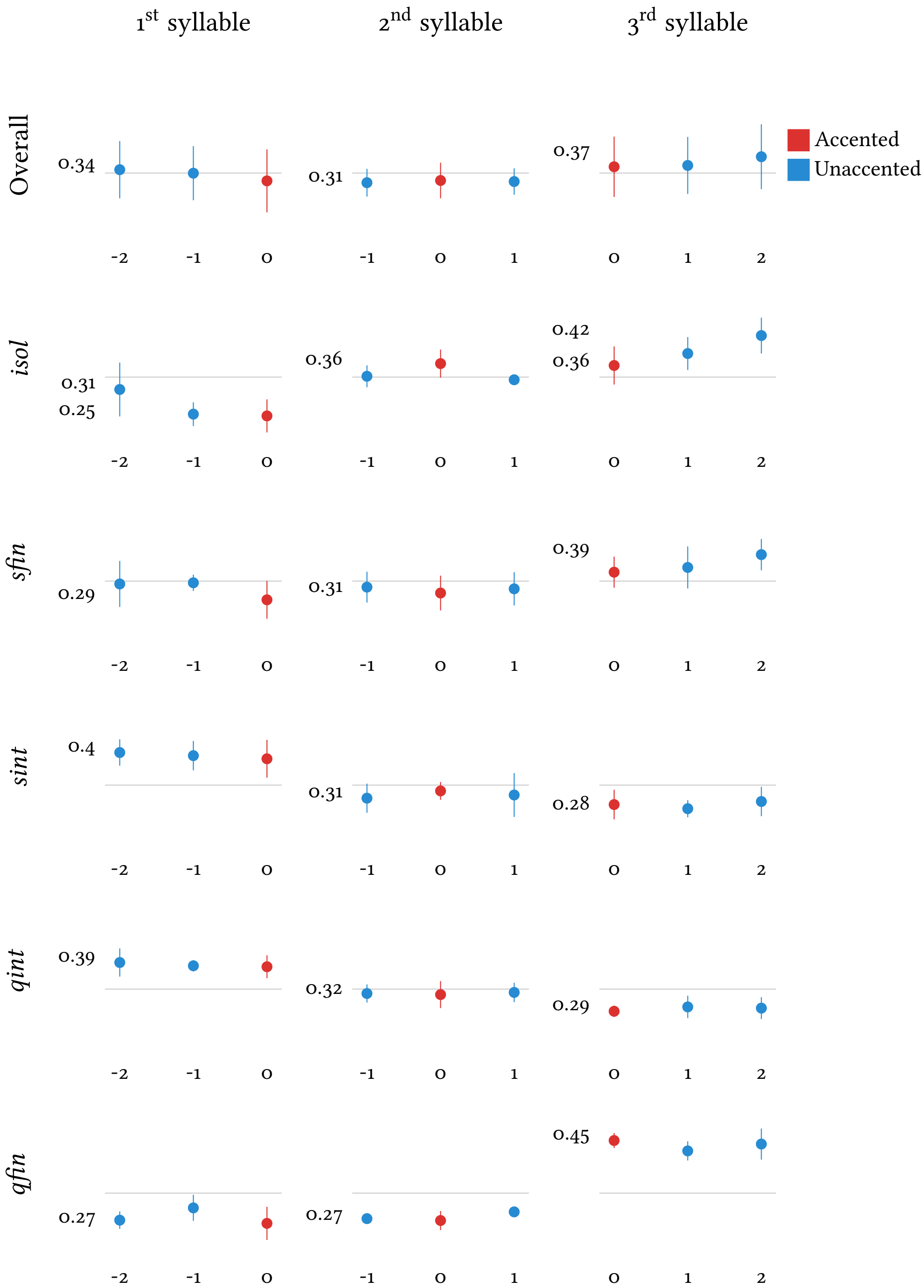


Figure D.3: Changes in the distribution of duration values per syllable for the Japanese older male speaker.

