EFFECTS OF L1 DURATION EXPERIENCE ON JAPANESE AND SPANISH LISTENERS' PERCEPTION OF ENGLISH HIGH FRONT VOWELS

by

Geoffrey Stewart Morrison

B.Sc., University of Dundee, 1990

M.T.S., Vancouver School of Theology, 1995

THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in the Department

of

Linguistics

SIMON FRASER UNIVERSITY

July 2002

APPROVAL

Name: Geoffrey Stewart Morrison

Degree: Master of Arts

Title of Thesis: Effects of L1 Duration Experience on Japanese and Spanish Listeners' Perception of English High Front Vowels

Examining Committee:

Chair: Dr. Trude Heift, Assistant Professor

Dr. Murray J. Munro, Senior Supervisor
Associate Professor

Dr. Zita McRobbie
Associate Professor

Dr. Juan Sosa
Associate Professor

Dr. Janet F. Werker, External Examiner
Professor & Canada Research Chair
Department of Psychology, University of British Columbia

Date Approved: July 16, 02

Abstract

Canadian English has two high front vowel phonemes (tense and lax) differing in spectral properties and duration, and, allophonically, Canadian English vowels are considerably longer when preceding voiced consonants than when preceding voiceless consonants. In contrast, Western Japanese has two high front vowel phonemes (long and short), traditionally described as differing in duration only, and allophonic duration differences due to post-vocalic consonant voicing are small. Japanese listeners' perception of the English and Japanese vowel contrasts was tested using multidimensional edited speech continua. The English continuum varied in terms of vowel spectral properties, vowel duration, consonant duration, and sentence duration. The Japanese continuum had the same duration properties but no spectral dimension. The Japanese listeners' English tense-lax and Japanese long-short categorical boundaries were found to be in the same location. This was consistent with the hypothesis that Japanese listeners assimilate the English tense vowel before a voiced consonant to the Japanese long vowel category; assimilate the English tense vowel before a voiceless consonant, and the English lax vowel before either a voiced or voiceless consonant to the Japanese short vowel category; and identify vowels assimilated to the Japanese long category as English tense vowels, and vowels assimilated to the Japanese short category as English lax vowels. Mexican Spanish listeners were also tested on the English continuum. Spanish has only one high front vowel and lacks any duration contrasts. The Spanish listeners did not have a categorical boundary between the two English vowels, but made some use of vowel duration, consistent with a category-goodness assimilation (Best's Perceptual Assimilation Model) in which the English tense vowel was a good match and the English lax vowel was a poorer match for the Spanish vowel. Longitudinal results (one versus six months residence in Canada) were compatible with Flege's Speech Learning Model: There was no change in the Japanese listeners' perception of the English vowels, consistent with English vowels being perceived as "similar" to Japanese vowels, and the formation of diaphone categories. Spanish listeners developed categorical perception based on duration or spectral properties, consistent with the formation of a "new" category for the English lax vowel.

Посвящена Илане

וַיִּלְכֹּ֥ד גִּלְעָ֛ד אֶת־מַעְבְּר֥וֹת הַיַּרְדֵּ֖ן לְאֶפְרָ֑יִם וְֽהָיָה֩ כִּ֨י יֹאמְר֜וּ פְּלִיטֵ֤י
אֶפְרַ֙יִם֙ אֶֽעֱבֹ֔רָה וַיֹּ֨אמְרוּ ל֧וֹ אַנְשֵֽׁי־גִלְעָ֛ד הַאֶפְרָתִ֥י אַ֖תָּה וַיֹּ֥אמֶר ׀ לֹֽא׃
וַיֹּ֣אמְרוּ ל֡וֹ אֱמָר־נָ֣א שִׁבֹּ֠לֶת וַיֹּ֨אמֶר סִבֹּ֜לֶת וְלֹ֤א יָכִין֙ לְדַבֵּ֣ר כֵּ֔ן
וַיֹּאחֲז֣וּ אוֹת֔וֹ וַיִּשְׁחָט֖וּהוּ אֶל־מַעְבְּר֣וֹת הַיַּרְדֵּ֑ן וַיִּפֹּ֞ל בָּעֵ֤ת הַהִיא֙
מֵֽאֶפְרַ֔יִם אַרְבָּעִ֥ים וּשְׁנַ֖יִם אָֽלֶף׃

Judges 12:5-6

Acknowledgements

I would like to express my thanks to my committee members Dr. Zita McRobbie, Dr. Juan Sosa, and especially Dr. Murray J. Munro for extensive feedback on earlier versions of this thesis. I am indebted to numerous other scholars with whom I have discussed my research and who have provided me with copies of their published and unpublished papers. My thanks also to the anonymous volunteers who participated in the research and to Miriam Y. Barranco Cedillo, Kenneth Gehrs, Lorna Marie Fadden, Kenichi Okuma, Cheryl Wilson, and Daniel Zimmerman. Finally, a special thank you to my parents for their continued support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis will investigate the influence of the Japanese long-short phonemic contrast on Japanese listeners' perception of the tense-lax phonemic contrast for high front vowels in Canadian English. It investigates the claim that L2 sounds are perceived at a context sensitive allophonic level (Brière, 1966; Flege, 1995). In particular, it examines the effect due to phonemic voicing of a consonant on the perception of the preceding vowel. Since vowel duration is a cue to phonemic voicing in a post-vocalic consonant in English (Chen, 1970), and the Japanese long-short vowel contrast is cued by vowel duration, the state of phonemic voicing in the post-vocalic consonant has the potential to affect Japanese listeners' perception of English vowels.

A specific hypothesis to be tested is that Japanese listeners identify English /i/ and /ɪ/ according to the same duration criteria they use to identify Japanese /iː/ and /i/; and that this results in English /i/ before a voiced consonant being assimilated to Japanese long /iː/ but English /i/ before a voiceless consonant being assimilated to Japanese short /i/; English /ɪ/ before both a voiced and voiceless consonant being assimilated to Japanese short /i/; and English vowels assimilated to Japanese long /iː/ being identified as English /i/, and English vowels assimilated to Japanese short /i/ being identified as English /ɪ/. This hypothesis was developed by Morrison (2002a). In this thesis, a comparison of Japanese listeners' Japanese /iː/-/i/ and English /i/-/ɪ/ categorical boundaries will be made to determine whether they are identical and thus consistent with this hypothesis.

Perceptual boundaries between vowel phonemes will be measured using multidimensional edited-speech continua. Potential acoustic cues to be tested in the English continuum will be vowel spectral properties, vowel duration, consonant closure duration, and speaking rate. The English stimulus consonants will be phonetically voiceless, (i.e. have silent consonant closures) so that only vowel duration and consonant duration are expected to be potential cues for phonemic consonant voicing. Consonant duration may also have a contrastive effect on the perception of vowel duration (Kluender, Diehl, & Wright, 1988). In addition, speaking rate cues recoverable from the larger rhythmic structure of the sentence may have an influence on the perception of vowel and consonant duration (Strange et al., 1998). Discriminant analyses will be used to construct statistical models of the listeners' stimulus-identification patterns. These models will provide a means of quantifying the position and the sharpness of the categorical boundaries between vowel phonemes.

If Japanese listeners have a duration based categorical boundary between English /i/ and /ɪ/, this could be due to transfer from the Japanese long-short vowel contrast, or could be due to a general L2 learning strategy applied by all learners of English irrespective of the existence of a duration contrast in their L1. In order to support the claim that the Japanese listeners' results are influenced by the Japanese long-short vowel duration contrast, a group of Spanish listeners will also be tested: Japanese and Spanish have similar five-vowel-quality systems, but Spanish, as opposed to Japanese, does not have a long-short contrast. Because of this difference, the Spanish listeners' results are predicted to differ from those of the Japanese listeners. Spanish listeners are expected to have difficulty perceiving the difference between English /i/ and /ɪ/, but, in line with Bohn's (1995) Desensitisation Hypothesis, they are expected to make non-categorical use of vowel duration.

The research in this thesis will also be longitudinal, allowing for an investigation of Flege's (1995) Speech Learning Model hypotheses regarding the development of "new" and "similar" categories. Since the Japanese listeners are expected to perceive the English vowels as similar to Japanese vowels, they are predicted to form diaphone categories whose eventual properties will be intermediate between the properties of the L1 and L2 vowels contributing to those categories. Hence, Japanese listeners are predicted to show only small changes in perception over the course of the longitudinal study. In contrast, Spanish listeners are expected to assimilate both English /i/ and /ɪ/ to Spanish /i/, but English /ɪ/ is expected to be a poorer match for the Japanese category than English /i/[1] (Flege, 1991b; Flege, Munro, & Fox, 1994). It is therefore predicted that the Spanish listeners will form a new category for English /ɪ/, whose properties may eventually be the same as for a native English speaker's English /ɪ/. Hence, Spanish listeners are predicted to show large changes in perception over the course of the longitudinal study.

[1]A category-goodness assimilation in terms of Best's (1995a) Perceptual Assimilation Model.

Chapter 2

Literature Review

In this chapter, Section 2.1 presents a description of current models of cross-language speech perception. This is followed in Section 2.2 by a review of studies dealing with vowel and consonant duration as cues to post-vocalic consonant voicing. Section 2.3 reviews studies dealing with the perception and production of high front vowels in English, Japanese, and Spanish. This section focuses especially on vowel duration effects in cross-language studies. Section 2.4 provides a summary of issues from the literature review which are central to this thesis. Finally, Section 2.5 presents a set of hypotheses to be tested in the remainder of the thesis.

2.1 Models for Cross-language Speech Perception and Production

2.1.1 Overview

The most influential paradigms in current cross-language speech perception research are Best's *Perceptual Assimilation Model* (PAM) (Best 1994, 1995a, 1995b) and Flege's *Speech Learning Model* (SLM) (Flege,1981, 1987, 1991a, 1992, 1995). The PAM is primarily concerned with how adult listeners will perceive the sounds of an unfamiliar language, and with the development of first language speech perception in infants. In contrast, The SLM is concerned with the development of perceptual categories as adult second language learners become more familiar with the second language (L2). The SLM also makes predictions regarding the production of L2 speech sounds. An additional model of particular interest to cross-language vowel perception is Kuhl and colleagues' *Native Language Magnet* model (NLM) (Kuhl, 1991; Kuhl et al., 1992; Iverson & Kuhl, 1995; Kuhl & Iverson, 1995). The *Perceptual Magnet Effect* in the NLM results in vowels being perceived as closer to their prototypes than might be suggested by acoustic properties alone. This implies that there is "reduced discrimination sensitivity" (Kuhl & Iverson, 1995, p131) in the area around a native, or first language (L1), vowel prototype, and that this hinders the listener's ability to distinguish two L2 vowels should both vowels be close to a single L1 vowel. Another hypothesis related to listeners' difficulty with L2 spectral differences is Bohn's *Desensitisation Hypothesis* (Bohn, 1995). The *Desensitisation Hypothesis* states that if a listener's L1 does not have multiple vowels in a particular part of the vowel space, the listener will not be sensitive to the spectral properties of multiple L2 vowels in that part of the vowel space, and will instead use duration properties (if available) to distinguish those L2 vowels. Bohn claims that this will be so whether or not phonemic duration is used in the L1. The remainder of section 2.1 presents a description of aspects of these models and hypotheses that are relevant to the present study.

2.1.2 The Perceptual Assimilation Model

According to the Perceptual Assimilation Model (PAM) (Best 1994, 1995a, 1995b), a non-native speech sound can be either assimilated to a native speech sound, assimilated as a speech sound but not assigned to any native category, or not assimilated as a speech sound (perceived as a non-speech sound). A non-native speech sound that is assimilated to a native category will either be a good, acceptable but not ideal, or notably deviant example of that category (Best, 1995a, 1995b). Best (1995a) argues that almost all non-native speech sounds will be assimilated as speech sounds, and only a few non-native sounds, such as clicks for speakers of non-click languages, are likely to not be assimilated as speech sounds (Best, McRoberts, & Sithole, 1988). Assimilation can be assessed by tests which require the listeners to categorise and give category-goodness ratings to non-native sounds (Best, 1995a).

The PAM also predicts that a pair of non-native speech sounds will be distinguished from one another according to one of the following set of assimilation patterns (Best, 1994, 1995a, 1995b):

Two-category assimilation: If each sound is assimilated to a different native category, discrimination is predicted to be excellent. Several studies have confirmed this prediction: Best & Strange (1992) found evidence that Japanese listeners assimilated an English syllable-initial /w/-/j/ continuum to Japanese /w/ and /j/, and discriminated at the same level as native English listeners. Polka (1992) found that some Farsi listeners assimilated Salish uvular versus velar ejectives (/q'/ versus /k'/) to voiced uvular and velar plosives in Farsi (/ɢ/ and /g/). These listeners reached a native-like perception criterion. A study by Best, Faber, & Levitt (cited in Best, 1995b) found that English listeners assimilated French /y/ versus /œ/ to English /u/ versus /ʊ/, and assimilated French /œ/ versus /ə/ to English /ʊ/ versus /ʌ/. In both cases discrimination was very good. Likewise, Norwegian high front in-rounded versus unrounded vowels were assimilated to English /ʊ/ versus /i/, and discrimination was perfect. However, there appear to be some exceptions to the excellent discrimination predicted for two-category assimilation: Polka (1991) found that English listeners assimilated initial position Hindi voiceless-unaspirated dental versus retroflex plosives (/t̪/ versus /ʈ/) to English /ð/ versus /d/ and discrimination was good. However, Hindi breathy-voiced dental versus retroflex plosives (/d̪ʰ/ versus /ɖʰ/) were assimilated to English /d/ versus /t/ and discrimination was moderate.

Single-category assimilation: If both sounds are equally well assimilated to a single native category, then discrimination is predicted to be poor. Several studies have found evidence in support of this prediction: Best & Strange (1992) found evidence that Japanese listeners assimilated an English syllable-initial /r/-/l/ continuum to Japanese /w/ and discrimination was poor. Polka (1991) found that English listeners assimilated initial position Hindi prevoiced dental and retroflex plosives (/d̪/ and /ɖ/) to English /d/ and discrimination was at chance levels. Hindi voiceless-aspirated dental and retroflex plosives (/t̪ʰ/ and /ʈʰ/) were both assimilated to English /t/ and discrimination was only slightly above chance levels. Best, Faber, & Levitt (cited in Best, 1995b) found that English listeners assimilated Norwegian high front out-rounded and unrounded vowels to English /i/ and discrimination was poor.

Category-goodness difference: If both sounds are assimilated to a single native category but differ in the degree to which they are good examples of that category, then discrimination is predicted to be moderate to very good. Several studies have found evidence in support of this prediction: Best & Strange (1992) found evidence that Japanese listeners assimilated an English syllable-initial /w/-/r/ continuum to Japanese /w/ but that English /w/ was a better match and discrimination was good. Polka (1992) found that some English listeners assimilated Farsi /ɢ/ and /g/ to English /g/, and those who reported that the Farsi /g/ was a better match reached a native-like perception criterion. Polka (1995) found that English listeners assimilated German /y/ and /u/ to English /u/, with German /u/ as a good match and German /y/ a poorer match, and discrimination was very good.

Both uncategorisable: If both sounds are assimilated as speech sounds but not assigned to native categories then discrimination is predicted to vary depending on how similar the sounds are to each other and to native sounds. Several studies have found evidence in support of the concept of uncategorisable sounds: Polka (1992) found that some English listeners assimilated Salish /q’/ and /k’/ to speech sounds but reported distinguishing the sounds on the basis of differences in the following vowel or on the basis of acoustic differences such as differences in duration. The use of such acoustic cues suggested that the sounds were not categorised. These listeners reached the native-like perception criterion used in Polka (1992). Guion et al. (2000) conducted classification and goodness rating tests, and discrimination tests. In contrast to Best & Strange (1992) (see above under single-category assimilation), they found that both English /r/[ɹ] and English /l/ were poor fits for any Japanese sound and discrimination was poor. Individual listeners assigned the same English sounds to different Japanese categories on different trials. The poor goodness of fit suggested that these sounds were uncategorised.

Uncategorised versus categorised: If one sound is assimilated to a native category and the other assimilated as a speech sound, but not assigned to a native category, then discrimination is predicted to be very good. Several studies have found evidence in support of this prediction: Some English listeners in Polka (1992) reported that for Farsi/ɢ/ versus /ɡ/, the Farsi /ɡ/ sounded like an English /ɡ/ and Farsi /ɢ/ sounded like a low-back vowel without an initial consonant. Best (1995b) suggested that this can be interpreted as evidence of a categorised versus uncategorised distinction. Discrimination was above chance. In classification and rating tests, Guion et al. (2000) found that English /w/ was a fair fit for Japanese /w/, whilst English /r/ was a poor fit for any Japanese category; discrimination was moderate to good. In the same study, English /s/ was a good fit for Japanese /s/ whilst English /θ/ was a poor fit, falling between Japanese /s/ and the [ɸ] allophone of /h/. Contrary to predictions, discrimination was poor. Guion et al. suggested that a revision of the PAM may be needed to account for this result, obtained when an uncategorised sound was close in phonological space to a categorised sound.

Nonassimilable: If neither sound is assimilated as a speech sound, discrimination is predicted to be good to very good depending on acoustic differences between the sounds. One study in particular has dealt with this situation: Best, McRoberts, & Sithole (1988) found that English listeners did not assimilate Zulu clicks as speech sounds, describing them instead as, "'clicks,' 'plops,' 'pops,' 'percussion instruments,' 'water drips,' 'finger snap,' and 'clap,' ... 'tongue popping,' 'tongue clucking,' and 'sounds coming from different areas in the mouth'" (Best, McRoberts, & Sithole, 1988, p. 352) Discrimination scores for pairs of clicks ranged from 80% to 99% and the differences in level of discrimination could be accounted for by the degree of acoustic difference between the members of each pair. Best, McRoberts, & Sithole argued that the reason listeners were able to make use of acoustic differences to differentiate the click sounds was precisely because the clicks were not assimilated as speech sounds. They cited examples of similar acoustic differences occurring in poorly identified non-native speech pairs that were single-category or category-goodness difference assimilated. They argued that because the clicks were not assimilated to speech sounds, they were not subject to phonemic influences that would have blocked the listeners from perceiving the acoustic differences. With respect to vowels, Best (1995b) notes that it is most probably impossible for a vowel contrast to be nonassimilable (she argues that since low vowels and stops, and labial and velar articulations are universal, all possible vowels fall within the gesture based articulatory space which forms part of the theoretical background of her model).

2.1.3 The Speech Learning Model

The Speech Learning Model (SLM) developed by Flege (1981, 1987, 1988, 1991a, 1992, 1995) claims that a major cause of a foreign accent in production is inaccurate perception of L2 sounds. The SLM states that the production of a sound will eventually correspond to its perception-based phonetic category, and that without accurate perception of the L2, the learner may be unable to establish accurate phonetic categories on which to base L2 production.

The SLM states that L2 sounds will be perceived at a context sensitive allophonic level, a claim previously made by Brière (1966). Munro, Wang, & Li (2000) therefore argued that research into the mapping of L2 sounds to L1 sounds should not be restricted to a single context.

Flege (1988) cited several studies that found no or little correlation between length of residence (LOR) in the L2 society and degree of foreign accent for adult L2 learners. Flege (1991b) found no significant difference in the perception of L2 learners who had been resident in the L2 society for 6.8 years compared to 0.8 years. Flege (1992) observed that several segmental production studies have found greater individual variation than variation between different groups with large differences in LOR. He speculated that individual differences in vowel production may emerge and stabilise within the first six months of learning. This implies that the properties of L2 perceptual categories will be established and show the most rapid change during the first few months of L2 learning. Flege & Liu (2001), however, proposed that LOR would be a significant predictor for L2 perception if the L2 learners interacted extensively with native speakers of the L2 and were thus exposed to large amounts of appropriate input. Chinese listeners' perception of English word-final consonants correlated with LOR for participants who were students, who had extensive interaction with native English speakers, but not for participants who were non-students (primarily researchers), who had relatively little interaction with native English speakers (participants were divided into short-LOR, 0.5-3.8 years, and long LOR, 3.9-15.5 years, groups, and orthogonally into student and non-student groups where there was no significant difference in LOR for student versus non-student groups).

2.1.3.1 Similar sounds

By a principle first called *phonological translation* (Flege, 1981, based on Catford, 1965) and later called *equivalence classification* (Flege, 1987, 1991a, 1992, 1995), L2 sounds that are "similar" to an L1 sound are identified as examples of that L1 sound. Although listeners may be able to distinguish the acoustic differences between the similar sounds, equivalence classification prevents them from making use of these differences in speech processing. When an L2 sound is assigned to the same category as an existing L1 sound, a *merger* will occur creating a *diaphone* category covering both the L1 and L2 sounds (Flege, 1995).

The SLM states that phonetic categories established for the L1 in childhood continue to develop over a person's life span. If *phonetic learning* continues, then experience with both L1 and L2 sounds will act on a diaphone category and modify its phonetic properties. Over time the cumulative effect of the differences between the L1 and L2 sounds will affect the boundaries or prototypical[1] values used to perceive the category. For this to occur, learners must be able to detect within-category differences whilst still assigning sounds to particular phonetic categories. There is evidence that within-category differences can be perceived. Flege (1984a) and Flege & Hammond (1982) found that monolingual speakers could detect and produce certain within-category differences which were indicative of foreign accents but still assigned the foreign accented speech sounds to the appropriate categories.

As a result of equivalence classification, the L2 learner will not develop L2 phonetic categories which match native-speaker norms. Instead, because of the merger of L1 and L2 sounds in the diaphone category, the learner will produce the L2 sound with properties intermediate between the properties of the L1 category and the norm for the L2 category. Equivalence classification will therefore impose an upper limit on the degree to which an L2 learner can approximate the phonetic norms for an L2 segment.

[1]"Prototypical" is used here as a convenient label and is not intended to imply that the theory or research results necessarily support prototype theory. This thesis will, however, take the position that categorical boundaries are a meaningful construct.

Several studies have found that similar L2 vowels are perceived and produced with properties intermediate between the L1 and L2 vowel. Munro (1993) found that experienced Arabic learners of English tended to produce English vowels with spectral values that were intermediate between those of native-English vowels and similar Arabic vowels. Bohn & Flege (1990) found that English /i/ and /ɪ/ were perceptually similar to German /i/ and /ɪ/, and Bohn & Flege (1992) found that German speaking learners of English tended to produce English /i/ and /ɪ/ with spectral values that were intermediate between those of German vowels and the native-English vowels. The duration of the Germans' /ɪ/ was also intermediate between German and English norms. Bohn & Flege (1992) also tested the production of English /ɛ/, which is similar to German /ɛ/. Experienced German learners of English produced English /ɛ/ with properties intermediate between German /ɛ/ and native-English /ɛ/. These findings are consistent with the theory that equivalence classification blocks the full approximation of L2 vowels to L2 norms.

2.1.3.2 New sounds

According to the SLM, "new" L2 sounds are not assimilated to L1 sounds and are therefore not subject to equivalence classification. The SLM predicts that L2 learners will set up new categories for these L2 sounds, and that the properties of a new category will be based only on an L2 sound, as produced by native speakers of the L2, rather than on a combination of an L1 and an L2 sound. The properties of the new category may match the properties of the equivalent category for a native speaker of the L2, in which case the L2 learner will succeed in producing new sounds in a native-like manner, in contrast to a similar sound for which they will only partially approximate native norms. The properties of the new category will not match those of the equivalent category for a native speaker of the L2, if the L2 learner uses different cue weighting (see Section 2.1.3.6) or if the new sound is deflected: The SLM posits that bilinguals have a single phonetic space covering both their L1 and L2, and that they strive to maintain contrasts between all speech sounds from both the L1 and L2; hence, the SLM predicts that the properties of a new category and neighbouring L1 categories will be subject to phonetic category dissimilation and will be deflected away from each other.

One study which examined the difference between sounds perceived as new and sounds perceived as similar was Flege (1987): He found that very experienced English speaking learners of French produced the similar French /u/ with F2 values which were intermediate between those of monolingual English speakers and monolingual French speakers. In contrast, all but the least experienced participants produced the new French /y/ with F2 values that were within the same range as those produced by monolingual French speakers. It was not clear, however, that French /y/ was indeed a new category; inexperienced learners may simply have been confused as to which word contained which phoneme. Flege, Shirru, & MacKay (in press) tested the production of English vowels by Italian immigrants to Canada. They found that Italian participants who had a late age of arrival in Canada and used a large amount of Italian (late-high) produced English /e/ with less formant movement than English /e/ produced by native English speakers. This was consistent with the late-high participants having perceived English /e/ as similar to Italian /e/ (which lacks formant movement). Italian participants who had an early age of arrival in Canada and made little use of Italian (early-low) produced English /e/ with significantly more formant movement than English /e/ produced by native English speakers. This was consistent with the early-low participants having established a new category for English /e/ with exaggerated formant movement in order to maintain phonetic contrast between this new English /e/ category and the existing Italian /e/ category.

2.1.3.3 Identical sounds

Early formulations of the SLM (Flege, 1987, 1988) included the category "identical" alongside "similar" and "new". Flege (1992) notes that:

> ... a "U" shaped rather than a linear function may describe better the effect of varying differences between L1 and L2 sounds. An L2 sound may be pronounced authentically if it is identical to a sound in the L1, or if it is *so similar* to an L1 sound that the differences between it and the nearest L1 sound will go unnoticed if the L1 sound is substituted for the L2 sound. (p. 566)

2.1.3.4 What constitutes new and similar?

One problem in the SLM is the question of what constitutes "new" and "similar" sounds. "New" and "similar" are discrete and there is no principled method for distinguishing the boundary between them. The formulation of the SLM in Flege (1995) deals with the problem by referring to likelihoods based on degree of perceived phonetic dissimilarity: A new category *can* be established if a learner perceives at least some of the differences between the L1 and L2 sounds, and the greater the perceived dissimilarity, the greater the chance that a new category will be established. Flege (1995) discussed whether the phonetic distance between sounds should be measured in terms of acoustic difference, phonological features, or articulatory gestures but did not reach a definitive conclusion. Flege (1991a) proposed that an accurate picture of what constituted new and similar sounds would require phonetic distance scaling by participants in a laboratory setting.

2.1.3.5 No uncommited space

Rochet (1995) argued that all L2 sounds that are assimilated as speech sounds are perceived as similar to some L1 sounds and so there is no uncommitted perceptual space. Flege (1992) had suggested that an L2 vowel that appeared in a part of the vowel space that was "uncommitted" in the L1 would be perceived as new. Rochet (1995) suggested that perceptual boundaries are more important than perceptual prototypes and that all vowel categories expand as far as the limits of the next category, leaving no uncommitted space and thus no room for new vowels. He found that Canadian English and Brazilian Portuguese listeners divided a high vowel continuum into /i/ and /u/, and did not have an uncommited space corresponding to Parisian French /y/. Note that the formulation of the SLM in Flege (1995) (see section 2.1.3.4) does not require an L2 sound to be in uncommitted perceptual space (i.e., assimilated as uncategorisable) in order for a "new" category to be established.

2.1.3.6 Different cues or different cue weighting

The SLM states that an L2 learner may base a new category on criteria that are different or weighted differently to those of a native speaker (Flege, 1995). Several studies have found that L2 learners give greater weight to vowel duration than vowel spectral properties (Bennett, 1968; Weiss, 1976; and Bohn & Flege, 1990). Bohn (1995) tested L2 learners on L2 vowel contrasts that did not occur in their L1. He found that both Spanish-speaking and Mandarin-speaking learners of English identified vowels on an English /i/-/ɪ/ continuum based primarily on duration, whereas native English speakers primarily used spectral cues. German speaking learners of English also primarily used duration to distinguish an English /ɛ/-/æ/ continuum where native English speakers primarily used spectral cues. Flege & Yang (unpublished study cited in Flege, 1995) found that Korean learners of English produced English /i/ and /ɪ/ tokens that were poorly identified and had substantial F1-F2 overlap, but that they produced a greater duration contrast between /i/ and /ɪ/ than did native English speakers. Wang (1997) found that Mandarin learners of English produced considerable spectral overlap between English /i/ and /ɪ/ but that their English /i/ was longer than the English norm (even though Mandarin /i/ was shorter) resulting in an exaggerated duration difference between /i/ and /ɪ/. Munro (1993) also found that very experienced Arabic speaking learners of English greatly exaggerated the duration differences between English tense and lax vowels. Burnham (1986) proposed that duration-based contrasts were more robust and were lost later by infants and more easily retrained than more fragile spectral contrasts.[2] He related robustness to psychoacoustic salience.

2.1.4 Desensitisation Hypothesis

Bohn (1995) accounted for L2 listeners' preference for duration cues over spectral cues via the *Desensitisation Hypothesis*:

> This principle states that whenever spectral differences are insufficient to differentiate vowel contrasts because previous linguistic experience did not sensitize listeners to these spectral differences, duration differences will be used to differentiate the non-native contrast. (p294)

Butcher (1976, cited in Bohn, 1995) found that English listeners perceived a greater difference between cardinal /ɛ/ and cardinal /a/ than did German listeners. L1 developmental changes were opposite in English and German: English speaking adults perceived a greater difference than English speaking children, German speaking adults perceived less of a difference than German speaking children. Butcher therefore suggested that English speakers are *sensitised* and German speakers *desensitised* to spectral changes in the low-front part of the vowel space because English vowels are acoustically closer together in that part of the vowel space than are German vowels. Gottfried & Beddor (1988) found that US English speakers made greater use of duration than French speakers when identifying a French /o/-/ɔ/ continuum. This may have been due to spectral desensitisation since US English has fewer back vowels. Bohn's (1995) finding that German learners of English identified an English /ɛ/-/æ/ continuum primarily on the basis of duration may have been due to L1 transfer: German makes greater use of duration than English in distinguishing vowel pairs. The finding that Mandarin and Spanish speaking learners of English identified an English /i/-/ɪ/ continuum primarily on the basis of duration, could not, however, be due

[2] Burnham made specific reference to VOT duration and spectral differences in place of articulation for stops. The concept that in general duration differences are more robust than spectral differences would also account for the results of the vowel studies cited above.

to L1 transfer: Neither Mandarin nor Spanish makes use of duration to distinguish vowels. Bohn therefore accounted for the latter result via the Desensitisation Hypothesis.

2.1.5 The Perceptual Magnet Effect

The *Perceptual Magnet Effect* and *Native Language Magnet Model* (NLM) (Kuhl, 1991; Kuhl et al., 1992; Iverson & Kuhl, 1995; Kuhl & Iverson, 1995), provide an alternative explanation for why listeners may be unable to perceive spectral differences in certain parts of the vowel space. Grieser & Kuhl (1989) and Kuhl (1991) found that infants and adult native speakers were good at identifying prototypical sounds for native speech categories. Kuhl (1991) had English speakers rate the prototypicality of over 100 /i/ tokens and then selected a prototypical and non-prototypical exemplar. By altering F1 and F2, variants, equidistant in mels, were created centred on each exemplar. The ability of six-month-old and adult native listeners to distinguish the variants from the exemplars was measured. For both groups, variants the same distance from their respective exemplars were more likely to be equated with the prototypical than the non-prototypical vowel. Kuhl explained this by positing the Perceptual Magnet Effect whereby linguistic experience causes the perceptual space around a category prototype to be distorted so that perceptual distance near the prototype is reduced. This results in reduced discrimination sensitivity around category prototypes. Kuhl et al. (1992) demonstrated that perceptual magnets may be established due to exposure to particular languages. Six-month old Swedish infants showed a magnet effect for Swedish /y/ but not English /i/, whilst infants in the US showed a magnet effect for English /i/ but not Swedish /y/.

The basis for vowel prototypes in the NLM has been challenged: Nakai (1997, 1999) had Japanese and Greek listeners choose, from a grid of synthetic vowels, those vowels which they perceived as closest to their L1 vowels /i e a o u/. With the exception of /a/, participants chose vowels that had spectral properties which were more extreme than the mean F1 and F2 values of their vowel productions. Thus the Perceptual Magnet Effect may be caused by vowel extremity rather than by an established L1 category. Lotto (2002) presented evidence suggesting that both factors are at work: in both human and animal tests he found that participants responded most strongly to stimuli that were furthest from categorical boundaries and also to stimuli at the statistical centroids for categories; e.g. peaks were found for the most peripheral /i/ stimuli and for stimuli with spectral values which might be called prototypical for /i/.

Although the Perceptual Magnet Effect is described in terms of prototypes, the NLM also makes reference to boundaries. Kuhl & Iverson (1995) posited boundaries to be innate. Rather than learning boundaries, infants learn to ignore innately specified boundaries which are not relevant to the language they are exposed to. The magnets dominate areas circumscribed by the remaining boundaries. This theory is therefore not necessarily incompatible with Rochet's (1995) assertion that perceptual boundaries are of more importance than prototypes. It also helps to explain why the single boundaries between the Portuguese and English participants' two high vowel categories fell in approximately the same place as one of the two boundaries between the French participants three high vowel categories, rather than falling midway so as to evenly divide the vowel space (the Portugese /i/-/u/ boundary was in approximately the same place as the French /y/-/u/ boundary, and the English /i/-/u/ boundary was in approximately the same place as the French /i/-/y/ boundary). Kuhl & Iverson (1995) claimed that the boundaries do not completely disappear but are accessible when the listener perceives sounds in non-speech mode, i.e., when the sounds are nonassimilated.

The establishment of perceptual magnets for the L1 explains why listeners have difficulty discriminating L2 vowels that are spectrally close to L1 vowels. This surfaces in the assimilation of L2 vowel to L1 vowels in the PAM, and in equivalence classification in the SLM.

2.2 Post-vocalic Consonant Voicing

Section 2.1 has presented a general description of current models of cross-language speech perception. Sections 2.2 and 2.3 will review studies dealing with the particular phenomena to be investigated in this thesis: Section 2.2 will review studies dealing with vowel and consonant duration as cues to post-vocalic consonant voicing; and Section 2.3 will review studies dealing with the perception and production of high front vowels in English, Japanese, and Spanish, focussing especially on vowel duration effects in cross-language studies.

Multiple cues have been found to signal post-vocalic consonant voicing: phonetic voicing (i.e., vocal cord vibration) (Wardrip-Fruin, 1982), consonant closure duration (Denes, 1955; Lisker, 1957; Suen & Beddoes, 1974), release burst properties (Wolf, 1978; Repp & Williams, 1985); preceding vowel duration (Denes, 1955; Zimmerman & Sapon, 1958; Peterson & Lehiste, 1960; House, 1961; Chen, 1970; Cochrane, 1970; Raphael, 1972; Wolf, 1978; Hogan & Rozsypal, 1980), vowel F1 offset (Wolf, 1978; Revoile et al., 1982; Walsh & Parker, 1983; Hillenbrand et al., 1984; Fischer & Ohde, 1990), vowel formant steady state (Wolf, 1978; Summers, 1988), and the speed of transition from vowel to consonant (Parker, 1974; Summers, 1987; Walsh, Parker, & Miller, 1987). Kingston (in press) argues that the perception of the phonemic voicing contrast is due to the perceptual integration of multiple cues: listeners do not hear the individual acoustic cues but perceive a higher order auditory cue resulting from the combination of the acoustic cues. He also argues that speakers control independent articulations in order to produce combinations of cues which optimise the contrast in each particular context.

Since this thesis is concerned with differences in L1 and L2 vowel duration perception, the remainder of Section 2.2 will focus on vowel duration and its relationship with post-vocalic obstruent voicing.

2.2.1 Vowel and consonant duration cues

Across languages, vowels preceding voiced consonants tend to be longer than vowels preceding voiceless consonants (Chen, 1970). Various possible explanations for this phenomenon are reviewed in Chen (1970) and Kluender, Diehl, & Wright (1988). In English the vowel duration difference is greater than in other languages and may in itself be sufficient to cue the voicing distinction. Fox & Terbeek (1977) found that US English vowels were significantly longer before flaps that were underlyingly voiced than before flaps that were underlyingly voiceless. In utterance final position English plosives may be phonetically devoiced and lack a release burst, but English listeners can use vowel duration (possibly in combination with spectral cues in the vowel) to reliably identify whether the plosive is underlyingly voiced (Halle, Hughes, & Radley, 1957; Raphael, 1972; Flege, 1989; Crowther & Mann, 1992). Parker (1974) presented anecdotal evidence that naïve native English speakers are even consciously aware that in whispered speech vowel duration is a cue to phonemic voicing. A vowel duration difference has also been observed for phonemically voiced and voiceless final plosives in Russian (Chen, 1970) and German (Port et al., 1981, cited in Keating, 1985) even though word-final phonemic voiced plosives are phonetically devoiced, and the phonemic contrast is usually considered to be neutralised in these languages.

For English, the vowel duration effect has been found to be greater before fricatives and plosive-fricative clusters than before plosives (Peterson & Lehiste, 1960; Raphael, 1972). Hogan & Rozsypal (1980) found that, in Canadian English, vowel duration was an effective cue to the perception of word-final consonant voicing for phonemically long and high vowels combined with fricatives and plosive-fricative clusters, but not generally so for other combinations of vowels and plosives.

Suen & Beddoes (1974) found that Canadian English speakers produced considerably longer closures for post-vocalic voiceless versus voiced plosives. Lisker (1957) found that US English intervocalic /p/ was longer than /b/ in the same context, and that listeners had a categorical boundary when identifying a continuum varying in closure duration. Crystal & House (1982), however, found that in connected speech there was no difference in the duration of the closure portion of voiced and voiceless stops.

Denes (1955) found that consonant duration was a relevant cue to voicing an English /jus/-/juz/ continuum. Consonant duration, however, varied systematically with vowel duration. Port (1981) and Port & Dalby (1982) proposed that it was the ratio of vowel to consonant duration that was relevant in the identification of post-vocalic consonant voicing in English. Port (1981) found that all his participants had a larger V/C ratio for syllables ending in a voiced consonant than for syllables ending in a voiceless consonant. In contrast, Luce & Charles-Luce (1985) found that neither V/C duration ratio nor consonant closure duration were reliable cues to US English post-vocalic plosive and fricative voicing in connected speech, but that vowel duration was a reliable cue.

Kluender, Diehl, & Wright (1988) cited studies attesting vowel duration differences before voiced versus voiceless consonants in fourteen different languages belonging to various language families. They reviewed earlier articulatory and auditory accounts for the vowel length effect, and proposed their own auditory account. They claimed that the primary cues used to signal voicing in medial and final consonants are the presence of glottal vibration during the closure interval, the length of the closure interval, and the length of the preceding vowel. They proposed that speakers could independently control each of these cues so that different languages could apply different weighting to each cue. They also proposed that the different cues are not perceived independently but have mutually reinforcing auditory effects. Of particular importance to Kluender, Diehl, & Wright's (1988) hypothesis was a contrast effect whereby a short consonant will be perceived as even shorter when preceded by a long vowel, and a long consonant will be perceived as even longer when preceded by a short vowel. They claimed that the vowel length effect exists in order to enhance perception of the consonant closure duration which was the primary cue to the voicing distinction in languages with a vowel duration effect. In support of their claim they cited Flege & Port (1981), who found that in Arabic there was no difference in closure duration or preceding vowel duration for voiceless versus voiced final consonants; if consonant closure duration were not a cue to voicing, then vowel duration would not be expected to vary either. Kluender, Diehl, & Wright (1988) claimed that English was not fundamentally different to other languages in its use of vowel duration as a cue to post-vocalic consonant voicing, and that the difference in degree of the effect between English and other languages had been exaggerated. However, their study used only English speakers and required the listeners to make a non-phonemic long-short response rather than a phonemic voiced-voiceless response. This raises two possible problems with their conclusions: first, native English listeners may behave very differently from natives of other languages; second, listeners may have perceived a non-phonemically relevant cue using raw acoustic information in non-speech mode. Perceptual tests requiring phoneme identification and monolingual participants from various language backgrounds would be needed to fully test their hypothesis. Kluender, Diehl, & Wright's (1988) hypothesis would fail to cover the situation in English where a final plosive may be devoiced and have no audible release, hence no consonant closure duration, but where phonemic voicing is still cued by vowel duration.

House (1961) hypothesised that longer vowel duration in tense versus lax vowels, and before voiced versus voiceless consonants is part of the phonology of US English, whilst longer vowel duration in open versus close vowels, and before fricatives versus stops is due to universal physiological constraints. Keating (1985) argued that whilst physiological constraints produce default patterns, all aspects of phonetic control must be learnt. Physiological constraints predict that utterance-final plosives will be voiceless. Inventory and allophonic preferences for word-final devoicing in utterance-medial position, however, suggest that a process initially motivated by physiological constraints has been transformed into a phonological rule. The physiological default is for shorter vowels before voiceless versus voiced consonants but individual languages exert active control to achieve three different patterns: 1. no duration difference (e.g., Czech and Polish [see Keating, 1985]); 2. vowels shorter before phonetically voiceless consonants (e.g., French [see Mack, 1982]); 3. vowels shorter before phonemically voiceless consonants (e.g., German and Russian [see Chen, 1970; Port et al., 1981, cited in Keating, 1985]). In addition a language may exaggerate the duration difference which could result in a language where, in certain contexts, vowel duration alone cues phonemic voicing for consonants (e.g., possible word finally in English). Such a situation may even lead to a consonant voicing contrast being rephonemicised as a vowel duration contrast as in Friulian (Baroni & Vanelli, 2000; Kingston, in press). No language is attested in which vowels are shorter before voiced consonants. This is as predicted since it would not be a natural extension of the physiological default.

2.2.2 Interaction with phonemic vowel duration

Flege & Port, (1981), and Mitleb (1984) found that native Arabic speakers did not produce a significant vowel duration difference before Arabic voiced versus voiceless obstruents. Similarly, Port & Mitleb (1983), Mitleb (1984), and Munro (1993) found that Arabic speakers produced a difference in vowel duration before English voiced versus voiceless consonants that was significantly smaller than that produced by native English speakers. Turning to perception, Flege (1984b) found that both English and Arabic listeners were quite sensitive to vowel duration as a cue to fricative voicing. Thus there appeared to be a lack of parallelism between perception and production. Flege (1984b) suggested that the perceptual results could be accounted for if the Arabic listeners had mapped English vowels before voiced fricatives to long Arabic vowels, and English vowels before voiceless fricatives to short Arabic vowels. In order to ascertain whether such a hypothesis is tenable, it would be necessary to determine whether in the *peas - peace* tokens used by Flege (1984b), the duration difference between English /i/ before voiced /z/ versus voiceless /s/ straddled the categorical boundary between Arabic long /iː/ and short /i/.[3]

The fact that Arabic does not have vowel duration differences before voiced versus voiceless consonants may be related to the existence of phonemic vowel duration, i.e., if a language uses vowel duration differences to cue phonemic vowel duration, then it may not be possible for that language to also use vowel duration to cue consonant voicing. Whilst the existence of phonemic vowel duration contrasts may block vowel duration differences due to consonant voicing, it is not a necessary condition: Keating (1985) found no significant duration difference before voiced versus voiceless word-medial consonants in Czech, a West Slavic language with contrastive vowel duration, or in Polish, a West Slavic language without contrastive vowel duration.

Flege & Hillenbrand (1986) had Swedish, and Finnish listeners identify members of an English *peace-peas* continuum in which vowel duration and consonant duration were varied independently. Neither Swedish nor Finnish has a /s/-/z/ contrast, but both have a phonemic contrast

---

[3]A similar hypothesis – that Japanese listeners' perception of English vowels before voiced and voiceless consonants is dependant on the position of the Japanese long-short vowel boundary – will be tested in this thesis (see Section 2.3.3.2.2).

between long and short vowels and consonants. Fricative duration had no systematic effect on the Swedish and Finnish participants' perception of voicing in the English contrast, but vowel duration was found to have an effect. Flege & Hillenbrand (1986) suggested that the Swedish and Finnish listeners may have assimilated the vowel duration differences in the stimuli to long and short vowel categories in their L1s, and then given responses based on this distinction. Finnish listeners who had never lived in an English speaking country were more sensitive to vowel duration differences than those who had. This suggests that the more experienced learners of English had learnt that differences in English vowel duration, such as those due to the voicing contrast, were not identical to the Finnish vowel duration contrast. Flege & Hillenbrand (1986) suggested that although fricative duration was also contrastive in Swedish and Finnish, the participants had not responded to fricative duration differences because vowel duration differences are intrinsically more salient and L2 learners may have difficulty integrating multiple acoustic cues.

2.2.3 English, Japanese, and Spanish post-vocalic obstruent voicing

In certain contexts such as in utterance final position, phonemic voiced English obstruents may be partially or fully devoiced so that phonetic voicing is not a cue to phonemic voicing. The phonemic voicing contrast is maintained, however, as a difference in vowel duration.

In Japanese, obstruents are always followed by vowels, and the phonetic voicing contrast is always maintained. Japanese also has a contrast between singleton and geminate consonants. Usually only voiceless obstruents can be geminate but voiced geminates occur in loan words and onomatopoeic words. The singleton - geminate contrast is realised phonemically by a difference in the duration of the consonant closure.

Spanish does not have a phonemic voicing contrast for fricatives. Intervocalically the phonetic voicing contrast for plosives is maintained and phonemic voiced stops are realised as voiced approximants, e.g., /bota/ → [bot̪a], /boda/ → [boð̞a]. Phonemic voiced word final plosives are realised as voiced or voiceless fricatives, e.g., /bit/ → [bit̪], /bid/ → [bið] or [biθ] (D'Introno, Del Teso, & Weston, 1995). Spanish does not have a singleton-geminate contrast for obstuents.

2.3 English, Japanese, and Spanish vowels

This section begins with a description of the vowel systems in General Canadian English, Western Japanese, and Mexican Spanish. This is followed by a review of research dealing with the perception and production of vowel duration in Japanese and Spanish, and with Japanese and Spanish speakers' perception and production of English vowels. The review focuses primarily on high front vowels.

2.3.1 Vowel systems

2.3.1.1 General Canadian English vowels

The term "General Canadian English" refers to a relatively homogeneous dialect spoken in western and central Canada. In terms of phonemic monophthongs, the vowel system has five tense vowels /i e ɑ o u/, five lax vowels /ɪ ɛ æ ʌ ʊ/, and a reduced vowel (schwa) /ə/ (Avis, 1975; Gregg, 1975; Walker, 1975). In identical contexts, the tense vowel /i/ is both more peripheral in the vowel space and longer than the lax vowel /ɪ/.[4] Avis (1975), and Gregg (1975) noted that /i/ may be diphthongised moving upwards in the vowel space. Nearey & Assmann (1986) found no significant formant movement in /i/ but did find significant formant movement in /ɪ/ moving downward and backward in the vowel space. This formant movement was such that when the nucleus and offglide sections of the vowel were reversed, listeners primarily identified the vowel as /e/ [ej] instead of /ɪ/. Formant movement, however, did not appear to be essential for /ɪ/ identification since correct identification scores were still relatively high when the offglide section was replaced by a repetition of the nucleus. Assman, Neary, & Hogan (1982) found that when isolated Canadian English vowels were "gated" so as to remove duration and dynamic spectral properties, native listeners' rates of confusion for vowel identification increased from 5 to 12%, a statistically significant but relatively small difference. In a discriminant analysis, a cross-validated correct classification score of 93% was obtained using (normalised) steady-state spectral information. The score rose by only one percentage point when duration and dynamic spectral information were included. On the basis of these results, steady state spectral properties appear to be the primary cue for Canadian English /i/ and /ɪ/ perception.

[4]Gregg (1975) claimed that tense-lax pairs do not differ in duration in Vancouver English, but failed to present any acoustic data in support of this claim. Duration data based on acoustic measurements of English /i/ and /ɪ/ produced by a Vancouver English speaker are given in Chapter 3 of this thesis. This speaker did produce duration differences between the lax and tense vowel.

2.3.1.2 Western Japanese vowels

Standard eastern and western dialects of Japanese have five short vowels /i e a o u/ and five long vowels /iː eː aː oː uː/, associated with one and two morae respectively. The high back vowels are unrounded in eastern dialects but slightly rounded in western dialects (Shibatani, 1990). Confusion of /i/ with /e/ has been reported in the Tokyo (eastern) dialect but not in western dialects (Shibatani, 1990). The long and short vowels are traditionally described as having identical spectral properties (Akamatsu, 1997). All these vowels are phonetically monophthongs with no evidence of F1 or F2 movement (Fitzgerald, 1996). Pitch accents may occur so that the second mora of a long vowel has a lower pitch than the first or vice versa (Shibatani, 1990).

2.3.1.3 Mexican Spanish vowels

Most Spanish dialects, including Mexican Spanish, have five vowels /i e a o u/ with no tense-lax or long-short distinctions. All these vowels are considered monophthongs. Vaquero de Ramírez (1996) noted that in colloquial American dialects both stressed and non-stressed /i/ may be pronounced as [e]. Navarro Tomás (1965) claimed that there were open and close allophones of the non-low vowels, and back and front allophones of the low vowel. Whilst his description of the allophones of the high front vowel suggest that Spanish speakers have experience with an allophonic contrast in Spanish that may be similar spectrally to the tense-lax contrast in English, there is little evidence to support the claim that these allophones exist. From spectral analysis Martínez Celdrán (1994) concluded that there was insufficient evidence to confirm the claims of Navarro Tomás (1965) although he did find that stressed vowels were slightly more peripheral than non-stressed vowels. Morrison (2002b) also failed to find a spectral difference between the putative mid vowel allophones, a difference which Navarro Tomás had claimed was more perceptable than that for the high vowel allophones.

2.3.1.4 Cross-language vowel inventory comparison

With respect to vowel quality, Japanese and Spanish speakers have similar five-vowel systems, and in learning Canadian English are faced with the problem of learning a system with twice as many vowels. Japanese and Spanish differ in that Japanese, but not Spanish, has a phonemic long-short vowel distinction. English /i/ and /ɪ/ differ in both spectral and duration properties in a part of the vowel space where Japanese has /i/ and /iː/ traditionally described as differing in duration but not spectral properties. Spanish has a single vowel /i/ not contrasting in duration or spectral properties with any other vowel in the high-front part of the vowel space.

2.3.2 English speakers' perception of English vowel duration

Hillenbrand et al. (2000) found that when the vowels in /hVd/ words were lengthened and shortened, US English listeners shifted their identification of /æ/-/ɛ/ and /ɑ/-/ɔ/-/ʌ/, but did not shift their identification of /i/-/ɪ/, /u/-/ʊ/, and /ɪ/-/e/-/ɛ/. They proposed that the reason for this was that high and front-mid vowels had considerable spectral separation making vowel duration differences redundant, whereas the low vowels had spectral overlap so that duration was a pertinent cue to vowel identification.[5] Hence although English lax-tense vowel pairs differ in duration and spectral properties, native English listeners are expected to ignore the duration cues as long as spectral cues are maintained.

2.3.3 Japanese speakers' vowel perception and production

2.3.3.1 Perception and production of Japanese vowel duration

2.3.3.1.1 Phonemic long-short distinction

Japanese vowels may be phonemically short, associated with 1 mora on the prosodic tier, or long, associated with 2 morae on the prosodic tier. Akamatsu (1997) analysed a long vowel as a sequence of two identical vowels. He emphasised that there is no difference in quality between a single vowel and a sequence of two identical vowels, and that there is no difference in quality or intensity between each of the two vowels in sequence. Fitzgerald (1996), however, found that, although there was considerable overlap, the mean F1-F2 frequencies for long vowels were more peripheral compared to those of short vowels. Short vowels were also more susceptible to allophonic variation due to consonant context; however, very little allophonic variation was observed for either /i/ or /iː/.

Japanese listeners have a sharp categorical boundary for the perception of long and short vowels (Enomoto, 1992). Fujisaki & Sugito (1977, cited in Takahashi, 1987) found that the categorical boundary between short and long vowels was in the region of 150-160 ms. The boundary

---

[5]There is little, if any, reason to expect a different result for Canadian English /i/-/ɪ/. Canadian English /æ/-/ɛ/, however, are better differentiated spectrally and so duration differences may not be as important as in US English (M. J. Munro, personal communication, 30 April 2002).

was longer when the word was embedded in a carrier sentence than when it was presented in isolation, e.g., 168 ms versus 156 ms for the /o/-/oː/ boundary. Toda (1999) had native Japanese listeners indicate the just noticeable duration difference on (C)VCV - (C)VCVV continua (presented both as "rising" series in which vowel duration was gradually increased, and as "falling" series in which vowel duration was gradually decreased). Doubling the duration of the first vowel had a statistically significant effect on the position of the categorical boundary for the final vowel; however, the categorical boundary value increased by a relatively small amount, in the order of 10%. When the first vowels were set to 70% of their natural durations, the long-short boundaries for the final vowels ranged from 231 to 245 ms, and when the first vowels were set to 140%, the range was 245 ms to 267 ms.

Several production studies have examined the difference in duration between long and short Japanese vowels. Han (1962) found that the long-short vowel duration ratio was 2 when not preceded by a consonant, 2.5 when preceded by a voiced consonant, and 3 when preceded by a voiceless consonant. Vowel durations were such that CVV syllables were twice as long as CV syllables. Fitzgerald (1996) found that for /C_Ca/ words in a carrier sentence, the mean long-short vowel duration ratio was 2.04 (73-150 ms), with individual speakers ranging from 1.97 (73-143 ms) to 2.10 (65-138 ms). Vowel duration also varied with vowel height: lower vowels were longer than high vowels. For a given long-short vowel pair, a minimum 1.5 duration ratio was maintained by each speaker. The mean duration ratio for /iː/-/i/ was 1.90 (137-72 ms). Across speakers and consonant contexts, however, duration overlaps were possible: the longest short vowel, a /g**i**ga/ by one speaker, was 116 ms, and the shortest long vowel, a /b**iː**ba/ by another speaker, was 102 ms. It may be that differences in speaking rate had contributed to this overlap (Fitzgerald did not provide speaking rate data).

The presence of a pitch accent on a vowel, realised as a drop in pitch in the middle of the vowel, has also been found to cue the perception of two morae (i.e., a long vowel) (Baba et al., 1996; Kozasa, 2002). Experiments in the present thesis will be restricted to stimuli words which do not have a pitch change in the vowel, and the thesis will not address pitch accent as a cue to vowel duration in Japanese.

In summary, the studies reviewed above indicate that Japanese speakers maintain a long-short vowel contrast ratio of 1.5 to 3 depending on phonetic context. Across speakers and contexts absolute durations for long and short vowels may overlap. Speaking rate has only a small effect on the location of the perceptual boundary between long and short vowels.

2.3.3.1.2 Allophonic duration differences due to consonant voicing

A small duration effect has been found for Japanese vowels preceding voiced versus voiceless plosives and fricative, with the effect being greater preceeding fricatives. Tsukada (1996, 1999) investigated the production of Japanese long and short vowels in isolated /C_to/ and /C_do/ words. Vowels were longer before /d/ than /t/; however, the difference was relatively small, the mean ratio being 1.06. Average absolute differences were 13.7 ms for female speakers and 3.4 ms for male speakers. The mean long-short vowel duration ratio was approximately 3. A large ratio was maintained even when phonemic duration was confounded by the consonant voicing effect: the mean durations for *kado* and *kaato* were 90 ms and 210 ms respectively. Takahashi (1987) found that in the Japanese words /suːsi suːzi isu izu/, the vowels preceding the voiced fricatives were longer than those preceding the voiceless fricatives by ratios ranging from 1.18 to 1.38, and the voiceless fricatives were longer than the voiced fricatives by a mean ratio of 1.45. Fricative voicing had no effect on the vowels following the fricatives.

Homma (1973, 1981), and Maeda (1979, cited in Takahashi, 1987) also found that vowels were longer before voiced versus voiceless consonants, but that the lengthening effect was greater for vowels that followed voiced versus voiceless consonants. In Homma (1981) the preconsonantal difference was 10 ms (12%), and the postconsonantal difference 28 ms (36%). Port et al. (1980) found that in Japanese CVCV words, both the preconsonantal and postconsonantal vowels were longer if the second consonant was a voiced as opposed to a voiceless stop. The voiced stop was also shorter than the voiceless stop. There was thus an inverse relationship between vowel and consonant duration such that the length of the two-mora word remained unchanged. This led Port et al. (1980) to posit word timing as well as segment timing control. The same effect was found across manners of articulation so that a word had almost the same duration whether the second consonant was intrinsically long, e.g., /s/, or short, e.g., /r/. Homma (1981) found the same compensatory lengthening and word timing in CVC(C)V words with voiced and voiceless consonants. In contrast, Beckman (1981, cited in Takahashi, 1987) and Takahashi (1987) did not find such compensatory lengthening, but Beckman did find that Japanese vowels were longer before voiced than voiceless consonants. According to Port el al.'s (1980) timing model, preconsonantal vowel duration does not vary to cue consonant voicing as in English but varies as part of a compensatory system which maintains equal duration for words with the same mora count. Since voiced consonants are shorter than voiceless consonants, vowel duration will vary with consonant voicing. Vowel duration, however, will be interpreted by native listeners as a cue to the moraic structure of the word rather than as a cue to consonant voicing.