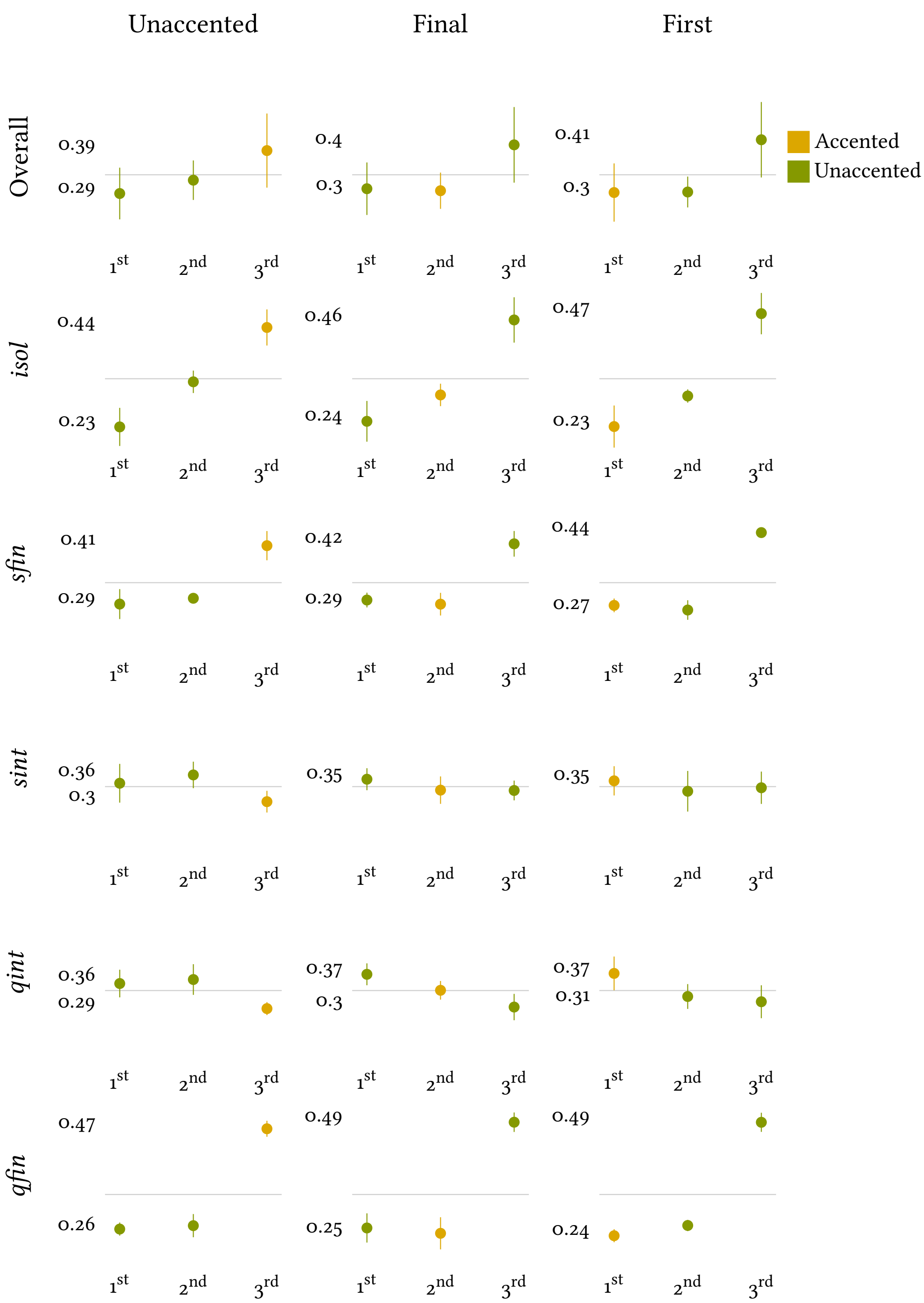


Figure D.4: Changes in the distribution of duration values per accent for the Japanese younger female speaker.