In summary, the studies above indicate that in Japanese, vowel duration effects due to postvocalic plosive voicing are small compared to phonemic vowel duration differences. Duration differences due to postvocalic fricative voicing are somewhat larger than those due to plosive voicing. Port el al. (1980) theorised that the vowel duration effect in Japanese is not particular to postvocalic voicing, but is part of a word-timing system that adjusts the duration of individual segments in order to maintain equal duration for words with equal mora counts.

2.3.3.2 Japanese Speakers' Perception and Production of English vowels

2.3.3.2.1 Assimilation of English vowels to Japanese categories

Ingram & Park (1997) had Japanese listeners classify Australian English vowels in terms of Japanese categories. The vowels were /i ɪ e æ a/ presented in isolated /h_d/ words. The stimuli had been produced by two native speakers of Australian English, one of whom spoke faster than the other resulting in different vowel durations: one speaker's /ɪ/ and /i/ vowels were approximately 75 and 255 ms respectively, and the other's were 165 and 390 ms. All vowels were consistently assimilated to long or short Japanese categories irrespective of the between-speaker duration differences. Japanese listeners assimilated English /i/ to Japanese /iː/ at a rate of 100% and assimilated English /ɪ/ to Japanese /i/ at a rate of 90% (10% as /iː/). Ingram & Park hypothesised that the Japanese listeners were able to normalise for inter-speaker speaking rate differences because of the phonemic vowel duration contrast in Japanese. Korean participants, however, were not able to normalise for speaking rate differences (modern Korean does not have a vowel duration contrast).

Strange et al. (1996) and Strange et al. (1998) had Japanese listeners with limited exposure to English categorise US English vowels in terms of Japanese categories. The listeners also rated the English vowels for goodness of fit with the Japanese category. The vowels were produced by four speakers and presented in a /h_bɑ/ context both in isolation and embedded in a carrier sentence. Strange et al. divided the English vowels into two groups according to duration; vowels in the long group (/i e æ ɑ ɔ o u/) were 30-45% longer than those in the short group (/ɪ ɛ ʌ ʊ/). Long English vowels were assimilated to Japanese long vowels more often in the sentence condition than the isolated word condition (85 versus 42%). Short English vowels were assimilated to Japanese short vowels more often in the isolated word condition than in the sentence condition but the magnitude of the difference was much smaller than for the long vowels (94 versus 83%).[6]

Strange et al. (1998) hypothesised that the difference in assimilation patterns between conditions might have been due to the final /ɑ/ vowel in /h_bɑ/ being longer in the isolated word condition than in the sentence condition. The test was repeated using isolated word stimuli in which the final /ɑ/ was truncated, but this had little effect on the assimilation of the target vowels. Guion et al. (2002) obtained similar results, finding no significant difference in the assimilation of English vowels to Japanese long and short categories in isolated /b_bo/ words in which the final /bo/ had two durations based on different speaking rates. There appear to be three possible explanations for these results:

1. The duration differences in the final part of the stimuli (/ɑ/ in Strange et al., 1998; and /bo/ in Guion et al., 2002) may not have been enough to signal a speaking rate against which the listeners could calibrate their Japanese long-short vowel boundary. They would then have presumably used some default boundary value applicable to words in isolation.

2. The duration differences in the final part of the stimuli may have been sufficient to indicate a speaking rate difference which affected the Japanese long-short vowel boundary, but the change in boundary duration did not affect the classification of the target vowels. For example, If there were two target vowels of duration 100 ms and 150 ms and a boundary of 140 ms at the slow speaking rate and 110 ms at the fast speaking rate, then there would be no difference in identification of these vowels across speaking rates.

3. Changes in speaking rate may not affect or may have only a minimal effect on the position of the Japanese long-short vowel boundary. As mentioned in section 2.3.3.1.1 above, Toda

[6] These findings are at odds with the finding of Fujisaki & Sugito (1977, cited in Takahashi, 1987) that the long-short boundary was longer when the word was embedded in a carrier sentence than when it was presented in isolation.

(1999) found that doubling the duration of the first vowel in Japanese CVCV words only resulted in the long-short boundary being 10% greater.

Strange et al.'s (1998) second hypothesis was that the difference in assimilation patterns between isolated word and sentence conditions was due to acoustic differences in the target vowel. Their acoustic analysis revealed that vowels in the sentence condition were on average longer than those in the isolated word condition. For stimuli produced by two of the four English speakers, however, differences in assimilation could not be attributed to relative duration differences between conditions, nor for one of these speakers could they be attributed to absolute duration. Strange et al. (1998) concluded that "the Japanese listeners appeared to have some uniform standard across speakers for judging the cross-language similarity in temporal structure" (p. 337). They suggested that duration differences in stimuli could only partially account for differences in assimilation and that the participants may have been using the larger rhythmic pattern of the sentence to calibrate their responses.

The pattern of assimilation of English /i/ and /ɪ/ to Japanese vowel categories is shown in Table 2.1 (the data extracted from Strange et al., 1996, are from the same experiments described in Strange et al., 1998). The duration based assimilation of English /i/ reflects the general assimilation pattern reported above for long English vowels. The high category goodness ratings for English /i/ likely indicate that it is spectrally close to Japanese /i/ and /iː/. English /i/ durations do not seem to have affected category goodness ratings once the English vowel was assimilated to either a long or short Japanese category. This may indicate that the Japanese listeners ignored duration when assigning category goodness scores, or that an acceptable vowel duration in Japanese is almost any duration that is not exactly at the categorical boundary; e.g., a long vowel is a good long vowel whether it be 25 or 50 ms longer than the long-short boundary. English /ɪ/ was assimilated to both Japanese /i/ and /e/ with relatively low category goodness ratings. This likely indicates that the spectral properties of English /ɪ/ are intermediate between these two Japanese categories. The rate of assimilation to Japanese /i/ was greater in the sentence condition. This may be because the listeners were able to adjust their responses relative to the spectral properties of the other vowels in the sentence.

Table 2.1 Assimilation of English /i/ and /ɪ/ to Japanese vowel categories in Strange et al. (1996). Assimilation rates given as percentages. Category goodness ratings (maximum 7) are given in parenthesis.

English Stimulus	Condition	Japanese Response		
		iː	i	e
i	Sentence	83 (6)	17 (6)	
	Isolated word	40 (6)	59 (6)	
ɪ	Sentence		77 (4)	16 (4)
	Isolated word		58 (3)	39 (4)

Guion et al. (2002) also had Japanese speakers with limited exposure to English classify US English vowels according to Japanese categories. /b_bo/ words were produced by eight different speakers of US English and presented in isolation. Effects due to speaking rate were not significant. The pattern for the assimilation of English /i/ and /ɪ/ to Japanese vowel categories is shown in Table 2.2. The assimilation pattern for English /i/ was similar to that in Strange et al. (1996) and Strange et al. (1988). The assimilation pattern for English /ɪ/ was, however, different in that it was almost always assimilated to Japanese /e/. The difference may have been due to dialectal or ideolectal differences between the speakers who produced the stimuli (/e/ responses in Strange et al., 1988, were disproportionally higher for the productions of one speaker). Alternatively, the difference may have been due to differences in the surrounding segments /b_bo/ versus /h_bɑ/.

Table 2.2 Assimilation of English /i/ and /ɪ/ to Japanese vowel categories in Guion et al. (2002). Assimilation rates given as percentages. Category goodness ratings (maximum 7) are given in parenthesis.

Condition	English Stimulus	Japanese Response			
		iː	i	eː	e
Isolated word	i	36 (4.3)	60 (4.8)		
	ɪ			7 (3.5)	87 (4.7)

Nozawa & Flege (2000) tested two groups of Japanese learners of English, one living in Japan and one living in the US, on their ability to discriminate US English vowels in /b_t/ words. For both groups, discrimination ability was high for the /i/-/ɪ/ contrast and the groups did not differ significantly. For the /ɪ/-/ɛ/ contrast, the discrimination ability of the group living in the US was significantly greater than that for the group living in Japan. This suggests that these English /ɪ/ stimuli may have been assimilated, along with English /ɛ/ stimuli, as the poorer member of a category-goodness assimilation to Japanese /e/, the more experienced listeners being better at distinguishing the within-category difference.

In summary, the studies reviewed above suggest that the spectral properties of English /i/ are such that it is assimilated to Japanese /i iː/. English /ɪ/ is spectrally a poor match for any Japanese category but is assimilated as Japanese /i/ or /e/. The choice of /i/ or /e/ may be affected by the phonetic context in which the English /ɪ/ occurs, and by the dialect of the English speaker producing the /ɪ/. The duration of English /ɪ/ is such that it is assimilated as a short Japanese vowel. English /i/ is assimilated to Japanese /iː/ at greater rates in a sentence context than in an isolated word, and Japanese listeners may calibrate their long-short responses using the larger rhythmic pattern of the sentence. Finally, speaking rate has little effect on Japanese listeners' assimilation of vowel duration. The latter statement can be interpreted in two contradictory ways:

1. Speaking rate has a minimal effect on the position of the long-short categorical boundary for Japanese vowels (Strange et al., 1998; Guion et al., 2002).
2. Japanese listeners are able to adjust their perceptual use of duration to normalise for different speaking rates (Ingram & Park, 1997).

The question of which of these interpretations is correct will be addressed below.

2.3.3.2.2 Identification of English vowels by Japanese listeners

Ingram & Park (1997) had Japanese listeners identify non-back Australian English vowels in terms of English categories. The vowel stimuli were the same isolated /h_d/ words used in the Japanese assimilation test described in Section 2.3.3.2.1. Experienced Japanese learners of English made only one error, and inexperienced Japanese learners of English made eight errors (an error rate of only 3.2%). Ingram & Park hypothesised that the Japanese listeners were able to obtain high correct-classification scores because they used duration to distinguish /i/ and /ɪ/, and further that they were able to compensate for speaking rate differences because of their experience with the Japanese long-short vowel duration contrast. Presumably the Japanese participants also made some use of spectral cues since they correctly identified the English /ɪ/ tokens at a higher rate (they made only one error) than the rate at which they were assimilated to Japanese short /i/.

Minnick Fox & Maeda (1999) had Japanese learners of English discriminate US English /i/ and /ɪ/ in a two-alternative forced choice task. The vowels were contained in words spoken by several speakers. The vowels were modified in different ways: *natural* (unmodified), *long* (double length), *short* (half length), and *uniform* (all vowels made 140 ms long). The participants were given computer mediated perceptual training using stimuli produced by a different set of speakers from those used in the test. One group of participants was trained using natural stimuli, and another group was trained using uniform stimuli. Some of the results were difficult to interpret given that there was an imbalance between group scores in the pre-test. For sake of simplicity, the numerical data presented here are the result of averaging the values which Minnick Fox & Maeda report separately for each group. Training appeared to be successful since the correct-identification rates for natural stimuli rose between the pretest and post-test (see Table 2.3). In the pretest, correct-identification rates for the long /ɪ/ and short /i/ stimuli were around chance level, i.e., close to 50%. Since correct-identification rates for natural stimuli were considerably higher, this indicates that the participants were using vowel duration as the primary cue to distinguish /i/ and /ɪ/. In the post-test, correct-identification rates for the long /ɪ/ and short /i/ stimuli were almost as high as those for natural stimuli. The improvement in correct identification rates indicates that the participants had learnt to distinguish English /ɪ/ and /i/ using spectral cues.

Table 2.3 Identification of English /i/ and /ɪ/ in Minnick Fox & Maeda (1999). Correct-identification rates given as percentages.

English Stimulus	Condition	Correct Identification Rate	
		Pretest	Post-test
i	Natural	75	88
	Short	45	82
ɪ	Natural	77	89
	Long	56	87

Morrison (2002a, 2002c) had Japanese and Spanish listeners identify Scottish / Northern English vowels in isolated CVC words. In one group of words, both consonants were (phonetically) voiced, and in another group, both were voiceless. The Japanese participants' response patterns for English /i/ and /ɪ/ are shown in Table 2.4. Japanese participants had a high correct-identification rate for English /ɪ/ in both voiced and voiceless consonant conditions. They also had a high correct-identification rate for English /i/ in the voiced consonant condition, but the majority of responses in the voiceless consonant condition were /ɪ/. To account for this finding, Morrison (2002a) proposed that English /i/ before a voiced consonant is assimilated to Japanese long /iː/ but assimilated to Japanese short /i/ before a voiceless consonant; that English /ɪ/ before both a voiced and voiceless consonant is assimilated to Japanese short /i/; that English vowels assimilated to Japanese long /iː/ are identified as English /i/, and that English vowels assimilated to Japanese short /i/ are identified as English /ɪ/. This requires that the Japanese long-short vowel boundary fall between the longest duration for an English /i/ stimulus in the voiced consonant condition or for an English /ɪ/ stimulus in either condition (whichever be the greater), and the shortest duration for an English /i/ stimulus in the voiceless consonant condition. Comparison with the data in Tsukada (1996) and Ingram & Park (1997), see Figure 2.1, suggested that the Japanese long-short vowel duration boundary did indeed fall within this range. Morrison's (2002a) English /i/ stimuli in the voiced consonant condition had durations of 262-275 ms, and English /i/ stimuli in the voiceless consonant condition and English /ɪ/ stimuli in both conditions had durations of 44-113 ms.

Table 2.4 Identification of English /i/ and /ɪ/ by Japanese listeners in Morrison (2002a). Identification rates given as percentages.

English Stimulus	Consonant Condition	English Response i	English Response ɪ	English Response ɛ
i	Voiced	95	5	
	Voiceless	27	72	
ɪ	Voiced		92	5
	Voiceless		93	5

In Morrison (2002a), Japanese participants did not always identify English /i/ stimuli before voiceless consonants as English /ɪ/. He hypothesised that this may have been the result of secondary use of spectral cues, i.e., if the duration was ambiguous, or the spectral properties were extreme, or the listeners' attention shifted for some reason, or there was a contrast effect with the vowel in the preceding stimulus, then the listeners would use spectral cues to identify the vowel instead of the duration cues which they normally used.

Morrison (2002a) suggested that the near 100% correct identification rate for English /i/ and /ɪ/ in Ingram & Park (1997) was not due to an ability to normalise for speaking rate differences as Ingram & Park had hypothesised. Rather, he proposed that the results were due to almost all the English /ɪ/ stimuli in Ingram & Park (1997) being short enough to be assimilated to Japanese /i/ and all the English /i/ stimuli being long enough to be assimilated to Japanese /iː/ (see Figure 2.1). This was so because they had used an /h_d/ word frame in which the vowels occurred before a voiced consonant. Morrison also noted that Ingram & Park had presented the words in isolation, a context in which there would be few cues to speaking rate even if the Japanese participants had the general ability to normalise for speaking rate.

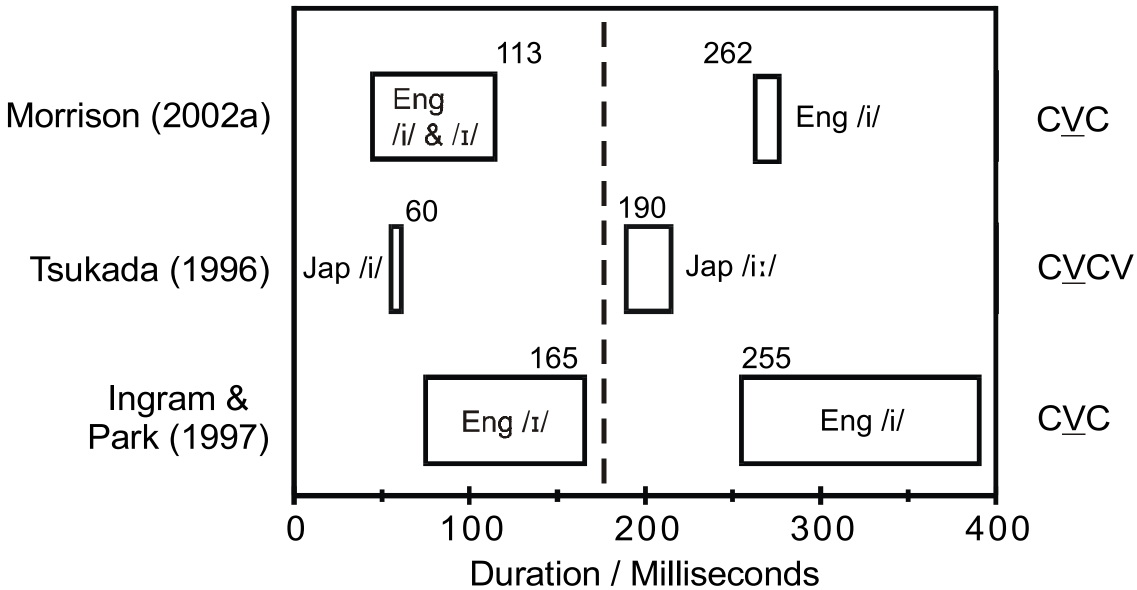


Figure 2.1 Adapted from Morrison (2002a). Duration ranges for English /i/ before a voiceless consonant and English /ɪ/ before voiced and voiceless consonants (left) and for English /i/ before a voiced consonant (right) from Morrison (2002a). Duration ranges for Japanese /i/ (left) and /iː/ (right) from Tsukada (1996); and for English /ɪ/ (left) and /i/ (right) from Ingram & Park (1997). The dashed line represents a categorical boundary that would account for the identification patterns of Japanese participants in Ingram & Park (1997) and Morrison (2002a).

Crowther & Mann (1992) found that Japanese learners of English produced a much smaller vowel duration difference (mean 26 ms) in the English words *pot* /pɑt/ and *pod* /pɑd/, than did native US English speakers (mean 76 ms). In their perception experiment, Japanese participants were affected more than native English participants when the closure and release burst of the final consonants in natural *pot* and *pod* stimuli were removed so that the only cues for voicing remained in the vowel. The Japanese participants' correct-identification scores, however, remained relatively high (78% in the edited condition compared to 89% in the unedited condition), indicating that they were able to extract consonant voicing information from the vowel. Their correct-identification score for the edited tokens was also significantly higher than for Mandarin participants. In a perception experiment using synthetic stimuli in which vowel duration varied, Japanese participants again outperformed Mandarin participants but did not perform as well as native English participants. Crowther & Mann (1992) suggested that Japanese and Mandarin participants made less use of vowel duration than English participants because of the lack of final consonants in their L1s and that the Japanese participants made more use of duration than the Mandarin participants because of experience with phonemic vowel duration in Japanese. It seems unlikely, however, that the Japanese participants used the Japanese long-short vowel contrast to distinguish the English consonant-voicing contrast: In the synthetic stimuli test, Japanese participants had a roughly linear relationship between vowel duration and percentage of voiced responses rather than an S-curve which would be expected for categorical perception (Coren, Ward, & Enns, 1994). In the production test, mean durations for the Japanese participants' English vowels before voiced versus voiceless consonants were 141 versus 116 ms; although it may be possible for the Japanese long-short vowel categorical boundary to fall within this range, the difference is not typical of either the ratio or absolute duration difference between Japanese long and short vowels (see review of literature above). An alternative interpretation of Crowther & Mann's suggestion is that the existence of phonemic duration distinctions in Japanese indirectly contributed to the Japanese participants' perception by giving them a general sensitivity to duration differences; however, this would be unexpected since it would suggest that Japanese listeners were sensitive to within-Japanese-category duration differences.

In summary, the studies reviewed above suggest that Japanese listeners identify English /i/ and /ɪ/ via assimilation to long and short Japanese vowels. English /ɪ/ is assimilated to Japanese /i/, English /i/ preceding a voiceless consonant is assimilated to Japanese /i/, and English /i/ preceding a voiced consonant is assimilated to Japanese /iː/. Further, that English vowels assimilated to Japanese /iː/ are identified as English /i/, and English vowels assimilated to Japanese /i/ are identified as English /ɪ/. These studies also suggest that Japanese listeners attend to duration differences at the onset of L2 English acquisition, but can be trained to attend to spectral differences.

2.3.3.2.3 Production of English vowels by Japanese speakers

When Ingram & Park (1997) had Japanese learners of English produce non-back Australian English vowels in isolated /h_d/ words, they observed considerable spectral overlap for /i/ and /ɪ/. A discriminant analysis with the normalised F1 and F2 values as independent variables correctly classified only 60% of /i/ tokens and 63% of /ɪ/ tokens. The Japanese speakers, however, clearly distinguished English /i/ and /ɪ/ using duration. This is compatible with the participants' two-category assimilation of English /ɪ/ and /i/ to Japanese /i/ and /iː/ (see section 2.3.3.2.1 above).

Kewley-Port, Akahane-Yamada, & Aikawa (1996) had native US English listeners identify English non-back vowels produced by Japanese learners of English. Vowels were spoken in isolation and embedded in /b_t/ words, and words were spoken in isolation and in a carrier sentence. Correct identification rates were higher for vowels in words than for isolated vowels (68%

compared to 58%), but there was no difference in correct-identification rates for words in sentences compared to isolated words. Identification rates for /i/ and /ɪ/ averaged over all conditions are shown in Table 2.5. Correct-identification rates for the Japanese speakers' English /i/ vowels were very high, but their English /ɪ/ vowels were primarily identified as /i/. In a two-way forced choice task, spectrally dissimilar vowel pairs (e.g., /i æ/) were correctly identified at higher rates than spectrally similar vowels (e.g., /ɪ e/). In contrast, vowel pairs of similar and dissimilar duration were both identified at relatively high rates. It therefore appeared that the English listeners primarily used spectral properties and made little use of duration when identifying the vowels. The Japanese participants had produced significant duration differences between English /i/ and /ɪ/, but had produced only small spectral differences. Kewley-Port et al. (1996) suggested that they had produced appropriate duration differences due to transfer of the Japanese long-short vowel contrast. Since the English listeners attended to spectral properties rather than duration, the Japanese speakers' use of duration to distinguish English /i/ and /ɪ/ was not effective.

Table 2.5 Identification of Japanese speakers' English /i/ and /ɪ/ by US English listeners in Kewley-Port, Akahane-Yamada, & Aikawa (1996). Identification rates given as percentages.

English Stimulus	English Response			
	i	ɪ	e	ɛ
i	97	1	2	
ɪ	64	28	6	1

Tsukada (1996, 1999) investigated the production of Australian English vowels by Japanese learners of English (participants were the same as in the Japanese vowel production tests described in section 2.3.3.1.2). Vowels were produced in isolated /C_t/ and /C_d/ words. The Japanese speakers produced a vowel duration difference before voiced versus voiceless vowels with a mean ratio of 1.25. This was smaller than the difference produced by native Australian English speakers (1.59) but greater than the difference they produced in Japanese (1.06). The English participants produced a greater duration difference for phonemic long English vowels, such as /i/, compared to phonemic short English vowels, such as /ɪ/, whereas the Japanese participants produced duration differences that were relatively similar across vowels. The durations of the Japanese and English high-front vowels are shown in Figure 2.2. The duration of the Japanese speakers' English /i/ productions were almost the same as their Japanese /iː/ productions; this would be consistent with assimilation of English /i/ to Japanese /iː/. The Japanese participants produced English /ɪ/ vowel durations that were quite similar to those produced by Australian English participants, but significantly longer than their Japanese /i/ and shorter than their Japanese /iː/. This suggests that English /ɪ/ may have been perceived as a new vowel.

With respect to /i/ and /ɪ/, Australian English is quite different to American and European English where spectral differences are of primary import. In a review of earlier work, Tsukada (1998, 1999) observed that spectral overlap is characteristic of Australian English /i/ and /ɪ/, and that duration is essential to distinguishing them. She also noted that Australian English /i/ is characterised by a long onglide which may also be a cue to vowel identity. Tsukada (2001) found less of an overlap for /i/ and /ɪ/ in spontaneous Australian English speech, consistent with the lax vowel being more susceptible to reduction than the tense vowel. Tsukada (1999) actually found that Japanese speakers produced greater spectral separation between /i/ and /ɪ/ than did native Australian English speakers. In Tsukada (1996, 1999), both Japanese speakers and Australian English speakers produced almost complete overlap in F1 and F2 for /i/ and /ɪ/. In addition, the native Australian English speakers' /ɪ/ before /d/ and /i/ before /t/ had similar durations: The mean durations were 153 and 163 ms respectively, and five of the twelve English participants produced an overlap in /ɪ/ and /i/ durations (similar results had been found by Peterson & Lehise, 1960; and Ichizaki, 1997, cited in Tsukada, 1999). One may propose that vowel duration may still be an effective cue to vowel identity for Australian English if it is a context sensitive cue: e.g., a high front vowel with a duration of around 155 ms will be identified as /i/ if it precedes a /t/, but as /ɪ/ if it precedes a /d/. Such an approach also requires that there be sufficient duration difference in vowels preceding voiced versus voiceless consonants: /ɪ/ preceding /t/ must be short enough that it not be confused with /i/ preceding /t/; /i/ preceding /t/ has the same duration as /ɪ/ preceding /d/; therefore /ɪ/ preceding /t/ must be short enough to be distinguished in duration from /ɪ/ preceding /d/. Since native Australian English speakers did maintain a relatively large vowel duration difference before voiced versus voiceless consonants, their vowels could be distinguished in the manner described above. Japanese speakers, however, produced relatively small duration differences, suggesting that they did not

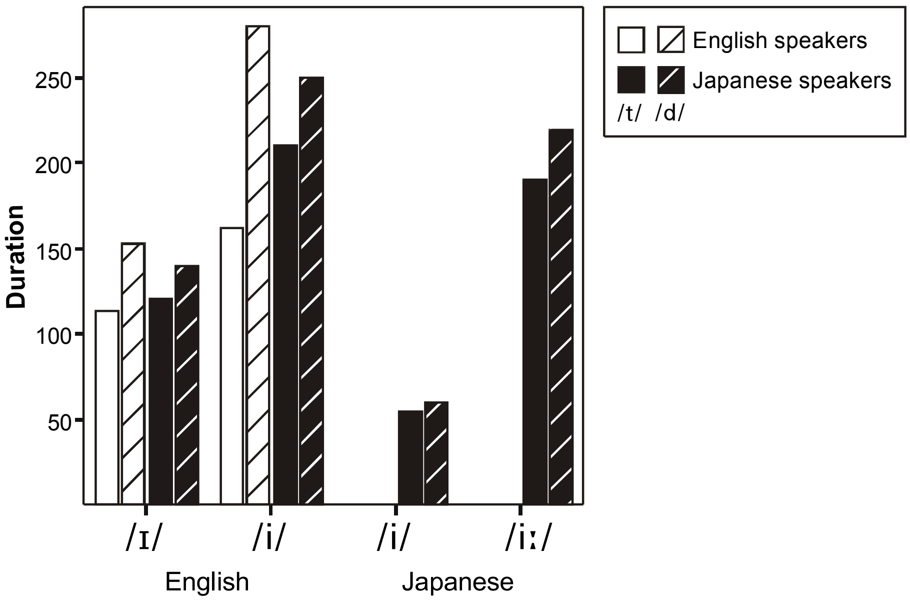


Figure 2.2 Durations (in milliseconds) of high front vowels produced in /C_t(o)/ (solid fill) and /C_d(o)/ (striped fill) context by Japanese (black fill) and English (white fill) speakers in Tsukada (1996).

perceive the English vowels in this way. The Japanese speakers' production pattern would instead be consistent with the perception pattern hypothesised by Morrison (2002a) (see Section 2.3.3.2.2 above): English /i/ preceding a voiced consonant is similar in duration to Japanese /iː/, resulting in productions that are intermediate in duration between the L1 vowel and L2 allophone; English /i/ preceding a voiceless consonant and English /ɪ/ preceding either a voiced or a voiceless are similar in duration to Japanese /i/, resulting in productions that are intermediate in duration between the L1 vowel and the three L2 allophones. This is consistent with the pattern seen in Tsukada's (1996, 1999) production data in Figure 2.2.[7]

Takahashi (1987) investigated the production of English vowel duration before voiced and voiceless fricatives. Japanese learners of English took part in a longitudinal study in which data was collected on five occasions over a period of ten months starting within a week of the participants' arrival in the US. In the first test, the Japanese speakers produced a vowel duration ratio before voiced versus voiceless fricatives of 1.14. This was smaller than the vowel duration ratio for native US English speakers (1.60), and for the vowel duration ratio for Japanese speakers' Japanese vowels

---

[7]The fact that the Japanese speakers' English vowels were longer than their Japanese vowels may in part be due to the difference in prosodic structure: CVC in English and CVCV in Japanese.