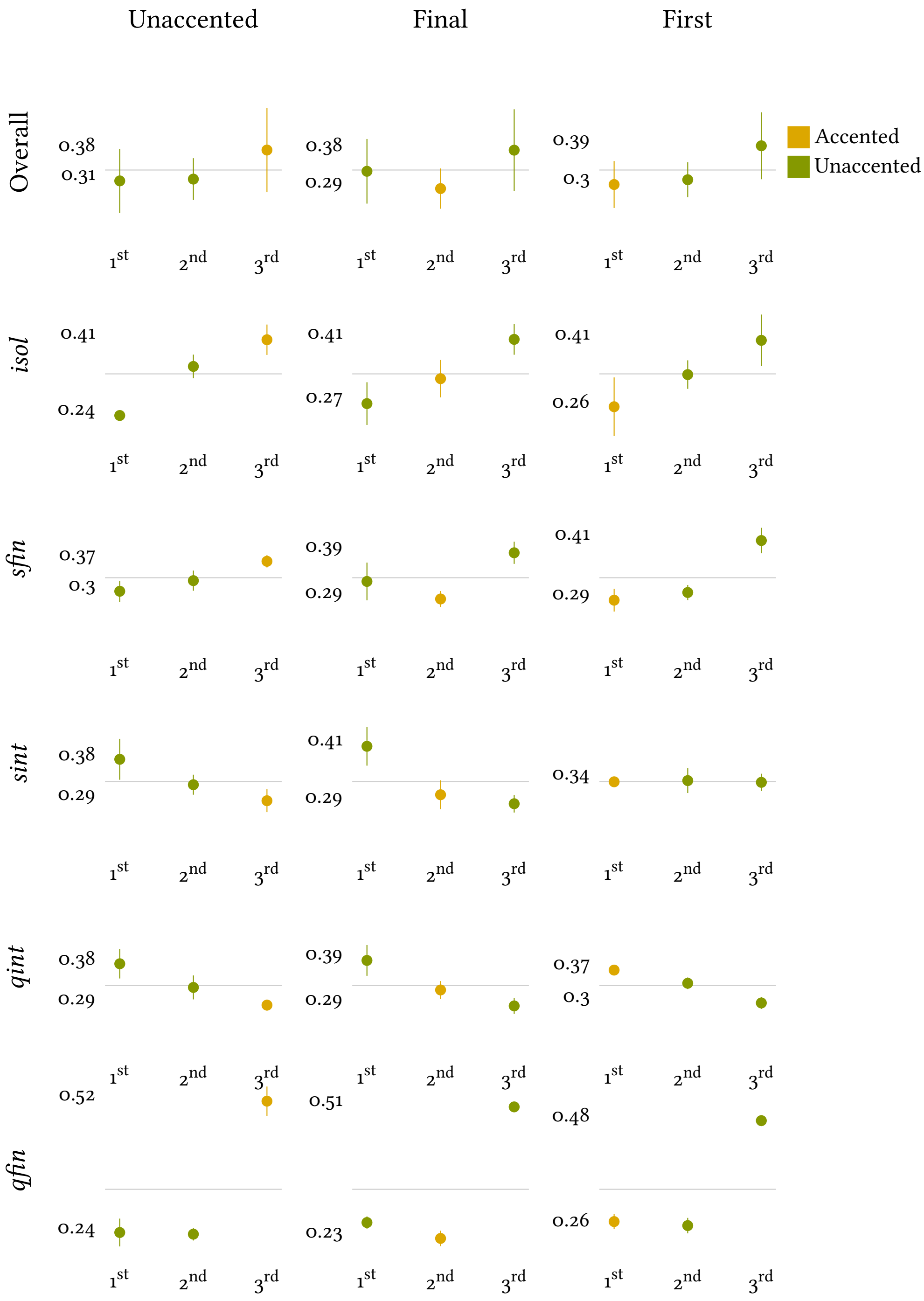


Figure D.5: Changes in the distribution of duration values per accent for the Japanese older female speaker.

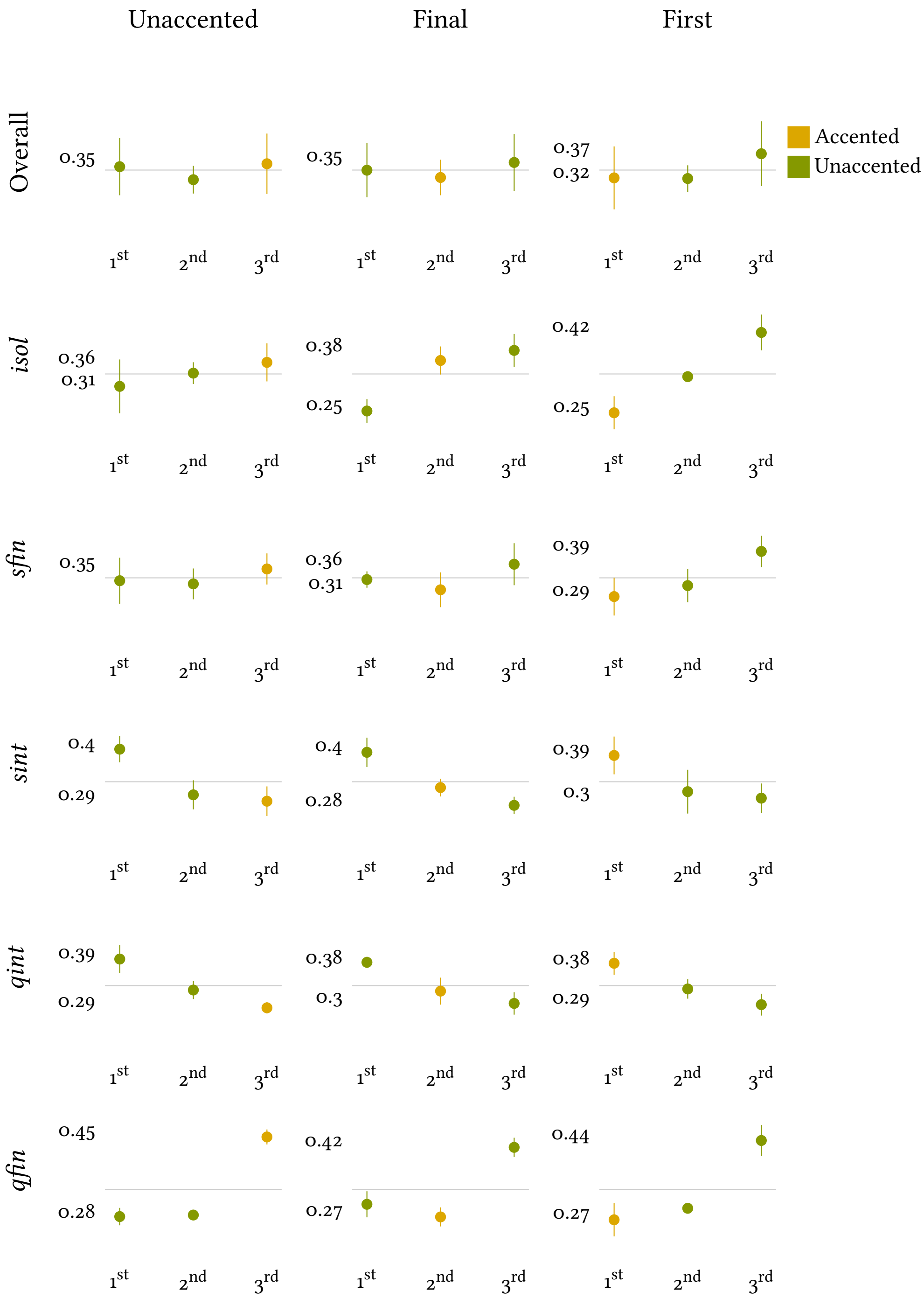


Figure D.6: Changes in the distribution of duration values per accent for the Japanese older male speaker.

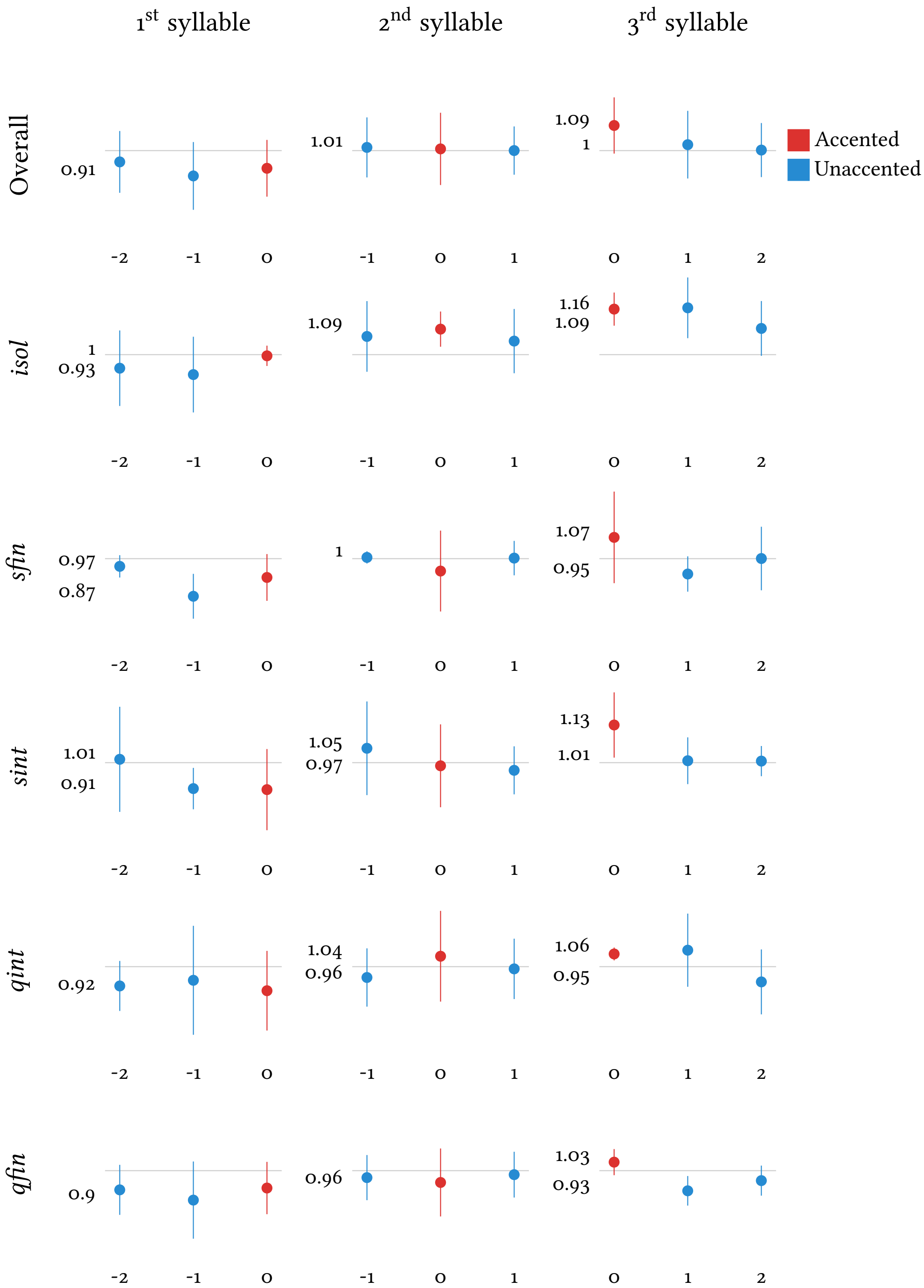


Figure D.7: Changes in the distribution of intensity values per syllable for the Japanese younger female speaker.