(1.23). Likewise, the Japanese speakers produced a voiced versus voiceless fricative duration ratio of 1.06 that was smaller than that produced by native English speakers (1.29), and that produced by the Japanese speakers for Japanese fricatives (1.32). The difference between the Japanese participants' English and Japanese production may have been related to the fact that the prosodic structures of the Japanese words were VCV and CVVCV whilst the English words were CVC. In terms of absolute duration, the Japanese speakers produced essentially the same vowel duration difference in English (26 ms) as they did in Japanese (27 ms). The absolute vowel durations of the Japanese speakers' English vowels were intermediate between the durations of their Japanese vowels and the native English speakers' English vowels. This suggests that the Japanese participants assimilated English vowels as similar to Japanese vowels and had developed a diaphone category.

There was no significant difference between the Japanese participants' vowel and fricative duration ratios in the first and final tests, and the non-significant difference was towards smaller ratios in the final test. There was therefore no evidence that, as a group, the Japanese participants had learnt to produce a more English-like vowel duration contrast before voiced versus voiceless fricatives. At the beginning of the study, one participant had a considerably lower score on an English proficiency test than the other participants. This participant produced more English-like duration ratios in the final test than in the initial test. It is therefore possible that the other participants had reached the most English-like state they would ever attain and *fossilised* (Selinker, 1972) prior to their involvement in the study.

In summary, the studies reviewed above indicate that Japanese speakers produce small spectral differences but large duration differences between English /i/ and /ɪ/. The duration contrast produced by Japanese speakers between English /i/ and /ɪ/ is consistent with them having established one diaphone category based on English /i/ preceding a voiced consonant, plus Japanese /iː/, and another diaphone category based on English /i/ preceding a voiceless consonant, plus English /ɪ/, plus Japanese /i/. Japanese speakers produce a small vowel duration contrast before voiced versus voiceless consonants and do not easily learn to produce an English-like contrast. Since US English listeners rely on spectral differences to distinguish /i/ and /ɪ/, and Australian English listeners distinguish /i/ and /ɪ/ on a context-sensitive basis that requires a large consonant voicing effect on vowel duration, Japanese speakers do not produce an effective contrast between the two vowels.

2.3.4 Spanish Speakers' vowel Perception and Production

2.3.4.1 Perception and Production of Spanish vowel duration

Several studies have found a duration difference for Spanish vowels preceding voiced versus voiceless consonants, but this difference is generally smaller than that for English. Navarro Tomás (1916) reported a duration ratio of 1.11 for /i/ before voiced versus voiceless fricatives in Peninsular Spanish. Marín Gálvez (1995) found a vowel duration ratio of 1.05 before voiced versus voiceless consonants in a read corpus of Peninsular Spanish. Quilis & Esgueva (1983) found that, in carrier sentences, Peninsular and American Spanish speakers' vowels were longer in b_bV(CV) words than in p_pV(CV) words by a ratio of 1.31. The larger ratio in this study is likely due to the fact that this is not only a post-vocalic voicing contrast: the preceding consonant also contrasted in voicing, and the following consonant contrasted in both voicing and plosive versus approximant manner, [b_β̞] versus [p_p]. Zimmerman & Sapon (1957) found a vowel duration difference before voiced versus voiceless consonants (including stop pairs realised as voiceless plosives and voiced approximants) that was much smaller for American Spanish than US English (18 versus 83 ms). Part of the difference may have been due to the fact that the prosodic structure was CVCV in Spanish and CVC in English, and that multiple vowels were included in Spanish test words but only /i/ was included in English.

2.3.4.2 Assimilation of English vowels to Spanish categories

Flege (1991b) had American Spanish listeners classify natural US English vowels (in isolated /bVt/ words) in terms of Spanish vowel categories. The results shown in Table 2.6 show that English /i/ was perceived as similar to Spanish /i/. Monolingual Spanish listeners primarily assimilated English /ɪ/ to Spanish /i/, but also gave a number of /e/ and "not a Spanish vowel" responses. This suggests a category-goodness assimilation in which English /i/ is a good match, and English /ɪ/ a poor match, for Spanish /i/. As Spanish speakers learn English they appear to realise that English /i/ and /ɪ/ are distinct: Compared to monolinguals, Spanish learners of English had a greater tendency to equate English /ɪ/ with Spanish /e/, or not equate it with any Spanish vowel. The most experienced learners also appeared to have realised that English /ɪ/ and Spanish /e/ are distinct;

they had a greater tendency to not equate English /ɪ/ with any Spanish vowel. Since the majority of English /ɪ/ stimuli were still assimilated to Spanish vowels, it did not appear that the Spanish participants had established a new English category. Since the number of "not a Spanish vowel" responses was relatively low, Flege concluded that there was no strong evidence that the participants had treated any English vowel as new.[8]

Imai, Flege, & Wayland (2002) presented Spanish listeners with natural Spanish and US English vowels (in isolated /bVt/ words), and within each trial had them classify and give goodness ratings in terms of Spanish vowel categories. There was no significant difference between the goodness rating given to English /i/ and Spanish /i/, indicating that Spanish listeners were not able to distinguish these vowels (the spectral properties of the two vowels were also very similar).

Table 2.6 Assimilation of English /i/ and /ɪ/ to Spanish vowel categories in Flege (1991b). Mean assimilation rates given as percentages. Spanish participants were grouped into a monolingual group and inexperienced and experienced learners of English. Figures not in bold are estimated from graphs in Flege (1991b).

		Spanish Response		
English Stimulus	Participant Group	i	e	not a Spanish vowel
i	Monolingual	**94**	5	
	Inexperienced	84	8	
	Experienced	88	3	9
ɪ	Monolingual	**68**	**19**	**12**
	Inexperienced	41	38	18
	Experienced	34	24	42

[8]This conclusion was based on the assumption that "new" vowels would be those that occurred in vowel space that was uncommitted in the L1; see Section 2.1.3.5 for an argument against this assumption.

2.3.4.3 Spanish Speakers' Identification of English vowels

Flege, Bohn, & Jang (1997) investigated the perception of US English /i/ and /ɪ/ in /bVt/ words by American Spanish learners of English. The Spanish participants were divided into two groups according to their level of experience with English. The participants' perception was measured using a synthetic /i/-/ɪ/ continuum which varied in spectral and duration properties. Spectral and temporal effect scores were calculated by subtracting the percentage of /i/ responses at one extreme of the continuum from the percentage of /i/ responses at the other extreme. If a Spanish participant had a reversal, i.e., gave more /i/ responses at the extreme where native English participants gave fewer /i/ responses, the Spanish participant's score was given a negative value. Table 2.7 shows mean spectral and temporal effect scores for the native English and Spanish groups. A large unsigned mean and small signed mean would indicate that members of the group had large effect scores but some members had reversals. As can be seen in Table 2.7, native English listeners identified vowels primarily on the basis of spectral cues. In contrast, Spanish participants gave approximately equal weight to spectral and duration cues and had a large number of reversals. On the basis of the inexperienced learners' unsigned scores, Bohn (1995) claimed that the Spanish listeners primarily relied on duration. The 6 point difference is, however, quite small and not statistically significant (standard deviations were 36 and 32 percentage points for spectral and temporal effects respectively). The results indicate that the Spanish participants' ability to distinguish English /i/ and /ɪ/ was poor. This would be compatible with a single-category or category-goodness assimilation of both English vowels to Spanish /i/. There was a great deal of individual variation, with some participants making considerable use of duration and little use of spectral cues, whilst other participants made considerable use of spectral cues and little use of duration cues. There was a trend for more of the experienced learners to make greater use of spectral than duration cues (6 of 10 for experience learners compared to 4 of 10 for inexperienced learners). This suggested that greater experience resulted in more English-like perception.

Table 2.7 Spectral and temporal effect scores (see text) for the identification of an English /i/ and /ɪ/ continuum by Spanish and English participants in Flege, Bohn, & Jang (1997). Spanish participants were grouped into experienced and inexperienced learners of English.

Participant Group	Spectral Effect		Temporal Effect	
	Signed	Unsigned	Signed	Unsigned
English	88	88	9	11
Spanish Experienced	-6	50	16	46
Spanish Inexperienced	13	44	46	50

Escudero (2000, 2001a) hypothesised that Spanish speakers learn the English /i/-/ɪ/ contrast in four stages: Stage 1, no contrast; Stage 2, exclusive use of duration; Stage 3, primary use of duration and secondary use of spectral properties; Stage 4, native-like primary use of spectral properties and secondary use of duration. She conducted an experiment using synthetic Scottish English /i/-/ɪ/ continua. The vowels occurred in /ʃVp/ words and both spectral and duration properties were varied (vowel durations were enhanced to cover the same range as Southern English vowels). There were a total of 30 participants from Spain and America. On the diagonal continuum (short duration and /ɪ/-like spectral properties, to long duration and /i/-like spectral properties) 21 Spanish listeners had a categorical boundary similar to that of native Scottish English listeners. The majority of the participants had perceptual patterns consistent with the developmental patterns hypothesised by Escudero; see Table 2.8. Four participants had reversals, i.e., identified long or spectrally /i/-like stimuli as /ɪ/. Since the study was not longitudinal it was not possible to demonstrate that the four patterns did indeed represent different developmental stages.

Scottish English makes almost no use of duration to distinguish /i/-/ɪ/. In contrast, Speakers from the south of England (hereafter “Southern English speakers”) produce relatively large duration differences and smaller spectral differences (Escudero & Boersma, 2001). Escudero (2001b) tested Southern English listeners’ perception of the synthetic /i/-/ɪ/ continua which had been used with the Scottish and Spanish listeners in Escudero (2000, 2001a). The relative weighting of spectral and duration cues for each group are shown in Table 2.9. Scottish listeners overwhelmingly used spectral cues to distinguish /i/ and /ɪ/, Southern English listeners primarily used spectral cues, and

Spanish listeners primarily used duration cues. Some English listeners, mostly Scottish English listeners, relied exclusively on spectral cues. As reported in Escudero (2000, 2001a) individual Spanish listeners had different perceptual patterns. The dialect of English that the Spanish listeners had been exposed to was found to be a possible factor affecting their use of perceptual cues in identifying /i/ and /ɪ/. The Spanish listeners who relied exclusively or primarily on duration had been exposed only to Southern English. Four (of six) Spanish listeners who relied exclusively on spectral properties had been exposed only to Scottish English. Escudero (2001b) therefore hypothesised two developmental sequences, one leading to spectral reliance and the other to duration reliance. She theorised that the developmental pattern that Spanish learners of English followed depended on the dialect of English to which they were exposed.

Table 2.8 Number of Spanish participants in Escudero (2000, 2001a) who fitted into hypothesised developmental stages for English high front vowel perception.

Stage	Number of Participants
1: No contrast	5
2: Exclusive use of duration	7
3: Primary use of duration and secondary use of spectral properties	6
4: Primary use of spectral properties and secondary use of duration	8
Reversals	4

Table 2.9 Relative weighting for spectral and duration cue use by Scottish English, Southern English, and Spanish participants in the identification of English /i/ and /ɪ/ in Escudero (2001b).

	Cue Weighting	
Group	Spectral	Duration
Scottish English	90	10
Southern English	70	30
Spanish	42	58

Flege, Munro, & Fox (1994) had Spanish and English listeners rate the degree of dissimilarity between English-English, Spanish-Spanish, and Spanish-English vowel pairs (US English and Mexican Spanish). For non-adjacent English-English and Spanish-English pairs (e.g., English /i/- English /ɑ/, English /ɪ/- English /ɑ/, and Spanish /i/- English /ɑ/) mean dissimilarity scores assigned by the English and Spanish listeners were not significantly different. This was as predicted because all groups were expected to perceive each member of the pair as a different L1 phoneme. For adjacent English-English and Spanish-English pairs (e.g., English /i/- English /ɪ/, English /e/- English /ɛ/, and Spanish /e/- English /ɛ/) mean dissimilarity scores assigned by the English participants were significantly greater than those assigned by the Spanish participants. This supported the prediction that the larger number of phonemic categories in English would cause the English participants to perceive greater dissimilarity between members of adjacent vowel pairs. Adjacent vowel pairs were expected to be perceived as different phonemes in English but not in Spanish. Comparison of Spanish /i/-English /i/ and Spanish /i/-English /ɪ/ results suggest that both Spanish and English listeners perceived Spanish /i/ as being more similar to English /i/ than to English /ɪ/. Although English experience did not appear to affect Spanish participants' perception of vowel dissimilarity, a correlation was found between perception of dissimilarity and an estimate of an individuals' overall proficiency in English pronunciation. This would be consistent with the more proficient learners of English having established new categories for English vowels.

Fox, Flege, & Munro (1995) used multidimensional scaling to analyse the data from Flege et al. (1994). They found that the English listeners used three dimensions to rate the dissimilarity of the vowels, whilst the Spanish listeners used only two dimensions. Both English and Spanish listeners had dimensions which correlated with vowel height and centrality in the vowel space. However, the English listeners' dimension which correlated with vowel height correlated more highly with vowel duration. The English listeners also had a dimension which correlated with vowel backness. Neither of the Spanish listeners' dimensions correlated with duration. This finding suggests that, contrary to the studies above, the Spanish listeners did not use duration to distinguish English /i/ and /ɪ/. This may, however, be an artifact of the experiment. Listeners were presented with a mixture of English and Spanish vowels and not instructed to attempt to differentiate them according to English criteria. Under such circumstances the listeners may have understood their task as rating the spectral dissimilarity of vowels, especially since their native vowel system does not include duration contrasts. Guion et al. (2002) found that almost half the Japanese participants in their study had understood that they were to differentiate vowels according to spectral differences even though the investigators had expected them to use both spectral and duration cues.

In summary, the studies reviewed above suggest that Spanish listeners at the outset of L2 English acquisition may be unable to distinguish English /i/ and /ɪ/. In the first developmental step in learning to distinguish the vowels, they may use duration differences alone. Later steps may involve a progressive increase in the use of spectral differences. The developmental pattern and ultimate weighting for the spectral and duration cues may be influenced by the cue weighting in the English dialect to which the Spanish listeners are exposed. The first two steps suggest a category goodness assimilation to Spanish /i/, with English /i/ as a good match, and English /ɪ/ as a poor match. The developmental pattern suggests that English /ɪ/ is identified as a new vowel category.

2.3.4.4 Spanish Speakers' Production of English vowels

Contreras Oller (1997) found that the mean F1, F3, and duration differences between English /i/ and /ɪ/ produced by Venezuelan learners of English were similar to the differences produced by native RP English speakers. The Venezuelan Spanish participants' mean F2 difference, however, was only 75 Hz compared to 250 Hz for the native English listeners. Their mean F1 values were also higher than those of native English speakers and close to F1 values for Spanish /i/ and /e/. This suggests that the Venezuelan Spanish participants had assimilated English /i/ and /ɪ/ to Spanish /i/ and /e/ respectively.

Hammond (1986) found that in Cuban Spanish speakers' spontaneous English speech, all occurrences of English /ɪ/ were mispronounced as /i/. This suggest that English /ɪ/ had been assimilated to Spanish /i/.

Flege, Bohn, & Jang (1997) investigated the production of US English /i/ and /ɪ/ in /bVt/ words by American Spanish learners of English. The Spanish participants were divided into two groups according to their level of experience with English. The participants read the stimulus words in a carrier sentence, and results were obtained by measuring the acoustic properties of the vowels and by having native English listeners identify the vowels. The identification results are shown in Table 2.10. Correct-identification rates were generally low. Compared to the inexperienced group, the experienced group had a higher correct-identification rate for /ɪ/ but a lower correct-identification rate for /i/. There was considerable spectral overlap between the Spanish participants English /i/ and /ɪ/, both vowels had mean spectral values close to native English speakers' /i/. The experienced learners' vowels were closer to native English speakers' /ɪ/ than were the vowels of the inexperienced learners. The perception and acoustic results are compatible with a single-category assimilation of English /i/ and /ɪ/ to Spanish /i/ resulting in a diaphone category which develops properties intermediate between the three vowels.

Table 2.10 Identification of Spanish participants' English /i/ and /ɪ/ by native English listeners in Flege, Bohn, & Jang (1997). Identification rates given as percentages. Spanish participants were grouped into experienced and inexperienced learners of English.

English Stimulus	Participant Group	English Response		
		i	ɪ	ɛ
i	Experienced	57	42	
	Inexperienced	69	24	7
ɪ	Experienced	39	61	
	Inexperienced	49	51	

Flege, Munro, & Skelton (1992) found that English speakers produced mean vowel duration differences in utterance final /CVt/ versus /CVd/ which were longer than those produced by experienced Spanish learners of English, which were in turn longer than those produced by inexperienced learners. Duration differences were 87 ms (48%), 55 ms (36%), and 37 ms (25%) respectively. Although not statistically significant, English speakers also produced a larger consonant closure duration difference than experienced Spanish learners of English, who in turn produced a larger consonant closure duration difference than inexperienced learners. Duration differences were 39 ms, 18 ms, and 7 ms respectively. Group results suggested that experienced learners had been partially successful at learning the English vowel and consonant duration distinctions which cue consonant voicing. There was a great deal of individual variation and there was evidence that some Spanish participants had learnt to produce duration distinctions within the range for the native-English norm.

McAllister, Flege, & McKay (1999) found evidence that Spanish speakers will not learn an L2 vowel duration contrast to L2 norms. They investigated the production and perception of long and short Swedish vowels by very experienced American Spanish learners of Swedish (at least 10 years residence in Sweden). Spectral differences are the most important cue in high and low Swedish long-short vowel pairs, but duration is the most important cue in mid vowels (McAllister, Flege, & McKay, 1999). Compared to native Swedish speakers, the Spanish participants produced only a small duration difference between long and short vowels (a ratio of 1.18, and an absolute difference of 26 ms for Spanish speakers, compared to 1.55 and 83 ms for Swedish speakers). Their ability to correctly identify long and short vowels was also considerably poorer: correct-identification scores were 62% for mid vowels and 74% for non-mid vowels compared to almost 100% for native Swedish participants. The Spanish participants' performance was better when they had access to a combination of duration and spectral cues as in non-mid vowels than when they only had access to duration cues as in mid vowels. Irrespective of which cues were available, the Spanish participants fell far short of reaching Swedish native norms.

2.4 Summary

The following issues raised in the literature review will be central to the thesis:

• Native English listeners have been found to distinguish English /i/ and /ɪ/ primarily on the basis of spectral cues (section 2.3.2).

• English /i/ is longer than English /ɪ/ in the same phonetic context. English vowels are considerably longer before voiced than before voiceless consonants (Section 2.2.1).

• Japanese listeners distinguish Japanese /iː/ and /i/ primarily on the basis of vowel duration (Section 2.3.3.1.1).

• Vowel duration differences due to post vocalic consonant voicing are small in Japanese compared to English (Section 2.3.3.1.2).

• Several studies have concluded that Japanese listeners assimilate English vowels to long and short Japanese vowel categories, and make use of the Japanese long short contrast when identifying English vowels (Section 2.3.3).

• Morrison (2002a) hypothesised that English /i/ before a voiced consonant is assimilated to Japanese long /iː/ but assimilated to Japanese short /i/ before a voiceless consonant; that English /ɪ/ before both a voiced and voiceless consonant is assimilated to Japanese short /i/; that English vowels assimilated to Japanese long /iː/ are identified as English /i/, and that English vowels assimilated to Japanese short /i/ are identified as English /ɪ/. This requires that the Japanese long-short vowel boundary fall between the longest duration for an English /i/ stimulus in the voiced consonant condition or for an English /ɪ/ stimulus in either condition (whichever be the greater), and the shortest duration for an English /i/ stimulus in the voiceless consonant condition. (Section 2.3.3.2.2)

• The above hypothesis claims that Japanese listeners' perception of English /i/ and /ɪ/ is influenced by the Japanese /iː/-/i/ long-short contrast. One form of evidence to support this claim would come from comparison with listeners whose L1 has a similar vowel system but lacks a long-short contrast, i.e., Spanish.

• Because of the Perceptual Magnet Effect (Section 2.1.5) and the Desensitisation Hypothesis (section 2.1.4), Spanish listeners are predicted to have difficulty distinguishing the spectral differences between English /i/ and /ɪ/ and are expected to use duration cues instead. Several studies have found results consistent with this prediction (Section 2.3.4).

• When identifying English /i/ and /ɪ/, if Spanish listeners' use of duration is different to Japanese listeners' use of duration, then this will be consistent with Morrison's (2002) hypothesis.

• The most rapid stage of learning for L2 sounds is expected to be during the first six months following arrival in the L2 society (Section 2.1.3).

• The Speech Learning Model predicts that L2 sounds that are perceived as similar to L1 sounds will form diaphone categories, the properties of which will eventually be intermediate between the norms for the L1 and L2 sound (Section 2.1.3.1). The results of several studies are consistent with Japanese listeners perceiving English /i/ and /ɪ/ as similar to Japanese /iː/ and /i/ (Section 2.3.3). Therefore, relatively little change the Japanese listeners' perception of English /i/ and /ɪ/ is expected during the first six months living in an English speaking society.

• The Speech Learning Model predicts that new categories will be formed for L2 sounds that are perceived as new, that sounds will be perceived as new if listeners perceive at least some of the differences between the L2 sound and the closest L1 sounds, and that the new categories will have properties close to the L2 norm for that sound (Section 2.1.3.2). Several studies have found results consistent with Spanish listeners assimilating English /i/ and /ɪ/ to Spanish via a category goodness assimilation to Spanish /i/, with English /i/ as a good match, and English /ɪ/ as a poor match (Section 2.3.4). Therefore, Spanish listeners are predicted to form a new category for English /ɪ/ and a relatively large change in their perception of English /i/ and /ɪ/ is expected during the first six months living in an English speaking society.

2.5 Research Questions and Hypotheses

The following hypotheses and research questions will be tested in this thesis.

Hypothesis 1: Japanese listeners with limited exposure to English will identify English /i/ and /ɪ/ according to the same duration criteria they use to identify Japanese /iː/ and /i/. English /i/ before a voiced consonant is assimilated to Japanese /iː/; and English /i/ before a voiceless consonant, and English /ɪ/ are assimilated to Japanese /i/. English vowels assimilated to Japanese /i/ are identified as English /ɪ/, and English vowels assimilated to Japanese /iː/ are identified as English /i/.

This is the hypothesis developed in Morrison (2002a) (see Section 2.3.3.2.2) and found to be consistent with the findings of Strange et al. (1998) on the assimilation of English vowels to Japanese categories (see Section 2.3.3.2.1), and the findings of Tsukada (1996, 1999) on the production of English vowels by Japanese speakers (see Section 2.3.3.2.3). The present study will test the first part of Hypothesis 1 via Research Question 1; an affirmative answer will be consistent with the hypothesis.

Research Question 1: Do Japanese listeners distinguish English /i/ and /ɪ/ using a duration-based categorical boundary in the same location as their categorical boundary between Japanese /iː/ and /i/?

Hypothesis 2. Experience with English will have little effect on Japanese listeners' perception of English /i/ and /ɪ/. The assimilation pattern in Hypothesis 1 results in two diaphone categories: (1) English /i/ before a voiceless consonant plus English /ɪ/ plus Japanese /i/; and (2) English /i/ before a voiced consonant plus Japanese /iː/ (see Figure 2.3). The properties of the diaphone categories will initially be close to the norms for the Japanese vowels, but will eventually be intermediate between the norms for the Japanese and English vowels which contribute to each diaphone. Since English /i/ and /ɪ/ allophones are assimilated to Japanese /iː/ and /i/ according to Japanese vowel duration criteria (Strange et al.,1998, see Section 2.3.3.2.1), English vowel allophones contributing to a diaphone have duration properties within the range for the Japanese vowel contributing to that diaphone. Hence, all vowel allophones contributing to this diaphone categories have the same duration properties as Japanese /iː/ or /i/, and the duration properties of the diaphone categories will continue to be identical to Japanese /iː/ and /i/. One diaphone consists of Japanese /iː/ and an allophone of English /i/; since these vowels have very similar spectral properties (Strange et al., 1998, see section 2.3.3.2.1; Kewley-Port, Akahane-Yamada, & Aikawa, 1996, see Section 2.3.3.2.3) there will be no change in the spectral properties of this diaphone. The other diaphone consists of Japanese /i/, an allophone of English /i/, and English /ɪ/; although English /ɪ/ differs spectrally from Japanese /i/, movement of the properties of the diaphone category towards English /ɪ/ will be tempered by English /i/ which is spectrally very similar to Japanese /i/. Hence the spectral properties of this diaphone category will remain close to Japanese /i/ and move only slightly towards English /ɪ/.

The present study will test Hypothesis 2 via Research Question 2; an affirmative answer will be consistent with the hypothesis.

Research Question 2: With increased exposure to English, will Japanese listeners categorical boundary between English /i/ and /ɪ/ maintain its initial duration properties and shift slightly towards more English-like use of spectral properties?

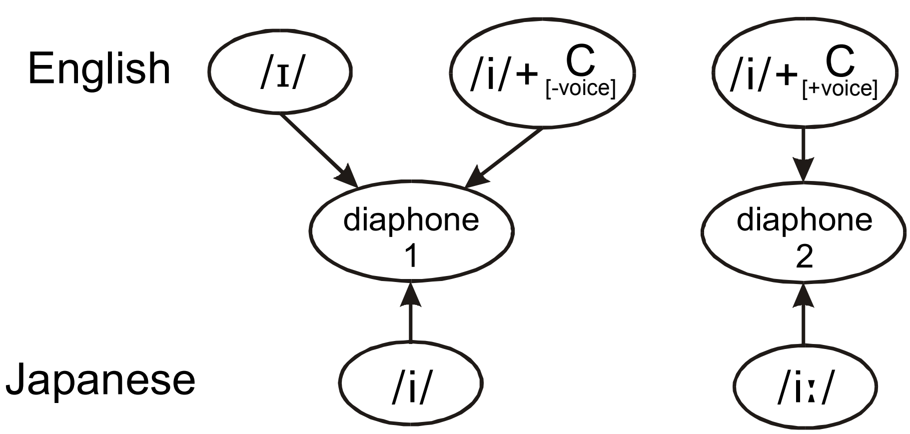


Figure 2.3 Hypothesised category formation for Japanese listeners.

Hypothesis 3. Spanish listeners with limited exposure to English will not perceive the English /i/-/ɪ/ contrast categorically. They will assimilate English /i/ and /ɪ/ to Spanish /i/ via a category-goodness assimilation pattern, with English /i/ as a good match for Spanish /i/ and English /ɪ/ a poor match (see Figure 2.4). Spanish listeners' ability to distinguish English /i/ and /ɪ/ will be poor.

The situation described in this hypothesis is the initial stage in the developmental pattern hypothesised by Escudero (2000, 2001a) (see Section 2.3.4.3). It is consistent with the findings of Flege, Bohn, & Jang (1997), and Flege, Munro, & Fox (1994) on the perception of English vowels by Spanish listeners (see Section 2.3.4.3), and with the findings of Flege, Bohn, & Jang (1997), on the production of English vowels by Spanish speakers (see Section 2.3.4.4). The present study will test Hypothesis 3 via Research Question 3; a negative answer will be consistent with the hypothesis.

Research Question 3: Do Spanish listeners have a categorical boundary between English /i/ and /ɪ/?