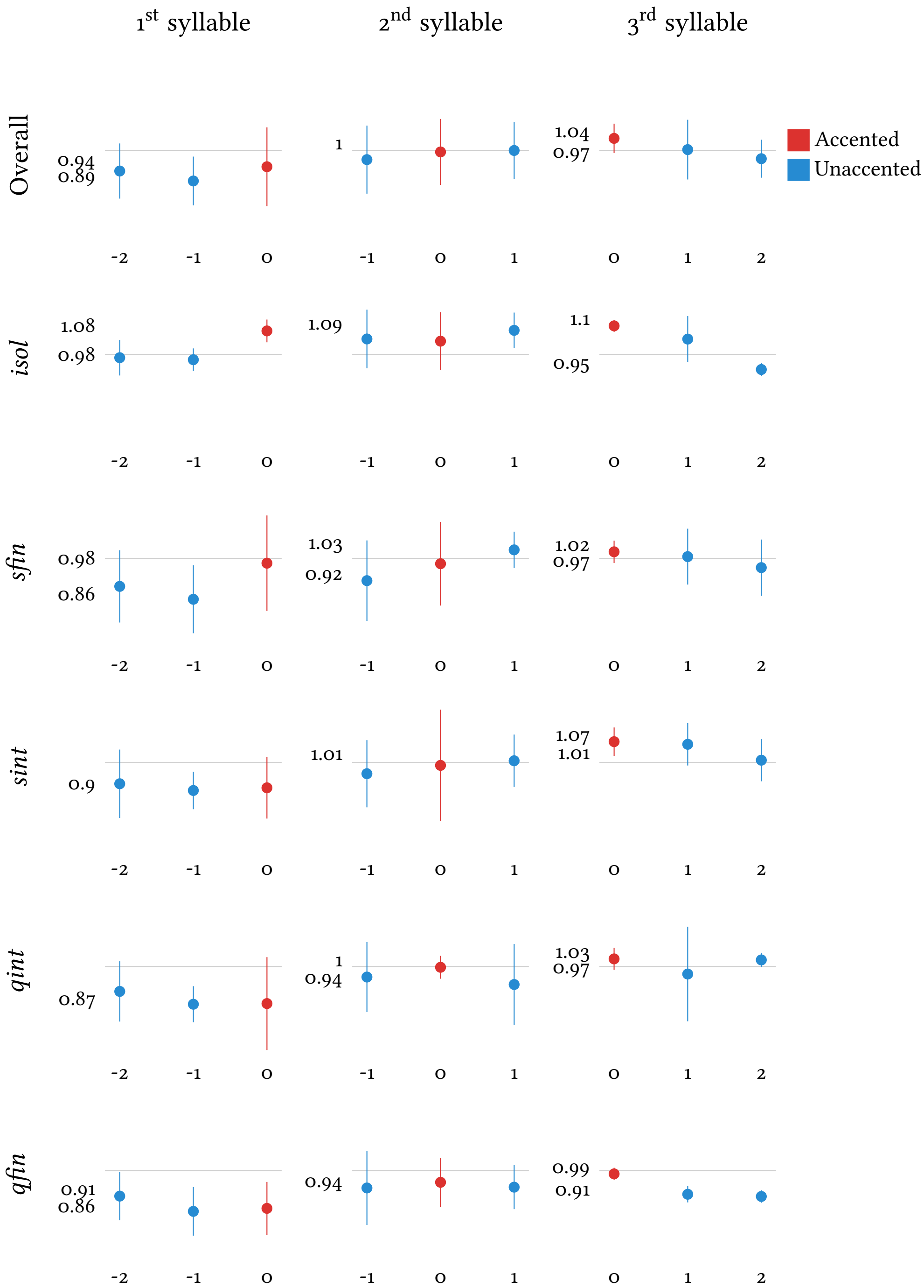


Figure D.8: Changes in the distribution of intensity values per syllable for the Japanese older female speaker.

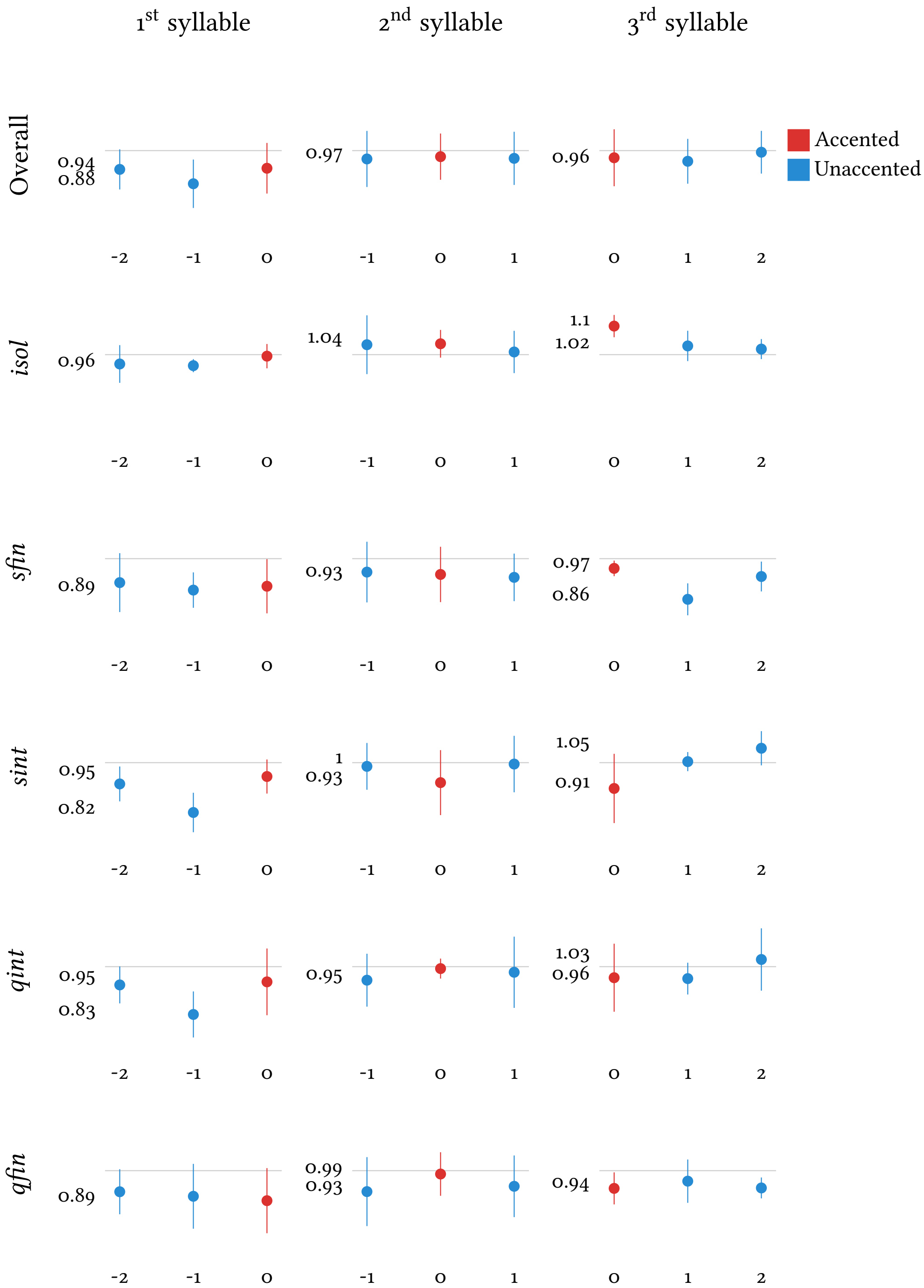


Figure D.9: Changes in the distribution of intensity values per syllable for the Japanese older male speaker.

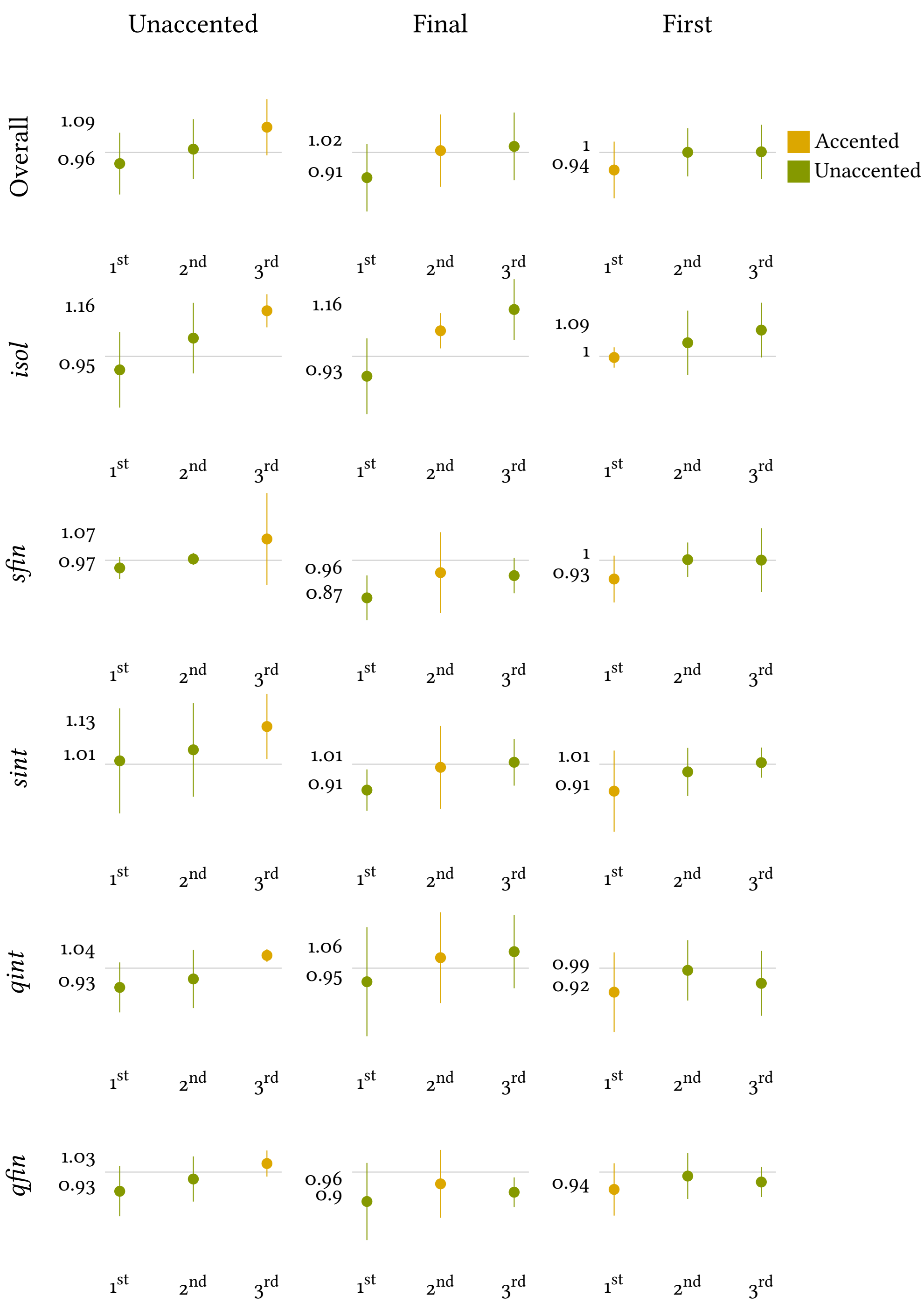


Figure D.10: Changes in the distribution of intensity values per accent for the Japanese younger female speaker.