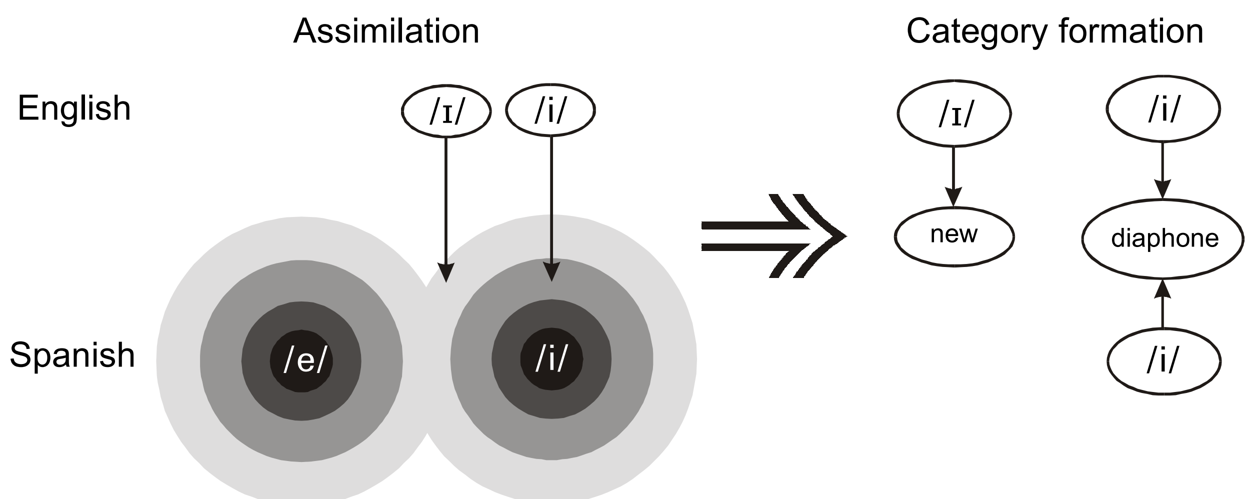


Figure 2.4 Hypothesised assimilation and category formation for Spanish listeners.

Hypothesis 4. Experience with English will improve Spanish listeners' ability to distinguish English /i/ and /ɪ/. Since English /i/ is very similar to Spanish /i/, these two vowels will form a diaphone category. Since English /ɪ/ is only a poor match for Spanish /i/, a new category will be established for English /ɪ/ (see Figure 2.4). The new category will develop first duration and then spectral properties that are close to the English norm for /ɪ/.

This hypothesis reflects the non-initial stages in the developmental pattern hypothesised by Escudero (2000, 2001a) (see Section 2.3.4.3). Following Flege's (1992) suggestion (see section 2.1.3), this category development is expected to occur primarily during the fist six months of massive exposure to English. The present study will test Hypothesis 4 via Research Question 4; an affirmative answer will be consistent with the hypothesis.

Research Question 4: During their first six months' of exposure to English in an English-speaking society, do Spanish listeners develop a categorical boundary between English /i/ and /ɪ/, and does this boundary match that of native English speakers?

Chapter 3

Methodology

3.1 Participants

Listeners were 7 native Canadian English speakers (CE), 7 Japanese learners of English (JP), and 5 Mexican Spanish learners of English (SP). Participants reported no hearing problems.

Canadian English participants (3 women and 4 men) were students and graduates recruited on the campuses of Simon Fraser University and the University of British Columbia. They had lived until age 16 in an Anglophone region of Canada west of Quebec, and immediately prior to data collection had lived for a period of at least 1 month in an Anglophone region of Canada west of Quebec. The group had a mean age of 30 (range 25-38).

Japanese and Spanish participants had lived until age 16 in Japan/Mexico. As children, they had not been immersed in a language other than Japanese/Spanish (e.g., they had not been brought up in English due to one or both parents being native English speakers, and they had not attended a school were English was the language of instruction). They had never lived in a non-Japanese/Spanish speaking country for a continuous period of more than three months, and had not lived in an English speaking country during the year prior to their arrival in Vancouver. They had lived in Vancouver for a period of less than 2 months prior to the first stage of data collection, and remained in Vancouver until the final stage of data collection approximately 5 months later. They had not received (and between the initial and final tests did not receive) pronunciation training on English vowels or the consonant-voicing effect.

Japanese participants (4 women and 3 men) were students attending the University of British Columbia on an exchange programme from Ritsumeikan University in Kyoto. Their academic programme included content-based language instruction, but no courses in English pronunciation. Four came from Nara, one from Osaka, one from Kyoto, and one from Kobe (all within the western dialect area). The group had a mean age of 20 (range 19-22). They had started studying English at age 13 or 14 and had studied English in school for 7-9 years.

Four of the Spanish speaking participants were students attending the University of British Columbia on an exchange programme from the Monterrey Institute of Technology (Instituto Tecnológico y de Estudios Superiores de Monterrey, ITESM) in Mexico. One participant was a former student at ITESM and an undergraduate student at Simon Fraser University. They all took regular undergraduate courses and did not take ESL classes. Three came from Mexico City and two from Monterey. The group consisted of 3 women and 2 men with a mean age of 20 (range 18-22). Four had started studying English at age 6 and one at age 4. They had studied English in school for 12-15 years.

Several studies (Flege, 1991b; Flege, Munro, & Skelton, 1992; Flege, Bohn, & Jang, 1997; Escudero, 2000, 2001a, 2001b) have used Spanish speakers from a number of countries; however, Godínez (1978) and P. Escudero (personal communication, 11 June 2001) found that there were considerable differences in the production and perception of Spanish vowels by speakers of different Spanish dialects. It was therefore decided to recruit Spanish speaking participants from a single country in order to minimise dialectal variation. Mexican Spanish was chosen because there were more Mexican participants available than participants of any other nationality.

Although the Japanese and Spanish groups were not matched for amount of English training in their home countries, they were matched for age of arrival and length of residence in Canada. Also, for both groups this was their first massive exposure to English.

3.2 Stimuli

Perceptual stimuli in the present study were constructed using edited natural speech. Edited natural speech allows for exact control over the acoustic properties of interest whilst holding all other acoustic properties invariant. The stimuli were based on the productions of two model speakers, a 34 year-old male monolingual English speaker who had lived most of his life in Vancouver, and a 38 year-old male native speaker of Japanese who had lived most of his life in Fukuoka. Recordings were made in a soundproofed room using a Sony MZS-R5ST Mini Disc recorder and a Sony ECM-MS907 microphone. The speakers read randomised lists of stimulus sentences at normal, slow, and fast speaking rates.

3.2.1 Target segments

English target words were “bit, beat, bid, bead” /bɪt bit bɪd bid/. These words were chosen to exemplify the English vowels /i/ and /ɪ/ in a stressed position followed by both a voiced and voiceless plosive. The consonants /d/ and /t/ following the target vowel were chosen since they both have the same place of articulation, differing only in terms of voicing, and allowed for the construction of real English words in combination with the target vowels.

The initial consonant of the word and the remainder of the carrier sentence were held constant so that any results attributable to phonetic differences in the stimuli could be due only to differences in the target segments. A carrier sentence was used since data obtained from citation forms may not be generalisable to continuous speech. As discussed in Chapter 2, the use of a carrier sentence, as compared to the presentation of target words in isolation, can have a significant impact on perception of the duration properties of stimuli (Strange et al., 1998).

Since responses to English and Japanese stimuli were to be compared, the segment immediately following the English target word had to match the parallel segment in the Japanese stimuli. Japanese consonants may be followed by vowels, but there is a tendency in Canadian English for intervocalic /t/ and /d/ to be flapped [ɾ], especially when the first vowel is stressed (Avis, 1975, Gregg, 1975; Wells, 1982). Pilot tests on potential model Canadian English speakers revealed that flapping eliminated the phonetic voicing distinction between /t/ and /d/, and the duration difference in the preceding vowel (contrary to the vowel duration findings of Fox & Terbeek, 1977 for US English). Because duration differences were of key interest to the present study a context which would induce flapping could not be used. Given Japanese phonotactics, the choice of segment following the target word was now highly restricted. /s/ was selected, resulting in voiceless affricated release for English /t/ and /d/. This corresponded to the Japanese affricate [t͡s] which is phonemically a single segment occurring as an allophone of Japanese /t/ before the high back vowel [ɯ]. It was not possible in Japanese to exactly match the English voiced alveolar plosive with voiceless affricated release, in Japanese both parts of the affricate are voiced, i.e. [d͡z].[1] The Japanese target words read by the model speaker were therefore “bitsu, biitsu, bittsu, bidzu, biidzu” exemplifying possible combinations of short and long vowels with voiced and voiceless singleton and geminate consonants. The model speaker read the long vowels with a high pitch, i.e. there were no pitch accent changes between the first and second mora. The words “biddzu, biittsu, biiddzu” were added as possible response categories in the perception test. The first word contains a geminated voiced consonant which may only occur in non-Chinese loan words (Itô & Mester, 1995,

[1]Although Japanese speakers often substitute [zɯ] for [d͡zɯ] (resulting in a merger of /zu/ and /du/) (Akamatsu, 1997), during the recording of the model sentences, the model Japanese speaker consistently pronounced the segments as [d͡zɯ].

e.g., *beddo* from English *bed*). The latter two words are of marginal acceptability in Japanese[2] but were included in the perception test since some of the stimuli had properties that combined typical long vowel and geminate consonant durations.

The English carrier sentence was "What they're wearing are beat suits." The carrier sentence put the target word in semantic and prosodic focus position. Most importantly, the carrier sentence was constructed in unison with a matching Japanese carrier sentence, "motteiru mono wa bitsuutsu" (What I have is a [nonsense word[3]]-suit). The English and Japanese sentences matched each other closely in terms of length, and certain acoustically similar events were synchronised: The non-release of the English /t/ before /ð/ in "what they're" parallelled prolonged closure of the Japanese geminate /t/ in "motteiru"; the English vowel in "are" /ɑ/ parallelled the low Japanese vowel /a/ in "wa"; the vowel of the final Japanese "tsu" was devoiced (Vance, 1987; Shibatani, 1990; Akamatsu, 1997) producing a similar acoustic/auditory effect to the affricated ending of the English "suits."

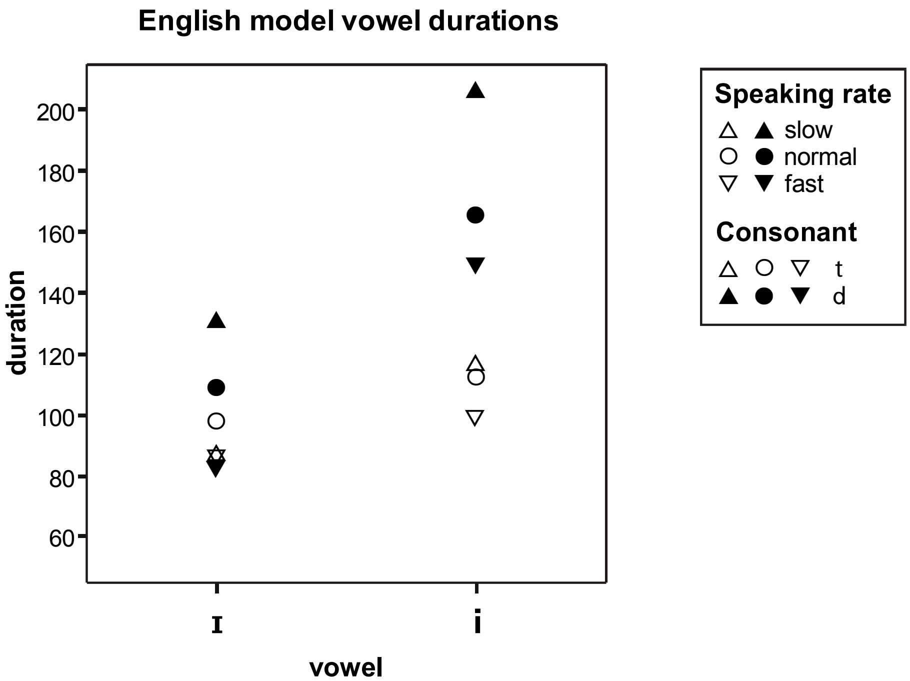
3.2.2 Acoustic properties of stimuli

The recordings of the model speakers were digitally transferred to computer and analysed using Praat (Boersma & Weenink, 2001). The investigator manually labelled the beginning and end points of the sentences, and the target vowels and consonants. An automated script then calculated the sentence, vowel, and consonant closure durations, and measured the first, second, and third formant values at the midpoint of the vowels using an LPC formant tracking algorithm. For each target word, the script displayed a spectrogram overlayed with the segment boundaries provided by the investigator and formant tracks calculated by the LPC tracking algorithm; this allowed the investigator to adjust algorithm parameters or take manual measurements if the automatic process did not produce formant tracks that coincided with the formants visible in the spectrogram. Duration and spectral properties for the model segments and sentences are given in Figures 3.1 through 3.3, and Table 3.1.

---

[2]The long vowel before a voiceless geminate is possible, e.g., *tootta* 'passed,' but rare. Vance (1987) pp. 72-73 notes that a long vowel in this context may shorten with the result that it does not contrast phonetically with a short vowel.

[3]Matching the phonetic context of the Japanese stimuli to the phonetic context of real words in English was given priority over using real Japanese words.



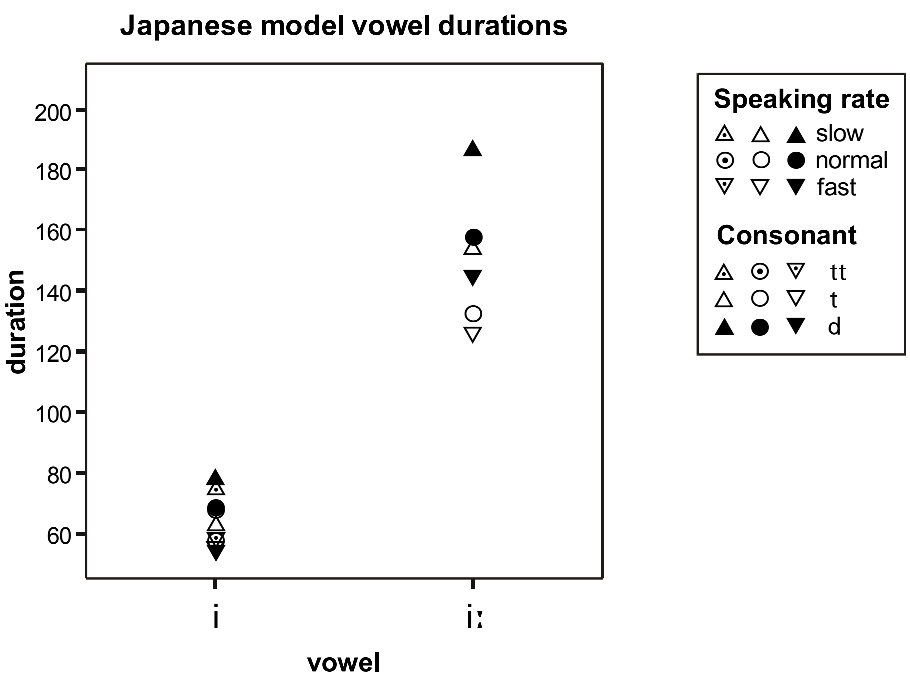


Figure 3.1 Mean vowel durations (in milliseconds) from model speakers' production. Sentences were read at slow (upright triangles), normal (circles), and fast (inverted triangles) speaking rates. The consonants following the vowels were /d/ (filled shapes), /t/ (non-filled shapes), and /tt/ (non-filled shapes with dots).

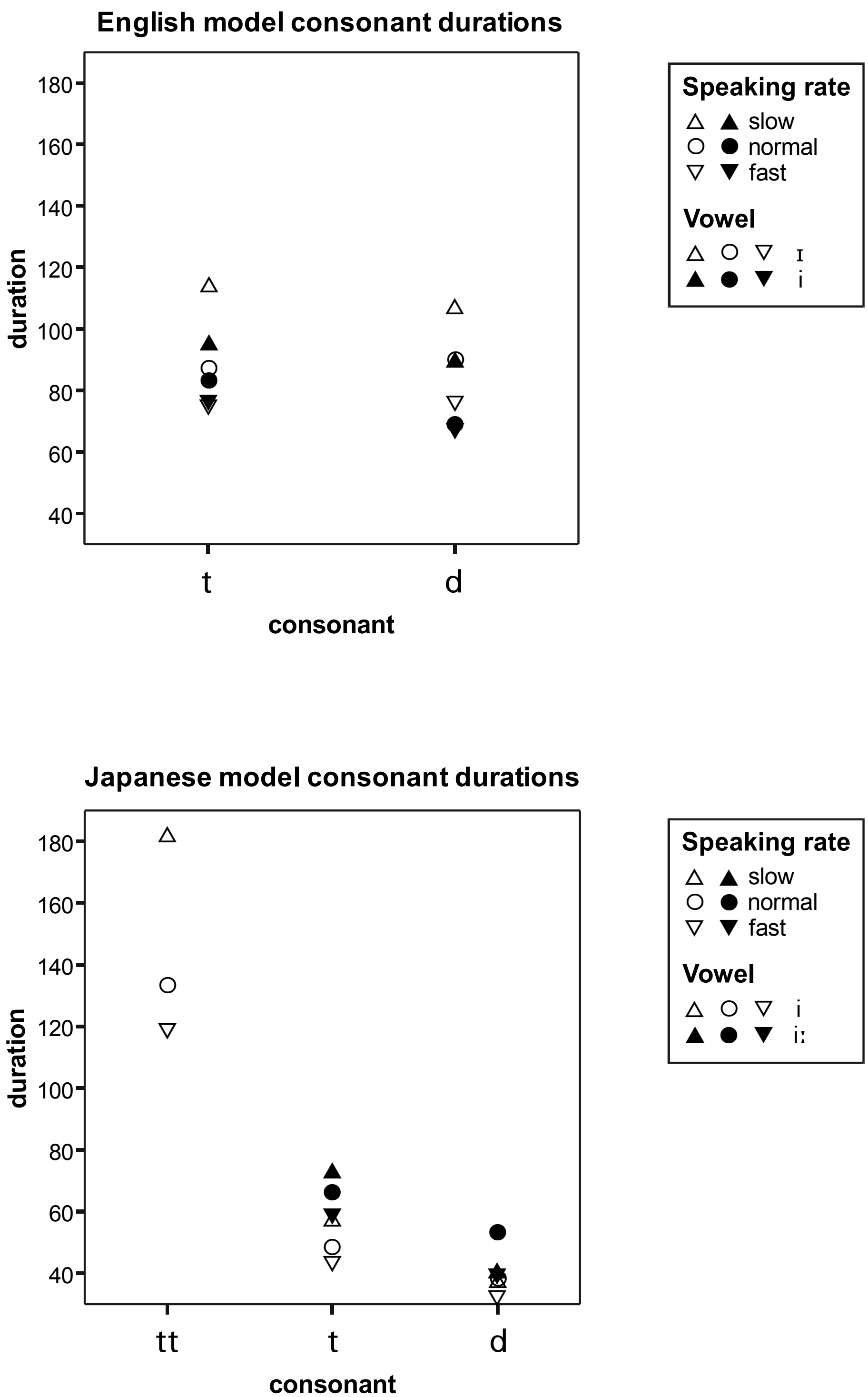


Figure 3.2 Mean consonant closure durations (in milliseconds) from model speakers' production. Sentences were read at slow (upright triangles), normal (circles), and fast (inverted triangles) speaking rates. The vowels preceding the consonants were English /i/ (filled shapes) and /ɪ/ (non-filled shapes), and Japanese /iː/ (filled shapes) and /i/ (non-filled shapes).

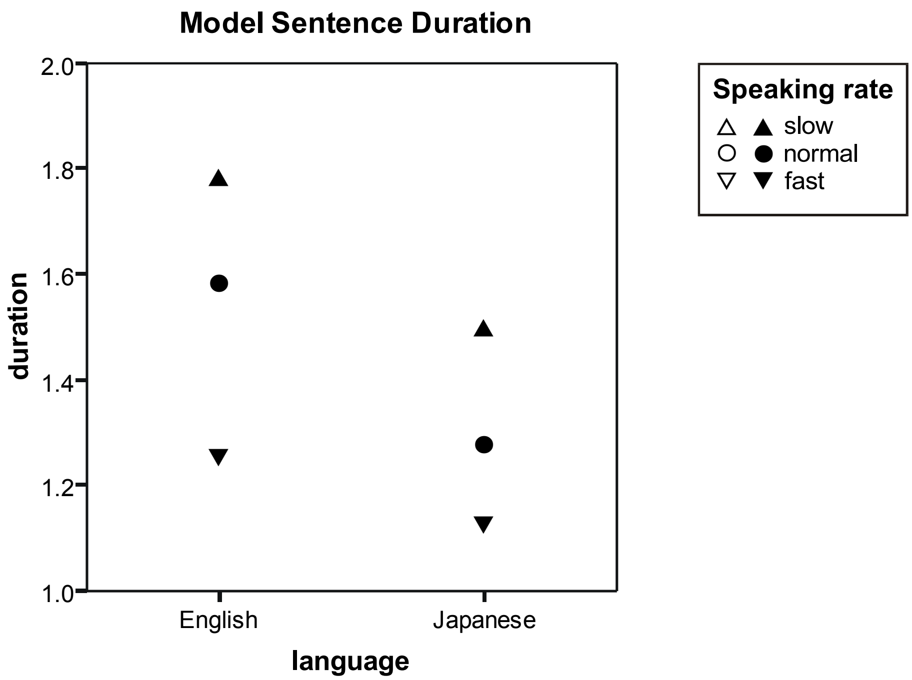


Figure 3.3 Model speakers' mean sentence durations (in seconds) excluding the durations of target segments. Sentences were read at slow (upright triangles), normal (circles), and fast (inverted triangles) speaking rates.

Table 3.1 Mean spectral values (in Hertz) measured at the mid-points of the target vowels produced by the model English speaker.

		Vowel			
		ɪ		i	
		consonant		consonant	
Formant	Speaking rate	t	d	t	d
F1	slow	410	372	250	253
	normal	413	371	256	256
	fast	406	369	257	256
F2	slow	1758	1819	2262	2290
	normal	1748	1821	2240	2275
	fast	1707	1761	2253	2253
F3	slow	2466	2567	2896	2963
	normal	2464	2583	2902	2896
	fast	2492	2605	2888	2876

A multidimensional continuum was created in which acoustic properties varied along the following dimensions: vowel spectra (5 points), vowel duration (7 points), plosive closure duration (5 points), speaking rate (2 points). The dimensions can be visualised as the cuboids shown in Figure 3.4:

- Vowel duration (durv) is on the front-to-back axis. Point 1 has the shortest vowel duration and point 7 the longest.
- Vowel spectral properties (spec) are on the right-to-left axis. Point 1 has /ɪ/-like spectral properties and point 5 has /i/-like spectral properties.
- Plosive closure duration (durc) is on the bottom-to-top axis. Point 1 has the shortest consonant closure duration and point 5 the longest.
- Speaking rate (spd) is represented by the series of two cuboids. The cuboid on the left has the slow speaking rate and the one on the right the fast speaking rate.

Typical values for the model English speaker's /bɪt bit bɪd bid/ are shown as the triangles in Figure 3.4.

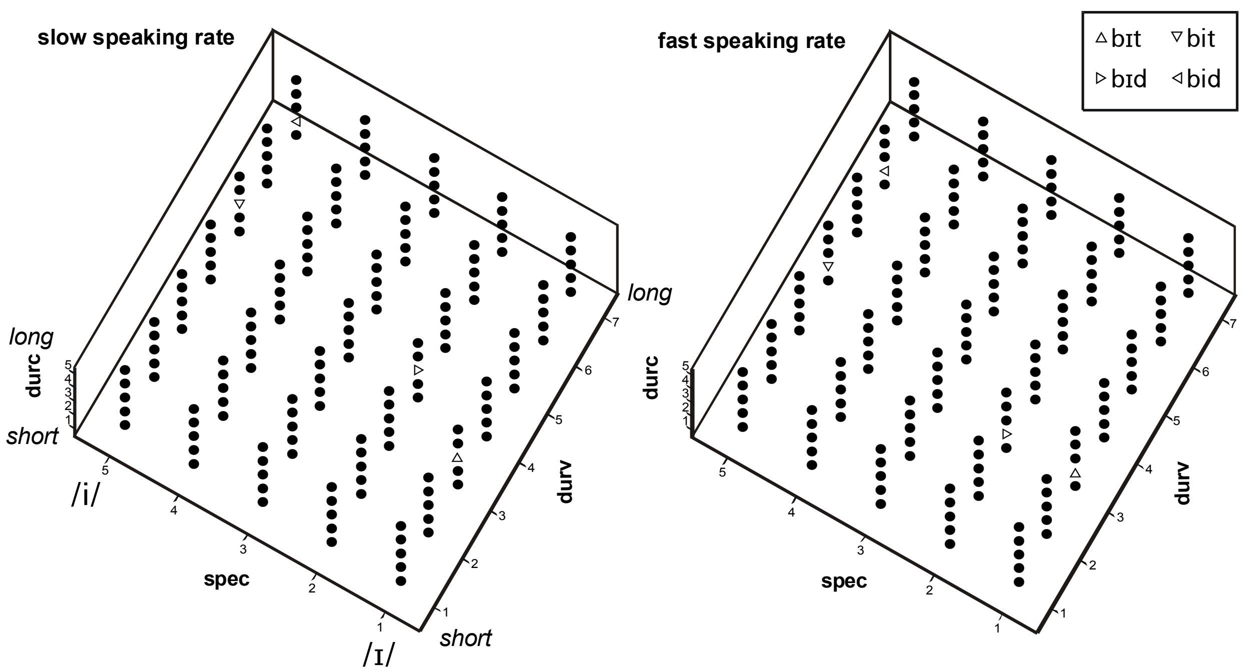


Figure 3.4 Cuboids representing the dimensions along which the English stimuli vary. Vowel duration (durv) from front to back: 1 is shortest and 7 longest. Vowel spectral properties (spec) from right to left: 1 is most /ɪ/-like and 5 most /i/-like. Consonant closure duration (durc) from bottom to top: 1 is shortest and 5 longest. Speaking rate: left cube is slow and right cube is fast. Dots represent individual stimuli. The triangles represent the position of typical "bit, beat, bid, bead" produced by the model speaker.

Table 3.2 Spectral values at the midpoints of the vowels in the English stimuli.

Endpoint vowel	Dimension point	F1		F2		F3	
		Hz	mel	Hz	mel	Hz	mel
ɪ	1	410	496	1700	1433	2465	1793
	2	370	454	1834	1503	2569	1836
	3	330	412	1974	1572	2676	1878
	4	292	370	2121	1642	2786	1921
i	5	255	328	2275	1712	2900	1964

The F1, F2, and F3 values at the midpoints of the vowels in the English stimuli are given in Table 3.2. The spectral values of the endpoints were the mean spectral values from the model English speaker's /i/ and /ɪ/. There were no differences in F1 offset other than those associated with differences in the F1 frequency over the whole vowel. The five points were evenly spaced on the mel scale (steps of 42 mel for F1, 70 mel for F2, and 43 mel for F3). The mel scale was used since it has a linear relationship to human frequency perception whereas the Hertz scale does not (Harrington & Cassidy, 1999). The mel - Hertz conversion formulae are:

$$f_{mel} = (1000 / \log 2) \log(f_{Hz} / 1000 + 1)$$
$$f_{Hz} = 1000 (10^{fmel / (1000 / \log 2)} - 1)$$

The vowel and consonant closure durations used in the English and Japanese stimuli are given in Table 3.3. The range of durations for the English stimuli was actually expanded to cover the same range as the Japanese durations. The values of the duration endpoints were based on the mean values from contexts resulting in the longest and shortest vowels and consonants produced by the Japanese model speaker. The duration points were evenly spaced on a logarithmic scale. This method of dividing the duration continuum was used by Escudero (2000, 2001a, 2001b) predicated on the concept that duration is perceived logarithmically (Turk, cited in Escudero, 2000). Conversions between stimulus point values and milliseconds can be made using the following formulae:

vowel duration	$t = 60 \times 1.2^{(P-1)}$	$P = \log_{1.2}(t / 60) + 1$
consonant closure duration	$t = 40 \times 1.45^{(P-1)}$	$P = \log_{1.45}(t / 40) + 1$

where t = time in milliseconds P = dimension point value

Differences in the English target consonants consisted only of differences in silent closure duration. No phonetic voicing was included during the closure.

Table 3.2 Durations (in milliseconds) of the vowels and consonants in the English and Japanese stimuli.

Dimension point	1	2	3	4	5	6	7
Vowel duration	60	72	86	104	124	149	179
Consonant closure duration	40	58	84	122	177		

Comparison of the model sentence durations revealed that the sentences spoken at normal speaking rate by the Japanese speaker were approximately the same duration as those spoken quickly by the English speaker (see Figure 3.3). Likewise the sentences spoken slowly by the Japanese speaker were most similar in length to those spoken at normal speaking rate by the English speaker. Stimulus sentence durations were therefore the means of the normal-rate Japanese and fast English pair, and the slow Japanese and normal-rate English pair. Excluding the duration of the target segments, sentence durations for the slow and fast speaking rates in the stimulus sentences were 1.541 and 1.265 seconds respectively.