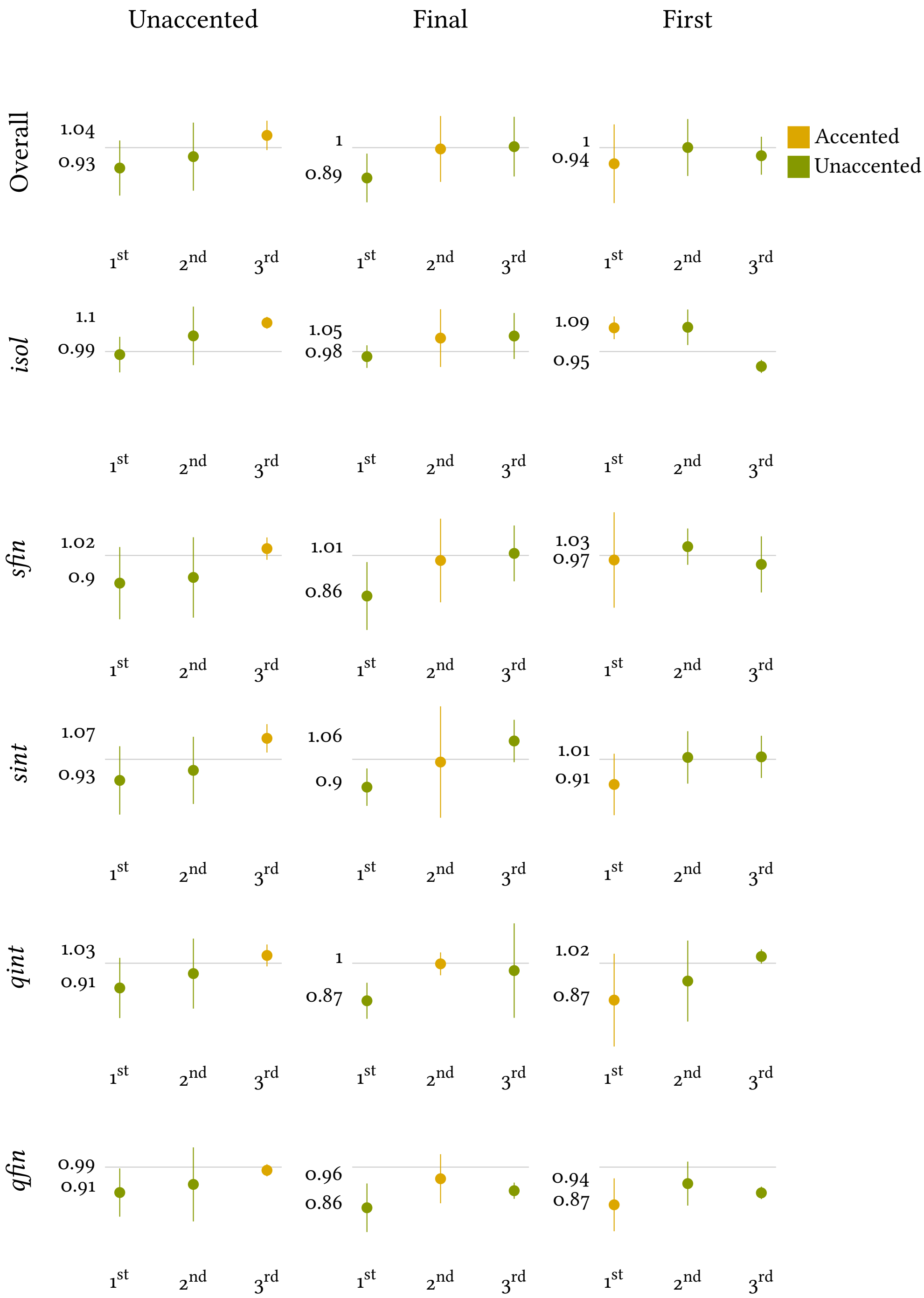


Figure D.11: Changes in the distribution of intensity values per accent for the Japanese older female speaker.