Multidimensional Japanese continua were created with the same duration properties as the English continuum. The Japanese continua had no spectral dimension since Japanese high front vowels differ only in duration. The spectral values for the Japanese stimuli (F1, F2, F3 = 318, 2126, 3115 Hz) were those of a model vowel which had spectral values close to the mean values across all vowels, consonant conditions, and slow and normal speeds. Since Japanese phonemic voicing is always cued by phonetic voicing, two continua were created, one with, and one without phonetic voicing. The dimensions can be visualised as the rectangular matrices shown in Figure 3.5:

- Vowel duration (durv) is on the left-to-right axis. Point 1 has the shortest vowel duration and point 7 the longest.
- Consonant closure duration (durc) is on the bottom-to-top axis. Point 1 has the shortest consonant closure duration and point 5 the longest.
- Consonant voicing (vce) is represented by the horizontal series of rectangles. The rectangles on the top are voiceless and those on the bottom voiced.
- Speaking rate (spd) is represented by the vertical series of rectangles. The rectangles on the left have the slow speaking rate and those on the right the fast speaking rate.

Typical values for the Japanese model speaker's "bittsu, bitsu, biitsu, bidzu, biidzu" are shown as the triangles in Figure 3.2.

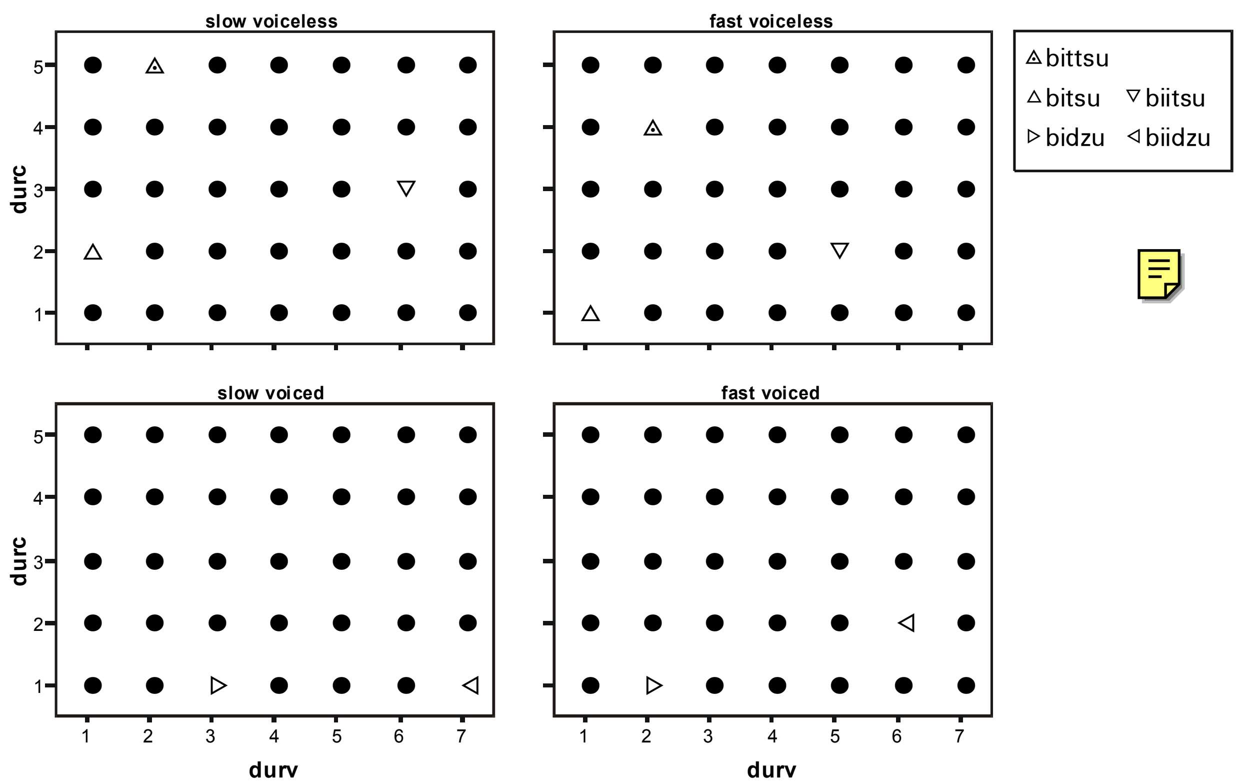


Figure 3.5 Rectangular matrices representing the dimensions along which the Japanese stimuli vary. Vowel duration (durv) from left to right: 1 is shortest and 7 longest. Consonant closure duration (durc) from bottom to top: 1 is shortest and 5 longest. Speaking rate: rectangles on the left are slow and those on the right fast. Consonant voicing: rectangles on the top are voiceless and those on the bottom voiced. Dots represent individual stimuli. The triangles represent the position of typical "bittsu, bitsu, biitsu, bidzu, biidzu" tokens produced by the model speaker.

3.2.3 Stimuli editing procedures

The edited natural speech stimuli were created using Cool Edit Pro LE (Johnston, 1999) and Praat (Boersma & Weenink, 2001) software. A single English sentence was chosen from which to create all the English stimuli. This sentence was spoken at normal speed and the target word was "beat." The target vowel had a clear steady state with smooth formant tracks as shown in Figure 3.6. The target vowel and consonant were extracted from the carrier sentence. A glottal source and a formant filter were derived from the target vowel using LPC analysis. The formants of the filter were raised or lowered to the midpoint values given in Table 3.2 (formants were raised or lowered by the same amount over the whole extent of the vowel, i.e., no special treatment was given to onset and offset transitions). The source and modified filters were then combined to produce vowels with the desired spectral properties. LPC echos were removed and the vowel intensity envelope was corrected to match that of the original vowel. Vowel and sentence duration were modified using the "manipulation" feature of Praat. The resulting vowels were deemed acceptable if their duration and spectral properties were within 2.5% of the values given in Tables 3.2 and 3.3. Consonant closure durations consisted of periods of silence. A Praat script was written which automatically concatenated the carrier sentences and target segments.

The Japanese stimuli were based on a Japanese sentence spoken at normal speed and containing the word "biitsu." The target vowel had a clear steady state with smooth formant tracks as shown in Figure 3.6. The latter part of the voiced consonant series was based on a sentence spoken at normal speed containing the word "biidzu." Stimulus editing procedures were the same as for the English stimuli except that there was no spectral modification of the vowel and the voiced consonant closure was based on the duration-modified closure extracted from "biidzu."

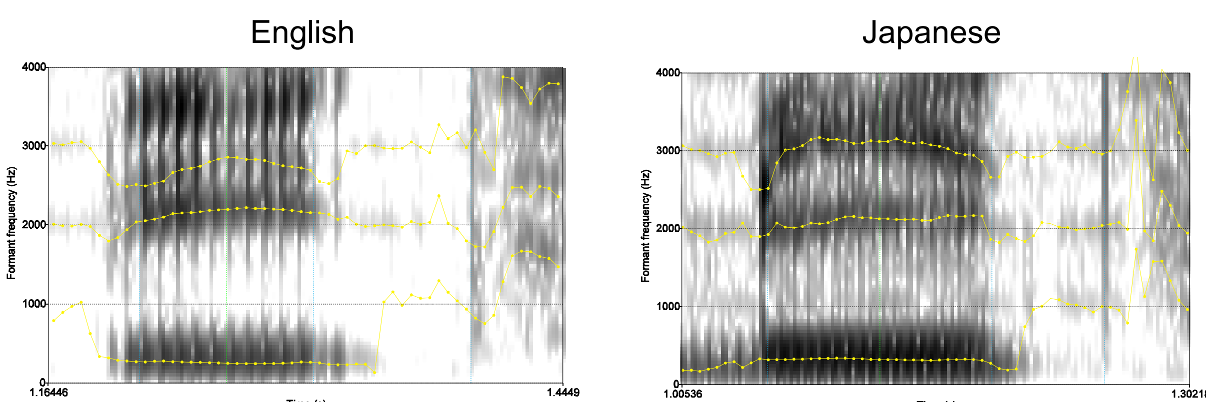


Figure 3.6 Spectrograms of original vowels and consonant closures used to create stimuli. Left: English "beat." Right: Japanese "biitsu."

3.3 Data Collection

Canadian English participants were tested once. Japanese and Mexican-Spanish participants were tested on English perception approximately one month after their arrival in Canada and again approximately five months later. Japanese participants were also tested on Japanese perception approximately one month after their first English test.

Participants were tested one at a time in a soundproofed room. English perception testing took approximately 30 minutes (preparation and a production test conducted during the same session took approximately 30 minutes for the initial test and 15 minutes for the final test). Participants listened to the English stimuli presented in random order via MEDS computer software (Kendall, 2001) over Optimus HP340 headphones; volume was adjusted to a comfortable level by the listener. Listeners used a mouse to respond to each stimulus by clicking on one of five pictures on the computer screen shown in Figure 3.7. Four of the pictures represented the words "bit, beat, bid, bead" (see below for details). The fifth picture, an "X," was a null response which the participants were instructed to use if they heard a word other than one of the four target words. Participants were told that almost all the words should sound like one of the four target words. They were instructed to choose the picture that was closest to the word they heard, and were encouraged to guess if not sure. Instructions were given in English both orally and in written form. The computer played a stimulus once and did not proceed until the participant had given a response.

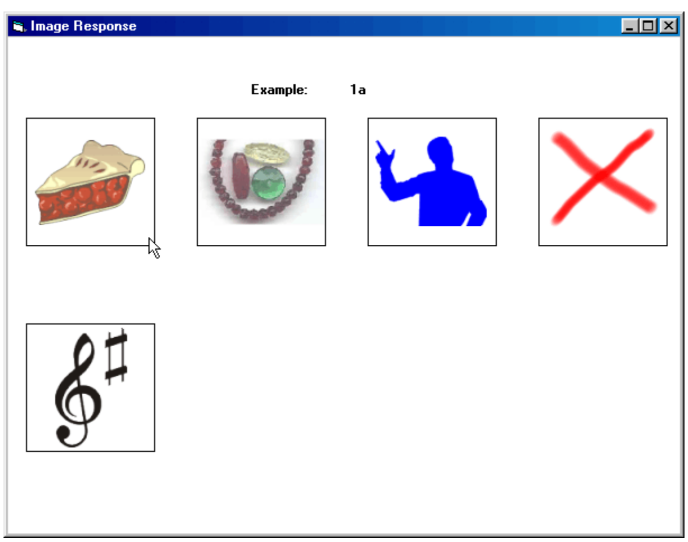


Figure 3.7 Screen shot of English perception test in MEDS software. The order of the pictures was assigned randomly in each trial.

There was a 500 ms pause between the response and the beginning of the next trial. The order of the pictures was assigned randomly for each trial. This prevented participants from associating raw acoustic properties in the stimuli with locations on the screen. A subset of 24 stimuli taken from different parts of the continuum was used as a warm-up before presenting the full set of 350 stimuli, each of which was identified once.

Pictures were used as response tokens in order to avoid confusion which may have arisen from orthographic representations of the target words: The letter “i” in English “bit” and “bid” represents English /ɪ/ but in Spanish and romanised Japanese orthography “i” represents vowels which are closer to English /i/. Participants were trained to interpret the pictures before the perception test. The investigator used the preliminary picture key shown in Figure 3.8 to elicit the words “ship” and “sheep.” The investigator then moved on to the main picture key shown in Figure 3.9 which contained a picture of a “bit” of pie, a musical “beat,” someone making a “bid” at an auction, and some “bead”s. Underneath each picture was written “b_t” or “b_d” as appropriate. The picture keys used with Japanese and Spanish participants had a translation of the target word beside each picture. The investigator elicited the target words from the participants using sentences such as: “this is a piece of pie, or a something of pie,” and “when someone wants to buy something at an auction, they make a . . .” If participants had difficulty guessing the correct words, the investigator pointed out what the first and last consonant of each word was, and indicated which words had the same vowels as one of the words from the preliminary picture key. At no point did the investigator say the target words, or the target vowels or consonants in isolation. Before the perception test, participants took part in a production test which involved reading randomised lists of target sentences in which the target words were replaced by the pictures. The results of the production test will be reported elsewhere.

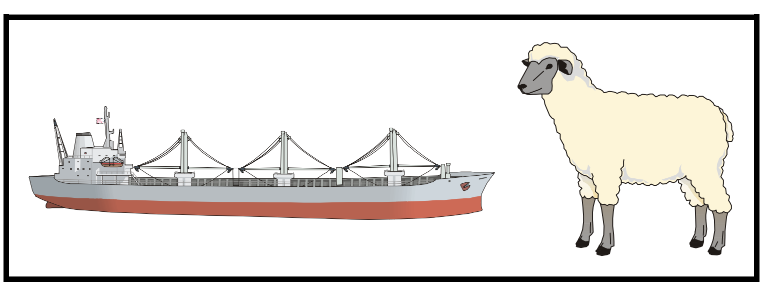


Figure 3.8 Preliminary picture key. Pictures of a *ship* and a *sheep* exemplifying the English vowels /ɪ/ and /i/.

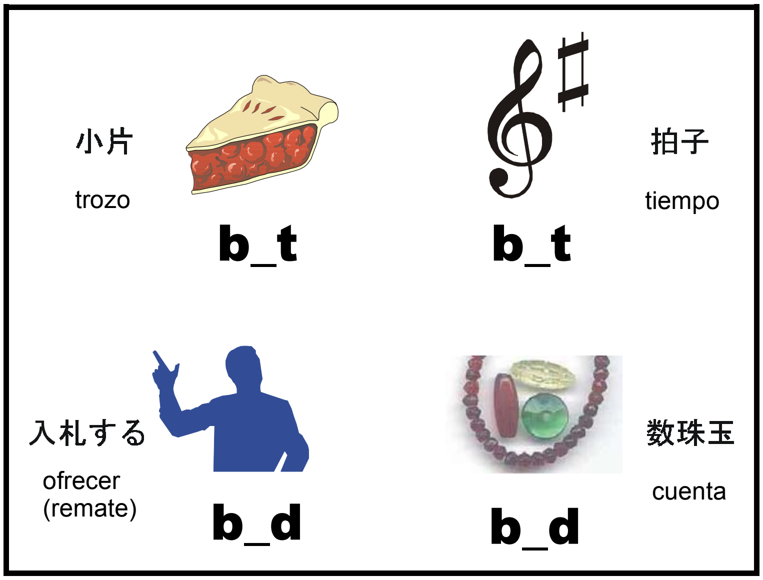


Figure 3.9 Main picture key. Pictures exemplifying the English words *bit*, *beat*, *bid*, and *bead* /bɪt, bit, bɪd, bid/.

The Japanese perception test was conducted using the same procedures as the English test except for the following differences: The eight response options "bittsu, biitsu, bitsu, biittsu, biddzu, bidzu, biidzu, biiddzu" written in katakana appeared on the computer screen shown in Figure 3.10. The order of pictures was systematic and fixed rather than randomised between trials. Randomisation may have caused unnecessary confusion since there were a relatively large number of response categories. The Japanese participants would have been aware of the systematic relationship between the Japanese response categories, but as native speakers were expected to identify the tokens phonemically rather than according to raw acoustic properties. The smaller number of stimuli in the Japanese continua meant that multiple identifications of each stimulus item could be obtained without fatiguing the listeners. The full set of 140 stimuli were presented as a series of two randomisations. Japanese perception testing took approximately 20 minutes (preparation and a production test conducted during the same session took approximately 20 additional minutes).

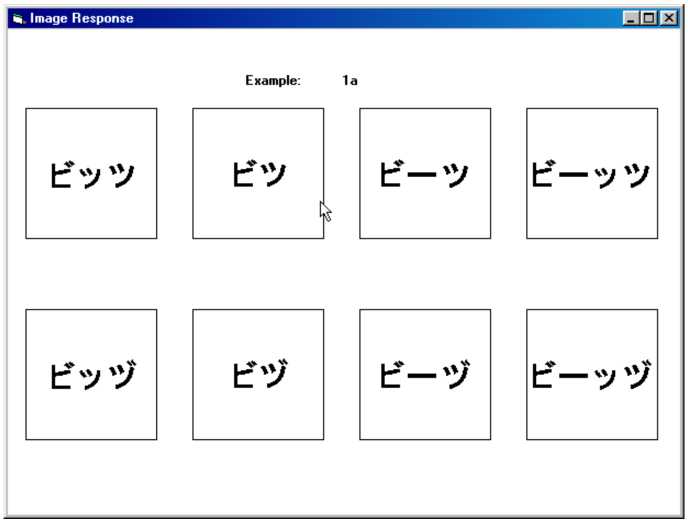


Figure 3.10 Screen shot of Japanese perception test in MEDS software. Response items are written in katakana.

Chapter 4

Results and Discussion

Only the results for the vowel perception tests are reported in this thesis. Results from consonant perception tests and production tests conducted as part of the larger study will be reported elsewhere. A full set of individual vowel perception results is given in the appendix.

4.1 Statistical Analysis Procedures

Research questions from Chapter 2 are repeated here for convenience.

Research Question 1: Do Japanese listeners distinguish English /i/ and /ɪ/ using a duration-based categorical boundary in the same location as their categorical boundary between Japanese /iː/ and /i/?

Research Question 2: With increased exposure to English, will Japanese listeners categorical boundary between English /i/ and /ɪ/ maintain its initial duration properties and shift slightly towards more English-like use of spectral properties?

Research Question 3: Do Spanish listeners have a categorical boundary between English /i/ and /ɪ/?

Research Question 4: Do Spanish listeners develop a categorical boundary between English /i/ and /ɪ/, and does this boundary match that of native English speakers?

In order to address these research questions it was necessary to determine whether each language group perceived a categorical distinction between English /i/ and /ɪ/. It was also necessary to determine the relative weighting of acoustic cues used to perceive this distinction. Of particular interest was a comparison of the cue weighting used by each language group and whether the cue weighting used by the Japanese and Spanish participants was more English-like in the final test than in the initial test. Also of interest was whether the categorical boundary for the Japanese participants' perception of English /i/ and /ɪ/ was in the same location as the categorical boundary between their Japanese /iː/ and /i/ perception.

Discriminant analyses were applied to each groups' data in order to determine the position of categorical boundaries between phonemes. The procedure described below (based on information in SPSS, 1999; Brown & Wicker, 2000; and Stevens, 2002) was carried out using SPSS software.

A discriminant analysis takes cases which are labelled as belonging to known groups, and estimates the coefficients of predictor variables that best characterise the differences between the groups. The linear combination of variables is known as a discriminant function. A single discriminant function describes the difference between two groups. Given predictor variables x, y, ..., n a discriminant function value z is calculated as follows:

$$z = C_x x + C_y y + ... + C_n n + C_{constant} \tag{1}$$

where C_x, C_y, ..., C_n and $C_{constant}$ are the coefficients generated by the analysis

The analysis also calculates the within-group means which are the values of the discriminant function at the group centroids. Let z_α be the value of the discriminant function at the centroid of group α, and z_β be the value of the discriminant function at the centroid of group β. If, for given values of x, y, ..., n the value of z is closer to z_α than z_β, then the model derived from the discriminant analysis will classify a case with those x, y, ..., n values as a member of group α.

Of interest in the present study is the categorical boundary between the two groups, which in the example above this may be defined as the line described by x, y, ..., n values such that the z value is exactly half way between z_α and z_β. For a model with two variables, this is the line[1] such that the probability that the case is a member of group α is 0.5 (and 0.5 for group β). To obtain an indication of the sharpness of the categorical boundary between the groups, lines can be calculated for x, and y combinations with 0.1 and 0.9 probability of being assigned to group α (0.9 and 0.1 for group β). A small distance between the 0.1 and 0.9 probability lines would indicate a sharp boundary (assuming that a linear model is appropriate), a large distance would indicate a fuzzy boundary. In the present study, listeners' perception will be deemed categorical only if the distance between the 0.1 and 0.9 probability points/lines/planes is less than half the range of spectral and duration properties tested (i.e. less than 3 for *durv* [range 1 to 7], and less than 2 for *spec* and *durc* [range 1 to 5]. The 0.1 and 0.9 probabilities were chosen as values which would clearly characterise robust identification of a given category whilst allowing for random errors due to non-linguistic

---

[1] a point for a model with one variable, a plane for a model with three variables etc.

factors such as fatigue. Posterior probabilities for assignment to group α can be calculated using the following formula:

$$p = \frac{e^{-(z_\alpha - z)^2}}{e^{-(z_\alpha - z)^2} + e^{-(z_\beta - z)^2}} \tag{2}$$

where p is the probability of a case being assigned to group α

Formula 2 can be transformed into Formula 3 in order to obtain the z value for a given probability p.

$$z = \frac{\ln(\frac{1}{p} - 1) - z_\alpha^2 + z_\beta^2}{2(z_\beta - z_\alpha)} \tag{3}$$

Formula 1 can be transformed into Formula 4 so that for given values of y, ..., n it is possible to calculate an x value:

$$x = (z - C_{constant} - C_y y - \ldots - C_n n) / C_x \tag{4}$$

An example of an application of the above process would be the derivation of a categorical boundary between English /ɪ/ and /i/ based primarily on spectral properties. The discriminant analysis is applied to the predictor variables *spec* (vowel spectral properties), *durv* (vowel duration), *durc* (consonant closure duration), and *spd* (speaking rate). *durv*, *durc*, and *spd* are systematically varied from 1-7, 1-5, and 1-2 respectively and substituted into Formula 5 to produce corresponding values for *spec*.

$$spec = (z - C_{constant} - C_{spd} \cdot spd - C_{durc} \cdot durc - C_{durv} \cdot durv) / C_{spec} \tag{5}$$

C_{spec}, C_{durv}, C_{durc}, C_{spd}, and $C_{constant}$ are coefficients generated by the discriminant analysis. z is calculated from Formula 3 in which z_α and z_β, the discriminant function values at the /ɪ/ and /i/ group centroids respectively, are generated by the discriminant analysis. The values 0.1, 0.5, and 0.9 are substituted for *p* (the probability of a case being assigned to the /ɪ/ group). The *spec*, *durv*, *durc*, and *spd* values are then graphed in the four dimensional space defined by the continuum dimensions (see Figure 3.4).

Other information supplied by the discriminant analysis includes Wilks's lambda, chi-square and classification results. Wilks's lambda (Λ) is the ratio of within-group variability to total variability in the discriminator variables. Values close to 1 indicate that almost all the variability is dues to within-group differences; values close to 0 indicate that almost all the variability is due to group differences. A chi-square (χ^2) test indicates whether the variability related to group differences is statistically significant. Classification results indicate the success of the model at assigning cases to the group to which they actually belong. High correct-classification scores would indicate that the boundary is sharp and linear. Low correct-classification scores may indicate that the boundary is fuzzy or that it is non-linear. Cross-validated correct-classification scores are derived by determining the classification of each case by a model including all cases except the case being tested (i.e. a jackknife, leave-one-out, or U-method procedure).

The discriminant analysis also supplies standardised canonical discriminant function coefficients which indicate the relative weight of the contribution of each variable in the model. Standardised coefficients will compensate for the differences in the range of values of the predictor variables in the present study, e.g., *durv* has values from 1 to 7 but *spd* only has values of 1 and 2.

In order to determine which variables were significant contributors to the listeners' vowel identification, variables were entered into the model in a stepwise manner. At each step the variable which maximised the smallest *F* ratio between the groups was entered or removed from the model. Variables were only entered if they had an *F* ratio significance of .05 or less and removed if they had an *F* ratio significance of .10 or more. Stepwise methods may include too many variables in the model (see Stevens, 2002, Section 7.7); therefore the last variables entered into the model (corresponding to the variables with the lowest weighting in the results below) may in fact not be relevant factors in the listeners' perception.

In order to focus on the categorical boundary between the two vowel categories in each continuum, and to maintain a simple two category model, "none" responses were not included in the calculation of the discriminant functions.

Discriminant analysis assumes multivariate normal distributions, although the procedure has been found to be robust to violations of this assumption. The data in the present study did not always conform to the normal-distribution assumption, e.g., the English listeners' *spec* data had an essentially bimodal distribution, and the distribution curve of the Spanish listeners' *durv* data rose towards the long and short extremes and was then truncated. The data, however, always had mirror image symmetry. This ensured that the calculation of the 0.5 probability boundary would be appropriate. 0.1 and 0.9 probability boundaries calculated on the assumption of normal distributions were closer to the 0.5 probability point than if the model had taken into account the skew in the distribution. Although for the Spanish *durv* data an arcsine transformation could reduce the degree of skew, no transformation was effective for the English *spec* data. It was therefore decided to use untransformed data throughout with the understanding that the model would consistently underestimate the distance between the 0.1 and 0.9 probability boundaries for the Spanish data.

4.2 Canadian English Participants

A discriminant analysis was conducted on the native English listeners' group results to determine their categorical boundary between English /i/ and /ɪ/. The variables entered into the model were spectral properties, vowel duration, and consonant closure duration, but not speaking rate. This suggests that speaking rate did not affect the English participants' vowel perception. The model was highly successful at categorising the English participants' responses, resulting in a cross-validated correct classification rate of 95.5%; see Table 4.1. $\Lambda = .234$ and $\chi^2(3, N = 2101) = 3045.365$, $p < .0005$.

Figure 4.1 shows the planes describing the categorical boundary derived from the discriminant analysis of the English participants' /i/ and /ɪ/ perception. The spectral values of the planes were calculated by substituting the coefficients and function values in Table 4.2 into Formulae 3 and 4. The centre plane represent 0.5 probability of /ɪ/ identification, indicating the position of the categorical boundary. The outer planes represent 0.9 and 0.1 probability of /ɪ/ identification. The distance between these planes provides a measure of the sharpness of the categorical boundary. The planes are almost perpendicular to the *spec* axis indicating that the native English participants based their vowel identification overwhelmingly on the spectral properties of the vowel. The spectral value at the centre of the planes, i.e. the mean spectral value of the boundary between /i/ and /ɪ/, was 3.07 (F1, F2, F3 = 415, 1568, 1875 mel; 333, 1964, 2668 Hz). The spectral distance between 0.1 and 0.9 probability of /ɪ/ identification was 0.44 (ΔF1, ΔF2, ΔF3 = 18, -30,

-19 mel; 17, -62, -47 Hz [Hertz values were calculated at the centre of the planes]) indicating a sharp boundary.

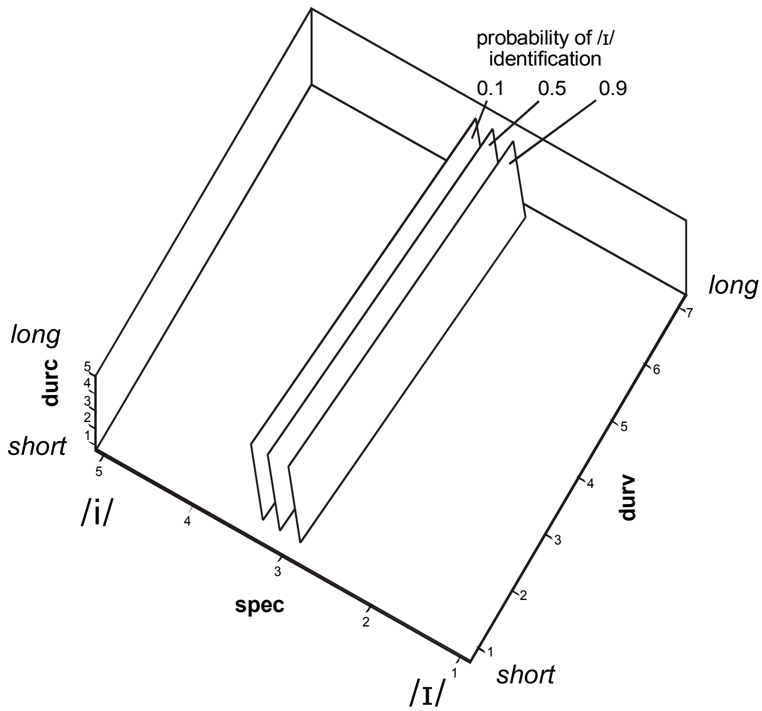


Figure 4.1 Planes describing the categorical boundary between /i/ and /ɪ/ derived from the linear discriminant function carried out on the identification responses of native English participants. Axes: *spec* = spectral properties, *durv* = vowel duration, *durc* = consonant closure duration; see section 3.2.2.

The standardised canonical discriminant function coefficients, see Table 4.2, indicate that the relative weighting given by the English participants to the different variables were 84% for vowel spectral properties, 11% for vowel duration, and 5% for consonant closure duration. The use of secondary cues is also apparent from the slight slopes visible in the planes in Figure 4.1. The small effect for vowel duration is predictable since tense /i/ is longer than lax /ɪ/, all else being equal (see section 2.3). The slight effect for consonant closure duration is also predictable since a long consonant closure will create the perceptual contrast effect of a shorter vowel (Kluender, Diehl, & Wright, 1988).

Further patterns of perception were observed by graphing the raw results of individual speakers. Figure 4.2 shows a typical vowel identification pattern for an individual native English speaker. *None* responses, i.e. none of bit, beat, bid, bead, were often given for stimuli with spectral value 3. These stimuli would have been ambiguous since they were on the categorical boundary between /i/ and /ɪ/. Five of the seven native English listeners also gave some *none* responses to stimuli with long vowel durations and spectral values of 1 and 2. These had /ɪ/-like spectral properties but vowel durations that were extremely long compared to natural /ɪ/ vowels. That native English speakers were systematic in their use of the *none* response is also reflected in the classification results from the discriminant analysis: The majority of *none* responses were assigned to the /ɪ/ group; see Table 4.1.

In summary, native English listeners clearly identified stimuli as containing /i/ or /ɪ/ according to the spectral properties of the vowel with a sharp categorical boundary between the two vowels. The results strongly suggested that they paid little attention to vowel and consonant closure duration differences, and no attention to speaking rate.

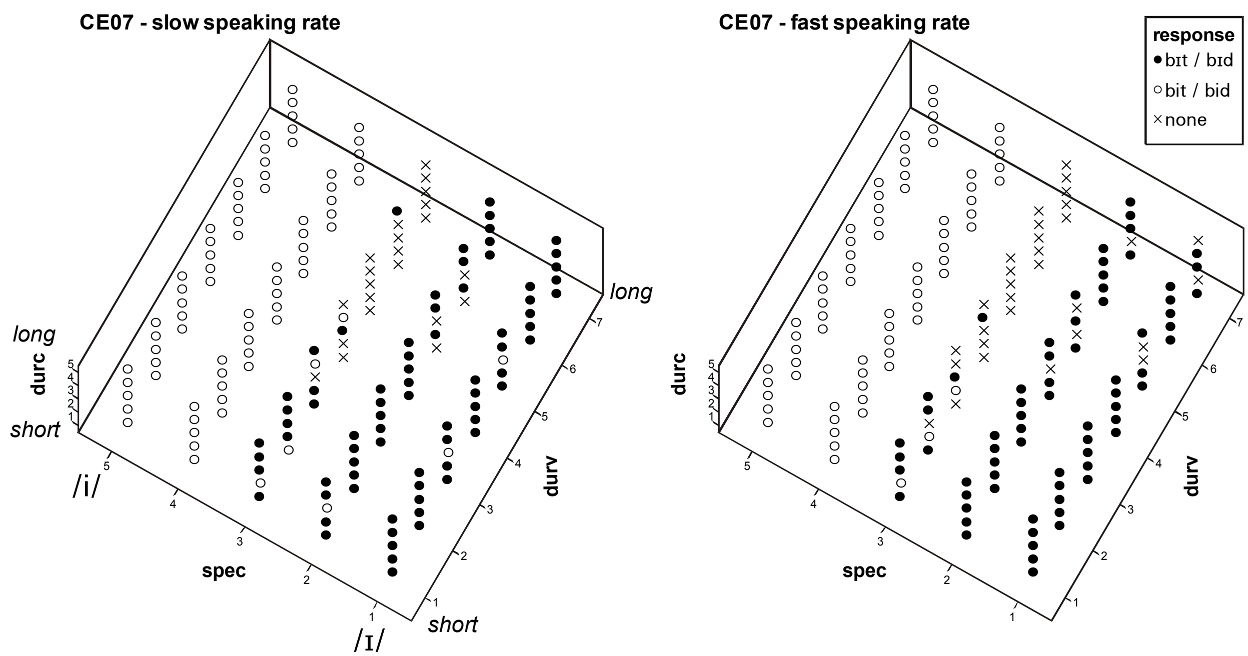


Figure 4.2 A typical vowel identification pattern for a native English participant in the initial test.

Table 4.1 Classification results for discriminant analysis carried out on the identification of /i/ and /ɪ/ by native English listeners. Model includes all variables.

Classification Results[b,c]

		Actual Group Membersip	Predicted Group Membership ɪ	i	Total
Original	Count	ɪ	1003	29	1032
		i	62	1008	1070
		none	265	83	348
	%	ɪ	97.2	2.8	100.0
		i	5.8	94.2	100.0
		none	76.1	23.9	100.0
Cross-validated[a]	Count	ɪ	1000	32	1032
		i	62	1008	1070
	%	ɪ	96.9	3.1	100.0
		i	5.8	94.2	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 95.7% of original grouped cases correctly classified.

c. 95.5% of cross-validated grouped cases correctly classified.

Table 4.2 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of /i/ and /ɪ/ by native English listeners.

	Coefficients	
	Unstandardised	Standardised
spec	1.395	1.002
durv	0.064	0.128
durc	-0.046	-0.065
constant	-4.431	

Centroid Values	
ɪ	i
-1.840	1.775

4.3 Japanese Participants - Initial English Test

A discriminant analysis was conducted on the Japanese listeners' group results to determine their categorical boundary between English /i/ and /ɪ/. The results of one participant, JP09, were radically different to the results of the remainder of the Japanese participants (details will be given in section 4.7.2.1 below) and were therefore not included in the general analysis of the Japanese participants' responses. The variables entered into the model were spectral properties, vowel duration, and consonant closure duration, but not speaking rate. This suggests that speaking rate did not affect the Japanese listeners' vowel perception, and is in accordance with the results of other studies (Strange et al., 1998; Guion et al., 2002, Ingram & Park, 1997) that found little effect due to speaking rate on Japanese listeners' identification of English lax versus tense vowels. The discriminant-analysis model was moderately successful at categorising the responses, resulting in a cross-validated correct classification rate of 85.9% (see Table 4.3). $\Lambda = .468$ and $\chi^2(3, N = 1907) = 1446.863$, $p < .0005$.

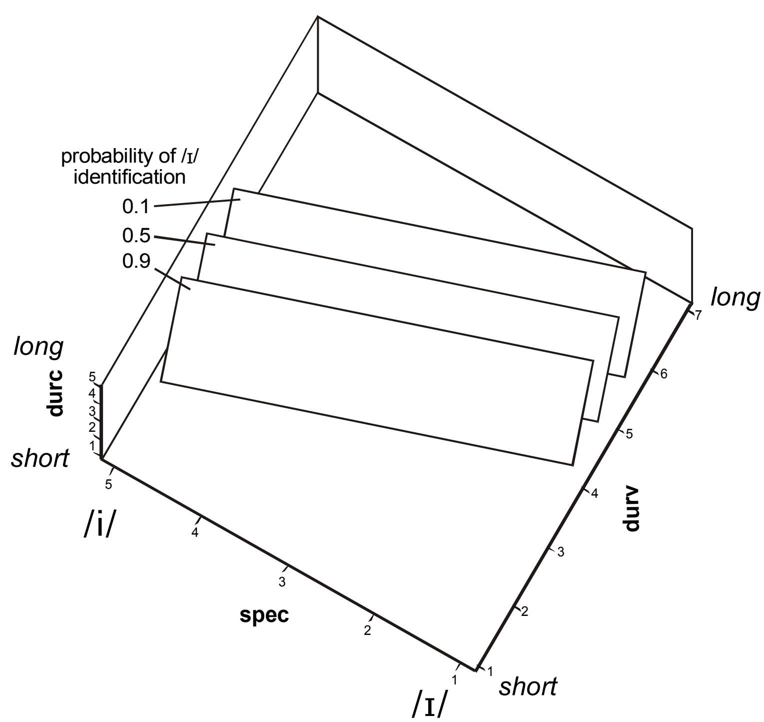


Figure 4.3 Planes describing the categorical boundary between English /i/ and /ɪ/ derived from the linear discriminant function carried out on the identification responses of Japanese participants in the initial test.

Figure 4.3 shows the planes describing the categorical boundary derived from the discriminant analysis of the Japanese participants' /i/ and /ɪ/ perception. The duration values of the planes were calculated from the coefficients and function values in Table 4.4. The orientation of the planes indicates that the Japanese participants based their vowel identification primarily on the duration properties of the vowel. The standardised canonical discriminant function coefficients, see Table 4.4, indicate that the relative weighting given by the Japanese participants to the different variables were 63% for vowel duration, 28% for vowel spectral properties, and 9% for consonant closure duration. The duration value at the centre of the planes, i.e. the mean duration of the boundary between /i/ and /ɪ/, was 4.22 (108 ms). The vowel duration distance between 0.1 and 0.9 probability of /ɪ/ identification was 1.49 (29 ms at the centre of the planes) indicating a relatively sharp boundary. The slopes due to the use of secondary cues resulted in a considerable range for the duration value of the boundary: The duration range for the 0.5 planes was 2.98 to 5.46 (86 to 135 ms), and for the 0.1 and 0.9 planes 2.24 to 6.20 (75 to 155 ms).

These results support the prediction in Hypothesis 1 that Japanese listeners would identify /i/ and /ɪ/ primarily according to vowel duration. The Japanese participants also made some use of vowel spectral properties. Inspection of the graphs of individual participants' data indicates that the linear solution derived from the discriminant analysis may not be an ideal description of the categorical boundary. The individuals' graphs, see example in Figure 4.4, showed an incursion of /ɪ/ responses into the long duration - /ɪ/-like spectral value quadrant but not into the long duration values - /i/-like spectral value quadrant. One explanation for this pattern may be that, like native English listeners, Japanese listeners consider long vowels with /i/-like spectral values to be the best examples of English /i/, and recognised long vowels with /ɪ/-like spectral values as having conflicting spectral and duration cues. However, relative to the native English speakers who gave priority to the spectral properties of these conflicting stimuli (resulting in either /ɪ/ or *none* responses), the Japanese listeners gave greater weight to the duration properties of the conflicting stimuli (resulting in /i/ as well as /ɪ/ and *none* responses). For shorter vowels the Japanese listeners gave priority to duration cues resulting in /ɪ/ responses irrespective of spectral properties. An alternative explanation may be that the Japanese listeners were sensitive to the spectral properties of the vowels, but only if the spectral information was available for a sufficient length of time. Long vowels provide a larger quantity of spectral information and also allow more time for the listener to process the information. Thus, they would only distinguish spectral properties and give both /ɪ/

and /i/responses if the vowels were sufficiently long. The latter possibility was suggested in a cross-language study by Weiss (1976). It would also be consistent with studies reviewed by Repp (1984) that indicated that monolinguals' discrimination performance was higher for vowel stimuli with longer durations than for otherwise identical vowel stimuli with shorter durations.

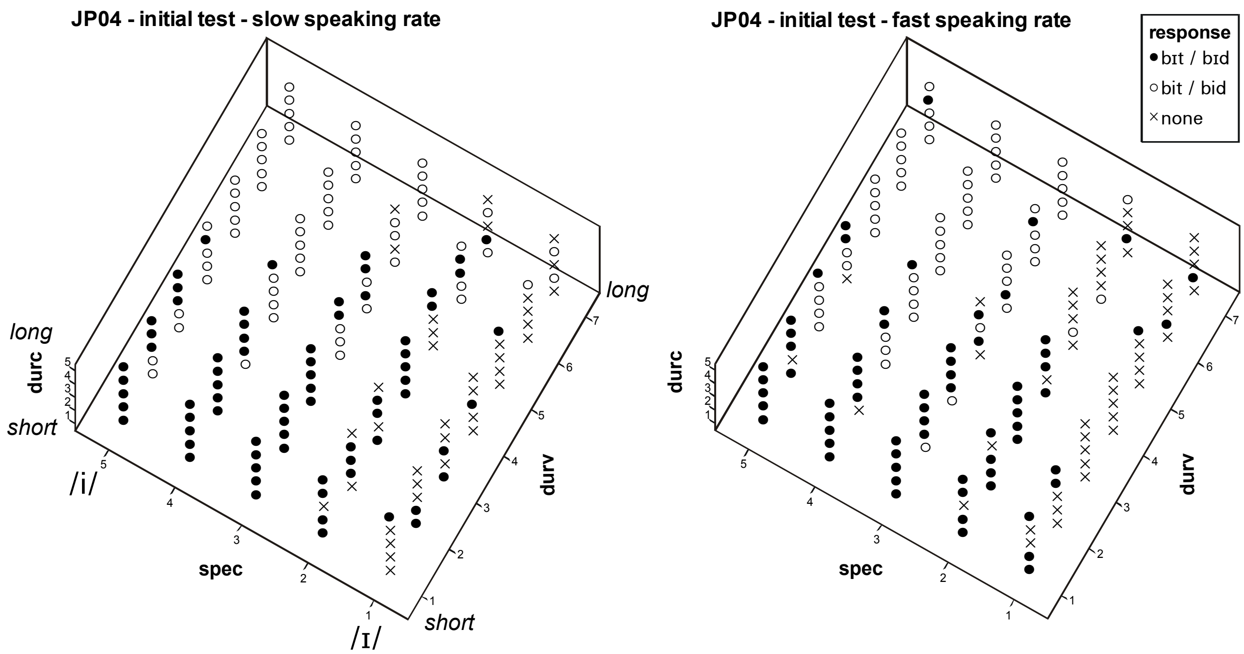


Figure 4.4 A typical English vowel identification pattern for a Japanese participant in the initial test.

The Japanese participants' use of spectral properties may be a result of learning English. Although still relying primarily on transfer of L1 duration criteria they may have realised that English /i/ and /ɪ/ differ spectrally and may have begun to shift their cue weighting towards greater use of spectral properties. L1 experience, however, may also have caused sensitivity to spectral properties: Fitzgerald (1996) found that Japanese speakers' long vowels were more peripheral in the vowel space than their short vowels. He found that, with only F1 and F2 as independent variables, a discriminant analysis on all vowels correctly classified 77% of /iː/ and 69% of /i/ vowels. The spectral differences may exist due to articulatory constraints (i.e. undershoot in short vowels). The existence of spectral differences leads to the possibility that spectral properties may in fact be a secondary cue differentiating long-short vowel pairs in Japanese. Further research is planned to determine whether this is the case.

A further pattern of perception was observed in the graphs of individual speakers' raw results. Four of the seven participants showed a tendency to give *none* responses to stimuli with spectral value 1, and to a lesser extent to stimuli with spectral value 2 (see Figure 4.4). When asked what they had heard when they gave *none* responses, these participants indicated that they had heard "bed." Hence, it appears that Canadian English /ɪ/ was sufficiently low to be perceived as a version of Japanese /e/, or of English /ɛ/ which may have formed a diaphone with Japanese /e/. Such a response is not unexpected given that Yamada, Akahane-Yamada, & Strange (1995) found that natural US English /ɪ/ tokens were transcribed as the phonetic symbol /e/ at a rate of 48% by monolingual Japanese listeners. Strange et al. (1998) also found that natural US English /ɪ/ tokens were assimilated to Japanese /e eː ei/ at a rate of 18% in a sentence condition and at 41% in an isolated word condition. Guion et al. (2002) found that US English /ɪ/ was assimilated as Japanese /e/ at a rate of 87% (and 7% as /eː/). Tsukada (1999) also found considerable spectral overlap of Japanese /e/ with Australian English /i/ and /ɪ/. It should be noted, however, that pictures of a bed, a bird, and a bat were provided in the perception-test instructions as examples of other words that might be heard, this may have biassed the participants towards reporting that they heard "bed" rather than reporting that they were merely unable to identify the vowel.

In summary, the Japanese listeners clearly identified stimuli as containing English /i/ or /ɪ/ according to the duration properties of the vowel but paid some attention to spectral properties. The predominantly duration-based identification pattern was radically different to the almost exclusively spectral identification pattern used by the native English speakers.

Table 4.3 Classification results for discriminant analysis carried out on the identification of /i/ and /ɪ/ by Japanese listeners.

Classification Results[b,c]

		Actual Group Membership	Predicted Group Membership ɪ	i	Total
Original / Cross-validated	Count	ɪ	900	164	1064
		i	105	738	843
		none	111	82	193
	%	ɪ	84.6	15.4	100.0
		i	12.5	87.5	100.0
		none	57.5	42.5	100.0

b. 85.9% of original grouped cases correctly classified.

c. 85.9% of cross-validated grouped cases correctly classified.

Table 4.4 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of /i/and /ɪ/ by Japanese listeners.

	Coefficients	
	Unstandardised	Standardised
spec	0.324	0.436
durv	0.688	0.996
durc	-0.102	-0.143
constant	-3.444	

Centroid Values	
ɪ	i
-0.949	1.198

4.4 Japanese Test

A discriminant analysis was conducted on the Japanese listeners' group results to determine their categorical boundary between Japanese /iː/ and /i/. The variables entered into the model were vowel duration, consonant voicing, and speaking rate, and the model had a cross-validated correct classification rate of 88.6% (see Table 4.5). $\Lambda = .409$ and $\chi^2(3, N = 1960) = 1749.070, p < .0005$. The mean intra-listener consistency rate, i.e. the proportion of Japanese stimuli that each Japanese listener identified as being the same vowel in both randomisations, was 86.6%. Hence the discriminant analysis was highly successful at modelling the Japanese listeners' perception. The standardised canonical discriminant function coefficients from the discriminant-analysis model, see Table 4.6, indicate that the relative weighting given by the Japanese participants to the different variables were 81% for vowel duration, 13% for consonant voicing, and 5% for speaking rate. This suggests that the Japanese listeners based their identification of /i/ and /iː/ primarily on vowel duration. The small effect due to consonant voicing is consistent with the findings of Homma (1973, 1981), Maeda (1979, cited in Takahashi, 1987), Port et al. (1980), Takahashi (1987), and Tsukada (1996, 1999) (see section 2.3.3.1.2). The small effect for speaking rate is consistent with the results of Toda (1999) who found only a small effect for speaking rate on Japanese listeners' identification of Japanese long versus short vowels.

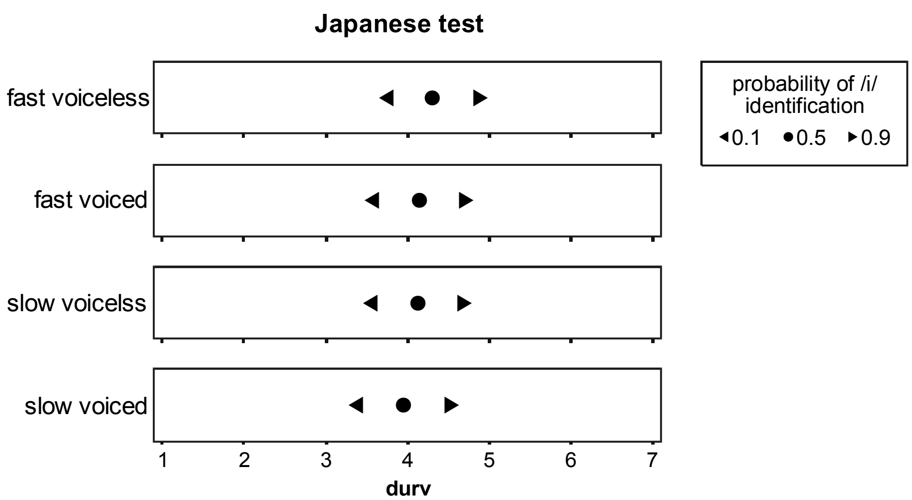


Figure 4.5 Points describing the categorical boundary between Japanese /i/ and /iː/ derived from the discriminant analysis carried out on the identification responses of Japanese participants.

The categorical boundary points between the Japanese participants perception of /i/ and /iː/ shown in Figure 4.5 were calculated from the coefficients and function values in Table 4.6. The mean duration of the boundary between /i/ and /iː/ was 4.07 (105 ms). The vowel duration distance between 0.1 and 0.9 probability of /i/ identification was 1.16 (22 ms when centred on the mean duration for the 0.5 probability boundary) indicating a relatively sharp boundary. Due to the effects of consonant voicing and speaking rate, the duration range for the 0.5 planes was 3.89 to 4.25 (102 to 108 ms), and for the 0.1 and 0.9 planes 3.31 to 4.83 (91 to 121 ms).

In summary, as might be expected, the Japanese listeners identified Japanese short /i/ and long /iː/ primarily according to vowel duration and had a sharp categorical boundary between the two vowels.

Table 4.5 Classification results for discriminant analyses carried out on the identification of /i/ and /iː/ by Japanese listeners.

Classification Results[b,c]

			Predicted Group Membership		
		response	i	iː	Total
Original / Cross-validated	Count	i	1002	141	1143
		iː	83	734	817
	%	i	87.7	12.3	100.0
		iː	10.2	89.8	100.0

b. 88.6% of original grouped cases correctly classified.

c. 88.6% of cross-validated grouped cases correctly classified.

Table 4.6 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of /i/ and /iː/ by Japanese listeners.

	Coefficients	
	Unstandardised	Standardised
durv	0.776	1.003
voicing	-0.149	-0.166
spd	0.130	0.065
constant	-2.928	

Centroid Values	
i	iː
-1.016	1.421

4.5 Comparison of English and Japanese Test Results

Research Question 1 asks whether Japanese listeners distinguish English /i/ and /ɪ/ using a duration-based categorical boundary in the same location as their categorical boundary between Japanese /iː/ and /i/. Whilst the Japanese participants did make some use of spectral cues in their identification of the English vowels, their primary cue was vowel duration.

Since the responses of JP09 were not included in the derivation of the Japanese group's categorical boundary between English /i/ and /ɪ/, their English boundary should be compared with the Japanese /iː/-/i/ boundary calculated excluding JP09's responses. Likewise, since the English stimuli all had silent consonant closures, comparison should be made between the English boundary and the Japanese /iː/-/i/ boundary based only on the responses to silent-consonant-closure stimuli. In addition, the Japanese boundary should be compared with the English /i/-/ɪ/ boundary based on stimuli with the same spectral properties as the Japanese stimuli. The formant frequencies of the Japanese stimuli (F1, F2, F3 = 318, 2126, 3115 Hz) were closest in terms of F1 and F2 to the English stimuli with spectral value 4 (F1, F2, F3 = 292, 2121, 2786 Hz).

A discriminant-analysis model based on responses to Japanese silent-consonant-closure stimuli and excluding the responses of JP09 resulted in a cue weighting of 92% for vowel duration and 8% for consonant closure duration. Discriminant function coefficients and centroid values for this model are given in Table 4.7, and classification results are given in Table 4.8. $\Lambda = .415$ and $\chi^2(2, N = 412) = 359.579, p < .0005$. The mean Japanese /i/-/iː/ categorical boundary (0.5 probability of /i/ identification) was 4.13 (106 ms). The vowel duration distance between 0.1 and 0.9 probability of /i/ identification was 0.99 (19 ms when centred on the mean duration for the 0.5 probability boundary) indicating a relatively sharp boundary. Due to the effects of consonant closure, the duration range for the 0.5 planes was 3.98 to 4.29 (103 to 109 ms), and for the 0.1 and 0.9 planes 3.31 to 4.83 (94 to 120 ms).

A discriminant-analysis model based on the Japanese listeners' responses to English /ɪ/ and /i/ stimuli with spectral value 4, excluding the responses of JP09, resulted in a cue weighting of 84% for vowel duration and 16% for consonant closure duration. Discriminant function coefficients and centroid values for this model are given in Table 4.9, and classification results are given in Table 4.10. $\Lambda = .359$ and $\chi^2(2, N = 840) = 856.666, p < .0005$. The mean English /ɪ/-/i/ categorical boundary (0.5 probability of /ɪ/ identification) was 3.82 (100 ms). The vowel duration distance between 0.1 and 0.9 probability of /ɪ/ identification was 1.20 (22 ms when centred on the mean duration for the 0.5 probability boundary) indicating a relatively sharp boundary. Due to the effects of consonant closure, the duration range for the 0.5 planes was 3.45 to 4.18 (94 to 107 ms), and for the 0.1 and 0.9 planes 2.85 to 4.78 (84 to 119 ms).

A graphical representation of the mean categorical boundary values for the Japanese listeners' Japanese /iː/-/i/ and English /ɪ/-/i/ perception is given in Figure 4.5. The 6 ms difference between the mean English and Japanese 0.5 probability points falls within both the English and Japanese 0.1 to 0.9 probability ranges. From these results it can be concluded that the Japanese participants' use of vowel duration in identifying English /ɪ/ and /i/ was similar to their use of vowel duration in identifying Japanese /i/ and /iː/. Their categorical duration boundary between English /ɪ/ and /i/ was very close to their categorical duration boundary between Japanese /i/ and /iː/.[2]

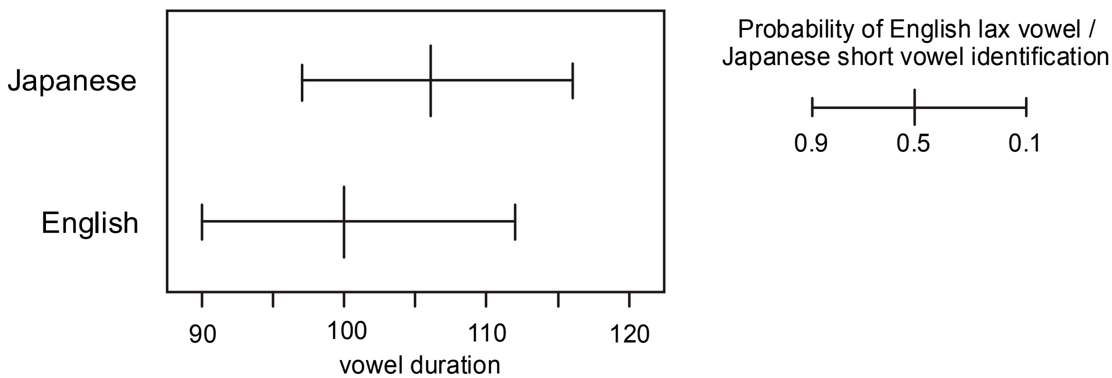


Figure 4.6 Comparison of the Japanese listeners' mean categorical duration boundary (in milliseconds) for Japanese /iː/ - /i/ (silent-consonant-closure stimuli only) with their mean duration boundary for English /i/ -/ɪ/ (spectral-value-4 stimuli only).

Table 4.7 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of Japanese /i/ and /iː/ silent-consonant-closure stimuli by Japanese listeners (excluding JP09).

	Coefficients	
	Unstandardised	Standardised
durv	0.832	1.001
durc	-0.065	-0.091
constant	-3.133	

Centroid Values	
i	iː
-1.227	1.450

[2] No economical method is available to determine the statistical significance of the difference.

Table 4.8 Classification results for discriminant analyses carried out on the identification of Japanese /i/ and /iː/ silent-consonant-closure stimuli by Japanese listeners (excluding JP09).

Classification Results[b,c]

		Actual Group Membership	Predicted Group Membership: i	Predicted Group Membership: iː	Total
Original / Cross-validated	Count	i	413	42	455
		iː	43	342	385
	%	i	90.8	9.2	100.0
		iː	11.2	88.8	100.0

b. 89.9% of original grouped cases correctly classified.

c. 89.9% of cross-validated grouped cases correctly classified.

Table 4.9 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of English /ɪ/ and /i/ spectral-value-4 stimuli by Japanese listeners (excluding JP09).

	Coefficients	
	Unstandardised	Standardised
durv	0.770	1.003
durc	-0.139	-0.196
constant	-2.657	

Centroid Values	
ɪ	i
-1.331	1.053

Table 4.10 Classification results for discriminant analyses carried out on the identification of English /ɪ/ and /i/ spectral-value-4 stimuli by Japanese listeners (excluding JP09).

Classification Results[b,c]

		Actual Group Membership	Predicted Group Membership: ɪ	Predicted Group Membership: i	Total
Original	Count	ɪ	164	18	182
		i	24	206	230
		none	4	4	8
	%	ɪ	90.1	9.9	100.0
		i	10.4	89.6	100.0
		none	50.0	50.0	100.0
Cross-validated[a]	Count	ɪ	164	18	182
		i	32	198	230
	%	ɪ	90.1	9.9	100.0
		i	13.9	86.1	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 89.8% of original grouped cases correctly classified.

c. 87.9% of cross-validated grouped cases correctly classified.