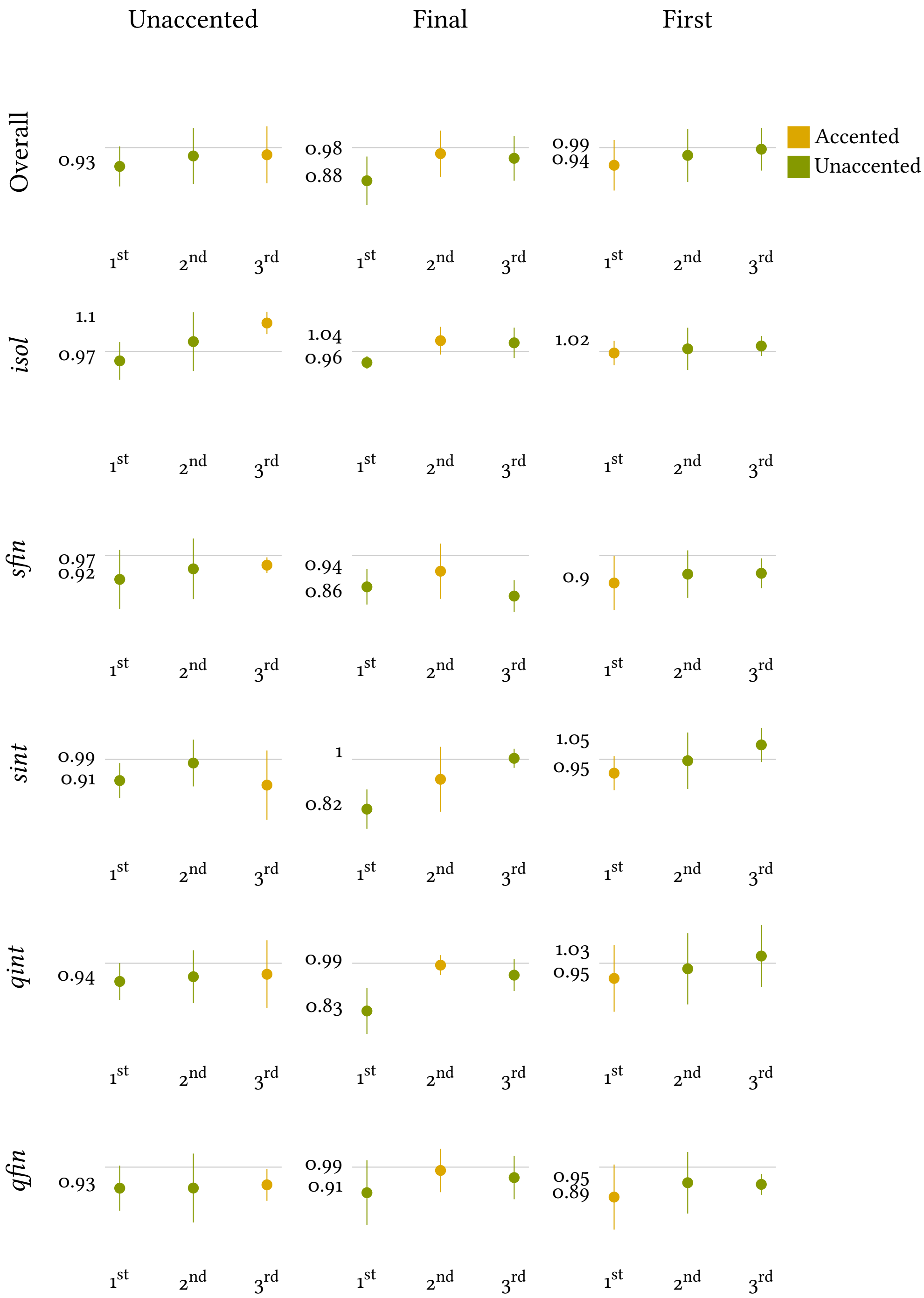


Figure D.12: Changes in the distribution of intensity values per accent for the Japanese older male speaker.

Appendix E

Duration tables per speaker

Keyword	Prom.	Speaker	Syllable						Total (s)
			1^{st}		2^{nd}		3^{rd}		
/balido/	1^{st}	CLF1	0.18	(0.35)	0.15	(0.30)	0.18	(0.36)	0.58
		CLM1	0.14	(0.30)	0.15	(0.33)	0.17	(0.37)	0.51
	2^{nd}	CLF1	0.09	(0.18)	0.19	(0.37)	0.23	(0.44)	0.52
		CLM1	0.14	(0.24)	0.21	(0.35)	0.25	(0.42)	0.55
	3^{rd}	CLF1	0.07	(0.15)	0.13	(0.28)	0.26	(0.57)	0.50
		CLM1	0.11	(0.22)	0.16	(0.31)	0.24	(0.48)	0.52
/numeɾo/	1^{st}	CLF1	0.10	(0.25)	0.16	(0.40)	0.14	(0.35)	0.52
		CLM1	0.15	(0.33)	0.14	(0.29)	0.18	(0.38)	0.46
	2^{nd}	CLF1	0.13	(0.25)	0.23	(0.43)	0.18	(0.33)	0.51
		CLM1	0.11	(0.21)	0.23	(0.44)	0.18	(0.35)	0.49
	3^{rd}	CLF1	0.08	(0.21)	0.15	(0.39)	0.16	(0.40)	0.43
		CLM1	0.16	(0.30)	0.17	(0.32)	0.20	(0.38)	0.42

(a) Spanish speakers

Keyword	Prom.	Speaker	Syllable						Total (s)
			1^{st}		2^{nd}		3^{rd}		
/haçi/	1^{st}	JPF2	0.16	(0.16)	0.29	(0.29)	0.20	(0.43)	0.41
		JPM2	0.11	(0.17)	0.29	(0.32)	0.18	(0.40)	0.54
	2^{nd}	JPF2	0.15	(0.16)	0.30	(0.31)	0.22	(0.39)	0.39
		JPM2	0.16	(0.16)	0.24	(0.37)	0.19	(0.39)	0.51
	3^{rd}	JPF2	0.16	(0.17)	0.28	(0.30)	0.25	(0.42)	0.51
		JPM2	0.15	(0.17)	0.31	(0.31)	0.21	(0.37)	0.46
/kaki/	1^{st}	JPF2	0.10	(0.16)	0.18	(0.31)	0.23	(0.51)	0.46
		JPM2	0.10	(0.16)	0.21	(0.32)	0.17	(0.47)	0.52
	2^{nd}	JPF2	0.11	(0.19)	0.20	(0.33)	0.23	(0.47)	0.53
		JPM2	0.11	(0.15)	0.22	(0.38)	0.16	(0.40)	0.46
	3^{rd}	JPF2	0.10	(0.17)	0.20	(0.35)	0.28	(0.44)	0.60
		JPM2	0.10	(0.16)	0.25	(0.36)	0.23	(0.39)	0.51

(b) Japanese speakers

Table E.1: Distribution of normalised syllable durations per speaker for Spanish (a) and Japanese (b) words used as the base of the synthetic stimuli discussed in chapter 5. Regular numbers indicate the duration of the syllable in seconds, while numbers in parentheses indicate the proportion of the word accounted by each syllable.