4.6 Mexican Spanish Participants - Initial English Test

A discriminant analysis was conducted on the Spanish listeners' group results to determine their categorical boundary between English /i/ and /ɪ/. The only variable entered into the model was vowel duration resulting in a cross-validated correct classification rate of 63.7% (see Table 4.11). This suggests that the Spanish participants based their identification of /i/ and /ɪ/ on vowel duration alone irrespective of other potential cues. The high Wilks's lambda indicates a high within-group variability: $\Lambda = .911$ and $\chi^2(1, N = 1577) = 147.407$, $p < .0005$. Visual inspection of graphs of individual participants' raw data (see example in Figure 4.7) did not reveal large contiguous subsets of stimuli that were clearly identified as one or other of the vowels. This suggests that the low correct-classification rate may in part be accounted for by a lack of consistency in intra-listener responses.[3]

[3]Since each stimulus was identified only once, it was not possible to calculate an intra-speaker consistency rate. Given the large number of stimuli, it was not possible to obtain multiple responses without fatiguing the listeners.

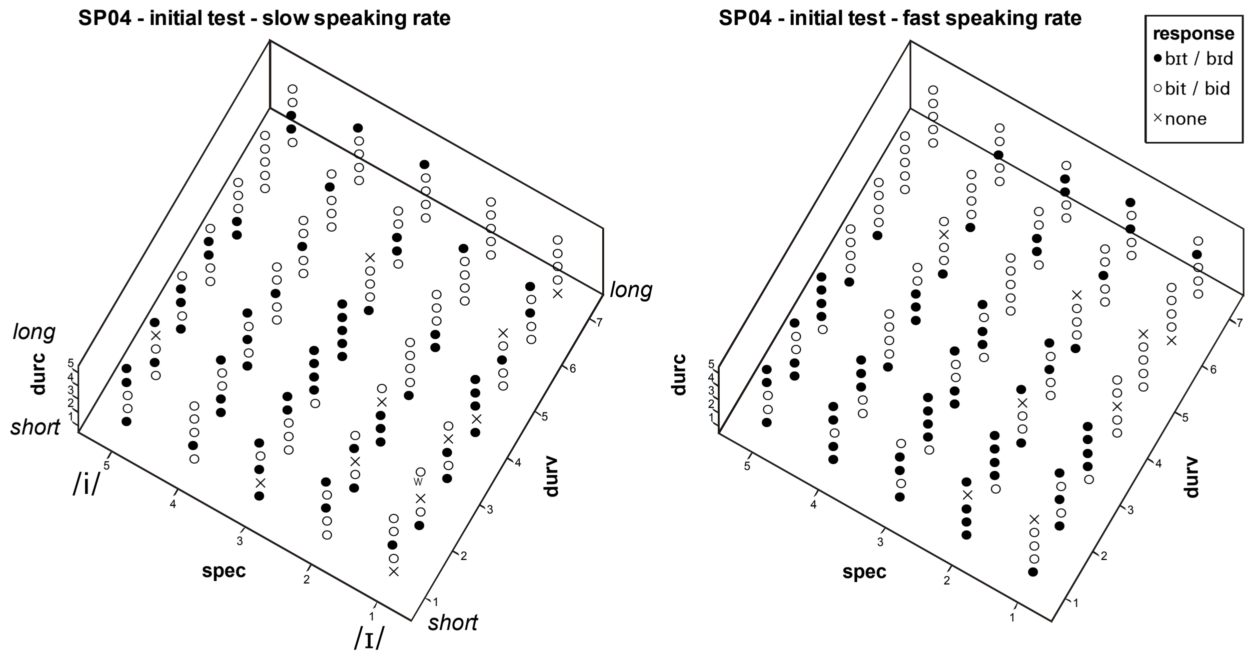


Figure 4.7 A typical English vowel identification pattern for a Spanish participant in the initial test.

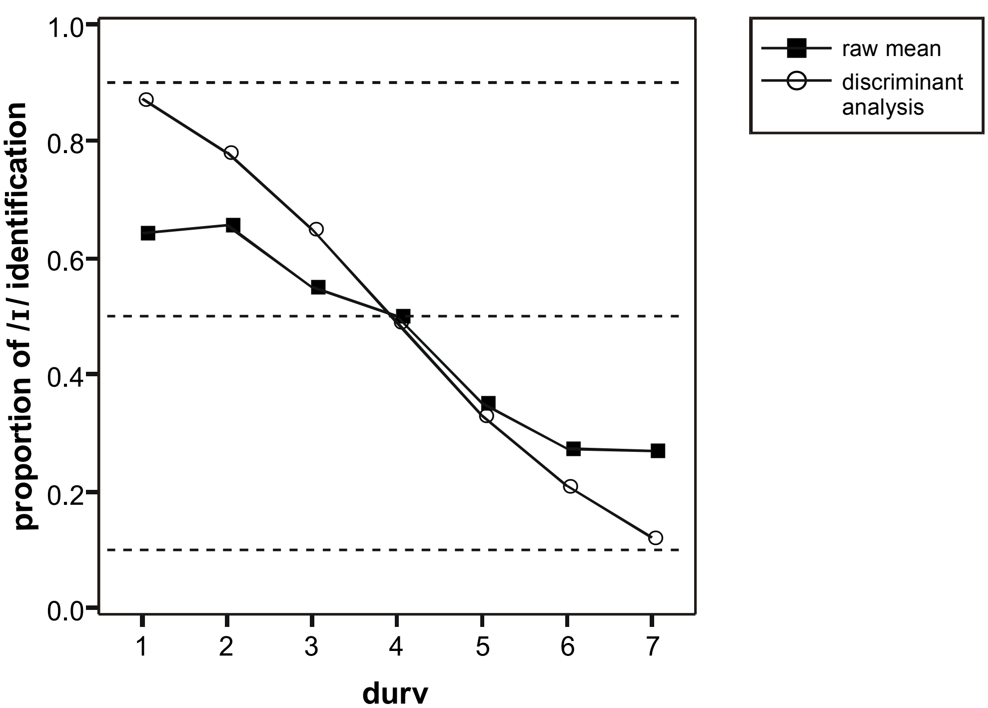


Figure 4.8 Proportion of stimuli identified as /ɪ/ by Spanish participants. Squares represent mean values for given durations. Circles represent probabilities derived from the discriminant analysis. Dashed lines represent 0.1, 0.5 and 0.9 probability thresholds.

The 0.1, 0.5, and 0.9 probability of /ɪ/ identification points were calculated from the coefficients and function values in Table 4.12. The Spanish participants' 0.5 probability of /ɪ/ identification was 3.95 (103 ms). The vowel durations for 0.1 and 0.9 probability of /ɪ/ identification were 0.59 (56 ms) and 7.30 (189 ms); these were outwith the range of vowel durations tested, indicating that there was no categorical boundary.[4] The Spanish participants' use of vowel duration in the perception of /i/ and /ɪ/ is shown graphically in Figure 4.8. The line representing the mean proportion of stimuli identified as /ɪ/ has a shallow slope, again indicating that the Spanish group's perception was not categorical. Separate discriminant analyses were also conducted on individual Spanish listeners' responses. No individual Spanish listener was found to have categorical perception of English /ɪ/ and /i/.

One Spanish participant (SP01) had a different pattern to the others, a more extreme version of the *none*-response pattern found above for Japanese participants. He gave *none* responses to the majority of stimuli with spectral values 1, 2, and 3 (see Figure 4.9), and reported hearing these stimuli as "bed." These spectral values may have fallen within this participant's Spanish /e/ category, or perhaps a diaphone category that included Spanish /e/ and English /ɪ/. Previous studies have found that Spanish listeners may identify English /ɪ/ as English /ɛ/, and assimilate both to Spanish /e/. Flege (1991b) found that 19% of US English /ɪ/ were identified as /ɛ/ by monolingual American Spanish listeners. The rate was higher for American Spanish learners of English who identified 39% of /ɪ/ as /ɛ/. Álvarez González (1980) found that 22% of English /ɪ/[5] were assimilated to Spanish /e/ by monolingual Peninsular Spanish listeners.

[4]The criterion for categorical perception established in section 4.1 was that the distance from the 0.1 to 0.9 probability boundaries was less than half the range of stimulus values tested, in this case less than 3, half the vowel duration range (1 to 7).

[5]Presumably produced in RP – the speaker was J. C. Wells.

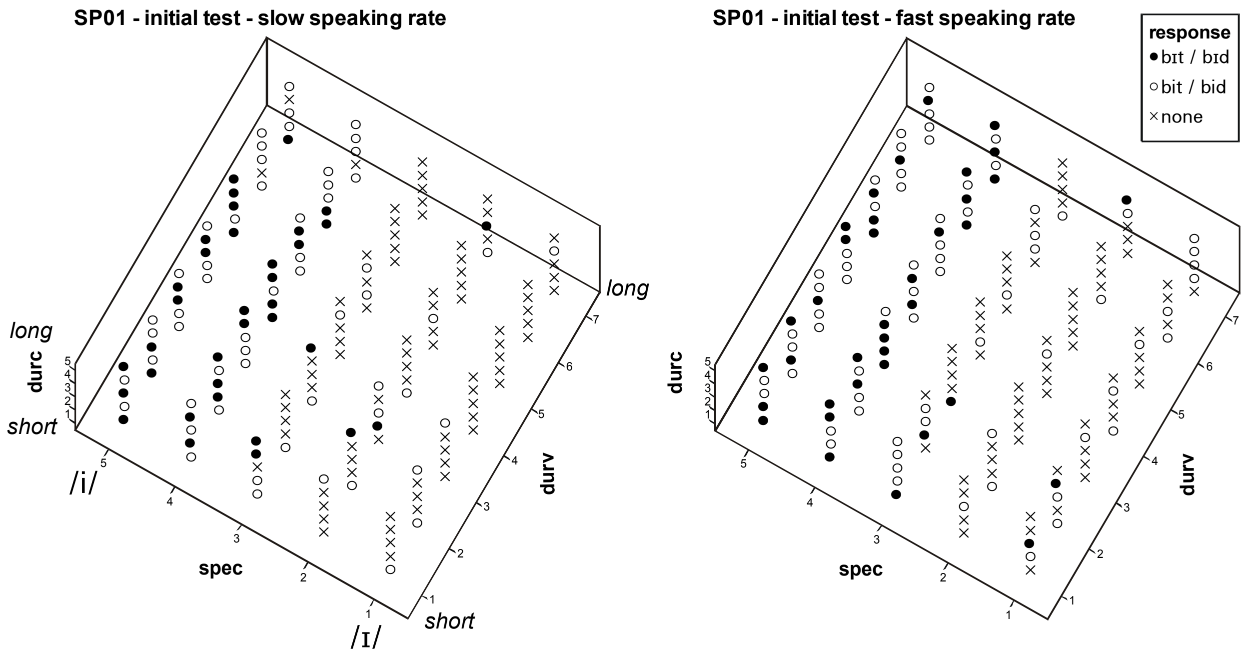


Figure 4.9 English vowel identification pattern for participant SP01 in the initial test.

Table 4.11 Classification results for discriminant analysis carried out on the identification of /i/ and /ɪ/ by Spanish listeners.

Classification Results[b,c]

		Actual Group Membership	Predicted Group Membership: ɪ	Predicted Group Membership: i	Total
Original / Cross-validated	Count	ɪ	421	311	732
		i	262	583	845
		none	67	106	173
	%	ɪ	57.5	42.5	100.0
		i	31.0	69.0	100.0
		none	38.7	61.3	100.0

b. 63.7% of original grouped cases correctly classified.

c. 63.7% of cross-validated grouped cases correctly classified.

Table 4.12 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of /i/ and /ɪ/ by Spanish listeners.

	Coefficients		Centroid Values	
	Unstandardised	Standardised	ɪ	i
durv	0.522	1	-0.336	0.291
constant	-2.079			

4.7 Final English Test

4.7.1 Predictions

The results of the Japanese test and the initial English test were consistent with the Japanese listeners assimilating English /i/ and /ɪ/ to Japanese /iː/ and /i/ in a form of two-category assimilation. Since each instance of an English vowel was perceived as similar to one of two Japanese vowels and subject to equivalence classification, the SLM predicts that two diaphone categories will be established. Hypothesis 2 of the present study (rephrased here as a prediction) predicts that one diaphone category will consist of English /i/ preceding a voiced consonant plus Japanese /iː/, and the other category will consist of English /i/ preceding a voiceless consonant, plus English /ɪ/, plus Japanese /i/. The eventual properties of each diaphone category would be expected to be intermediate between the properties of the L1 and L2 sounds. Since Japanese /iː/ and English /i/ are very similar spectrally, little spectral change is expected. Spectral change in the diaphone category of which Japanese /i/ is a member is expected to be small since, as well as English /ɪ/, the category contains instances of English /i/ which differs little from Japanese /i/. The Japanese participants assigned instances of English /i/ and /ɪ/ to Japanese categories according to Japanese duration criteria. Instances of English vowels assimilated to Japanese categories therefore have durations within the acceptable range for Japanese vowels in each vowel category. Therefore the duration properties of the English vowels will not influence the duration properties of the diaphone categories. Hence no change in Japanese listeners' use of duration cues is expected unless they first learn to correctly identify the English vowels using spectral cues. The Japanese participants are therefore predicted to show no or minimal changes in vowel identification between the initial and final tests.

The results of the initial English test were consistent with the Spanish listeners assimilating English /i/ and /ɪ/ to Spanish /i/. Some differentiation between /i/ and /ɪ/ was made, suggesting a category-goodness assimilation. This leaves open the possibility that the Spanish listeners will establish a new category for English /ɪ/ (the vowel which is least similar to the Spanish vowel [Flege, 1991b; Flege, Munro, & Fox, 1994]). If a new category is established, then the properties of that category are expected to eventually match the L2 properties for that sound (should there be no deflection of difference in cue weighting). In contrast with the Japanese participants, the Spanish participants are therefore predicted to show relatively large changes in vowel identification between the initial and final tests.

4.7.2 Japanese Participants

A discriminant analysis was conducted on the Japanese listeners' group results (excluding the results of JP09) to determine their categorical boundary between English /i/ and /ɪ/ in the final test. The variables entered into the model were vowel duration, spectral properties, consonant closure duration, and speaking rate). The discriminant-analysis model was moderately successful at categorising the responses resulting in a cross-validated correct classification rate of 87.6%; see Table 4.13. $\Lambda = .408$ and $\chi^2(4, N = 1983) = 1772.409$, $p < .0005$.

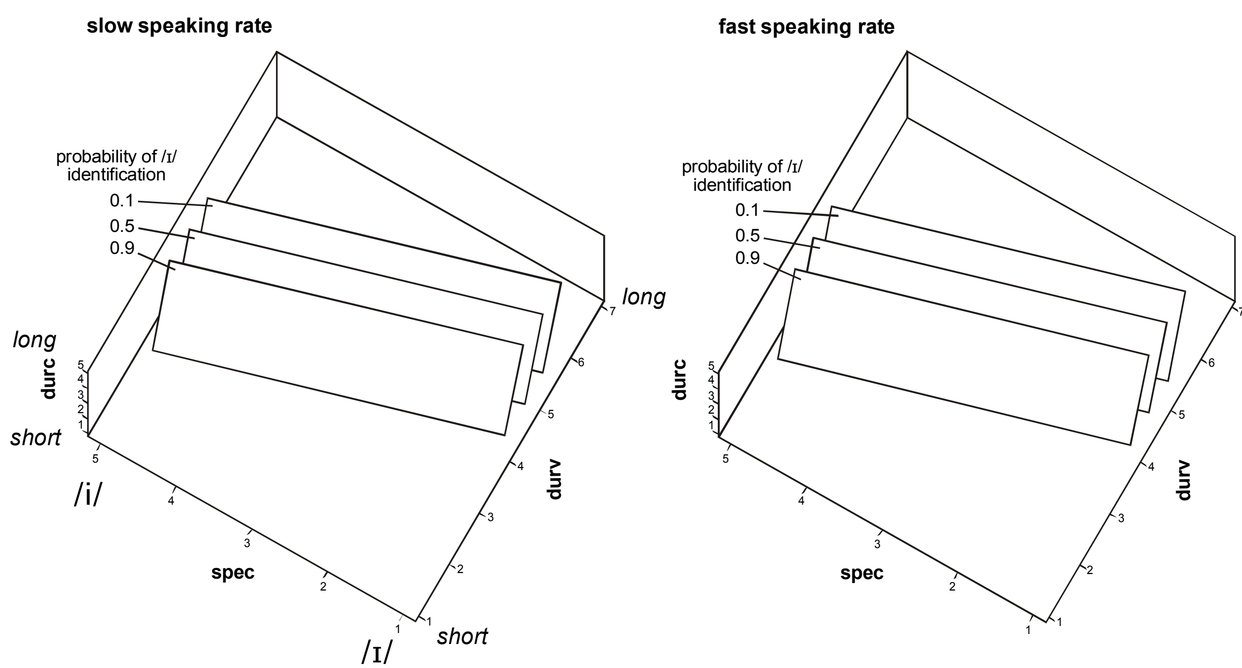


Figure 4.10 Lines describing the categorical boundary between /i/ and /ɪ/ derived from the linear discriminant function carried out on the identification responses of Japanese participants in the final test.

Figure 4.10 shows the planes describing the categorical boundary derived from the discriminant analysis of the Japanese participants' /i/ and /ɪ/ perception. The duration values of the planes were calculated from the coefficients and function values in Table 4.14. The slope of the planes indicates that the Japanese participants based their vowel identification primarily on the duration properties of the vowel. The standardised canonical discriminant function coefficients, see Table 4.14, indicate that the relative weighting given by the Japanese participants to the different variables were 62% for vowel duration, 26% for vowel spectral properties, 9% for consonant closure duration, and 4% for speaking rate. The duration value at the centre of the planes, i.e. the mean duration of the boundary between /i/ and /ɪ/, was 4.23 (108 ms). For a single speaking rate, the vowel duration distance between 0.1 and 0.9 probability of /ɪ/ identification was 1.61 (24 ms at the centre of the lines) indicating a relatively sharp boundary. Across speaking rates, the difference was 1.69 (27 ms). The slope due to the use of secondary cues resulted in a vowel duration range of 3.03 to 5.44 (87 to 135 ms) for the 0.5 plane, and of 2.42 to 6.05 (78 to 151 ms) for the 0.1 and 0.9 planes.

A comparison between the Japanese participants perception patterns for the initial and final English tests reveals little change other than the inclusion of speaking rate in the model for the final test but not the initial test. The contribution due to speaking rate, was, however, very small in the final test. The weighting of this cue was only 4%, and the weighting of other cues was very similar between tests (63, 28, and 9% for vowel duration, vowel spectral properties, and consonant closure duration respectively in the initial test, compared to 62, 26, and 9% in the final test). Measured to the nearest millisecond, the position of the mean value of the categorical boundary remained unchanged at 108 ms. The distance from the 0.1 to 0.9 probability boundaries was also similar, 29 ms in the initial test and 27 ms in the final test. For a visual comparison of the results of the initial and final tests, Figure 4.8 may be compared with Figure 4.3. As predicted in section 4.7.1, six months (compared to one month) living in an English speaking society had no effect on the Japanese listeners' perception of English /i/ and /ɪ/.

Table 4.13 Classification results for discriminant analysis carried out on the identification of /i/ and /ɪ/ by Japanese listeners in the final test.

Classification Results[b,c]

		Actual Group Membership	Predicted Group Membership ɪ	i	Total
Original	Count	ɪ	973	150	1123
		i	92	768	860
		none	69	48	117
	%	ɪ	86.6	13.4	100.0
		i	10.7	89.3	100.0
		none	59.0	41.0	100.0
Cross-validated[a]	Count	ɪ	970	153	1123
		i	92	768	860
	%	ɪ	86.4	13.6	100.0
		i	10.7	89.3	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 87.8% of original grouped cases correctly classified.

c. 87.6% of cross-validated grouped cases correctly classified.

Table 4.14 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of /i/and /ɪ/ by Japanese listeners in the final test.

	Coefficients: Unstandardised	Coefficients: Standardised
spec	0.309	0.418
durv	0.741	1.009
durc	-0.106	-0.150
spd	0.122	0.061
constant	-3.766	

Centroid Values: ɪ	Centroid Values: i
-1.053	1.375

4.7.2.1 Exceptional individual

In the initial English test the response pattern of participant JP09 differed radically from all other Japanese participants. He had a categorical boundary between English /i/ and /ɪ/ based on spectral properties. The only variable entered into the discriminant analysis on JP09's results was vowel spectral properties. The cross-validated correct classification rate for the model was 89.2% (see Table 4.15). $\Lambda = .432$ and $\chi^2(1, N = 296) = 246.201$, $p < .0005$. The spectral value of the boundary between /i/ and /ɪ/, was 2.72 (F1, F2, F3 = 400, 1592, 1890 mel; 319, 2014, 2707 Hz), and the spectral distance between 0.1 and 0.9 probability of /ɪ/ identification was 0.90 (ΔF1, ΔF2, ΔF3 = -38, 63, 39 mel; -35, 131, 99 Hz). The spectral values were calculated from the coefficients and function values in Table 4.16.

As can be seen in Figure 4.11, however, JP09 differed from the native English speakers in showing a reversal in labelling: Stimuli with spectral values near the /i/ end of the continuum were identified as /ɪ/, and stimuli with spectral values near the /ɪ/ end of the continuum were identified as /i/. Such reversals have been observed in previous studies, e.g., Crowther & Mann (1992); Flege (1993); Flege, Bohn, & Jang (1997); Wang (1997); and Escudero (2001a).

In the final test, participant JP09's perception pattern had changed radically so that it was the same as that of the other Japanese participants: primary use of the vowel duration cue with secondary use of spectral cues (see Figure 4.11).

There are several possible explanations for the reversal in the initial test. One possibility is that the participant perceived the spectral differences between the two vowel phonemes in an English-like manner, but was confused as to which spectral properties were associated with which phoneme, or as to which words contained which phoneme. Were this possibility true, then the participant might be able to easily correct the reversal and have English-like perception by the time of the final test. In the final test, however, the participant's perception pattern was similar to that of the other Japanese participants. A second possibility is that the participant noticed the spectral differences without associating them with phonemes and arbitrarily assigned one spectral extreme to one vowel response category and the other spectral extreme to the other vowel response category. This would be typical of a both-uncategorisable or non-assimilable pattern in the PAM (Best, 1995a). Given the acoustic similarities between English /i/ and /ɪ/, and Japanese /i/ and /iː/, however, a single-category or category-goodness assimilation would seem much more likely. A third possibility is that participant JP09 identified the English vowels according to a category-goodness assimilation, assimilating English /i/ to Japanese /iː/ and English /ɪ/ to Japanese /i/ according to spectral properties, but made an ‘error’ in associating the Japanese phonemes with the response categories (i.e. he associated the pictures of a “bit” and a “bid” with the long Japanese vowel and the pictures of a “beat” and a “bead” with the short Japanese vowel). A fourth possibility is that JP09 had a two-category assimilation in the initial test, assimilating English /i/ to Japanese /i iː/ and English /ɪ/ to Japanese /e eː/ according to spectral properties, and had an ‘error’ in associating the Japanese phonemes with response categories (i.e. he associated the pictures of a “bit” and a “bid” with Japanese /i iː/ and the pictures of a “beat” and a “bead” with Japanese /e eː/). The latter possibility would be most consistent with the sharp spectrally-based categorical boundary in the initial test. It would also be consistent with the change in response pattern from the initial to the final test: The participant may have realised that the spectral properties of English /ɪ/ were different to Japanese /e/ (or English /e/ or /ɛ/ which may have been assimilated to Japanese /e/) and so abandoned his initial perceptual strategy in favour of a strategy which assimilated English /i/ and /ɪ/ to Japanese /i/ and /iː/ primarily according to vowel duration, i.e. the same perceptual strategy as used by the other Japanese listeners.

Since the responses of JP09 were radically different to those of the remaining Japanese participants in the initial English test, they were not included in the discriminant analysis. Including the responses from JP09 would have resulted in a poorer model of the behaviour of the other participants and would not have added to the understanding of the behaviour of JP09 himself.

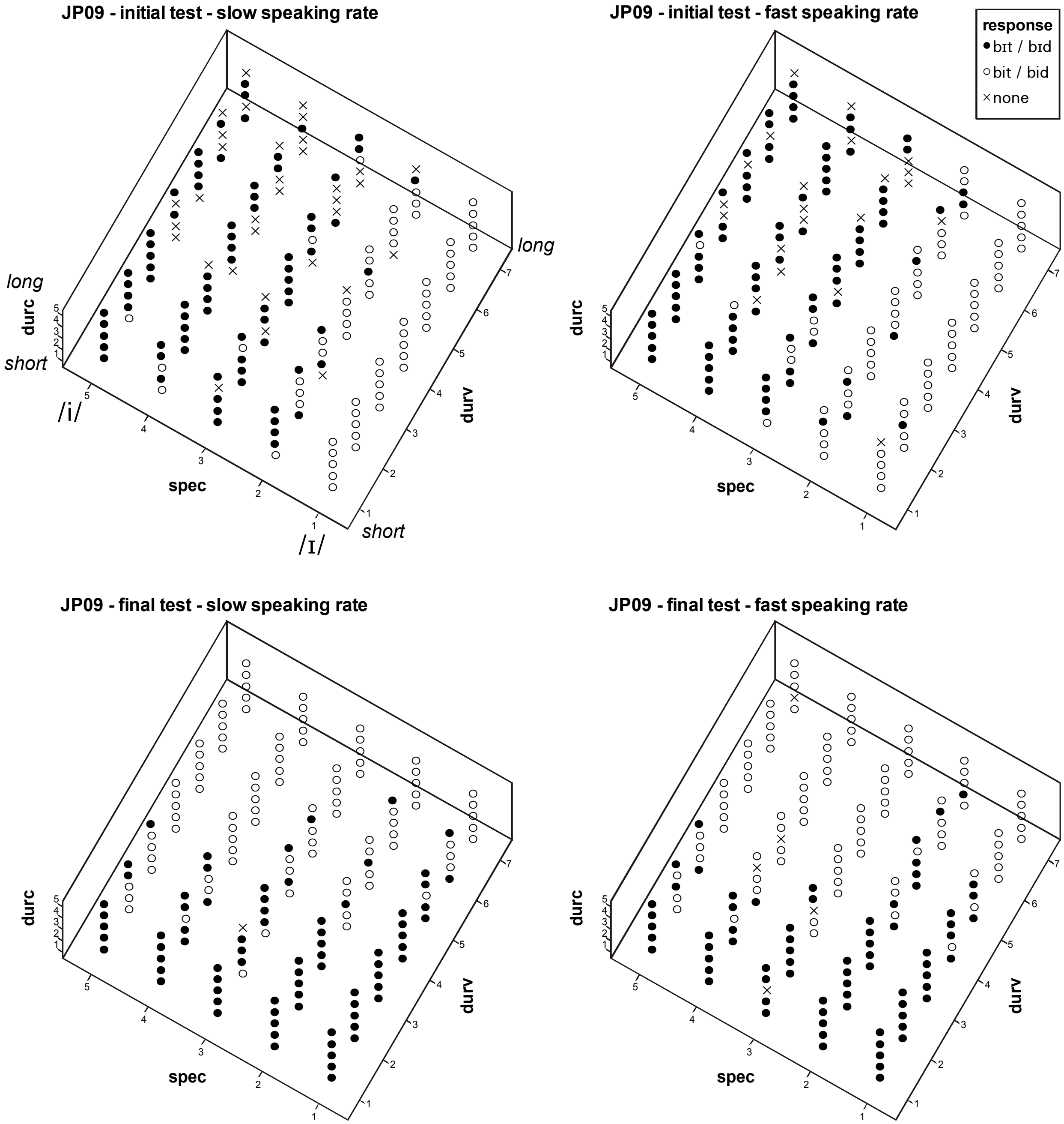


Figure 4.11 English /i/ - /ɪ/ identification by Japanese participant JP09 in the initial and final test.

Table 4.15 Classification results for discriminant analysis carried out on the identification of English /i/ and /ɪ/ by participant JP09 in the initial test.

Classification Results[b,c]

		Actual Group Membership	Predicted Group Membership ɪ	i	Total
Original / Cross-validated	Count	ɪ	149	19	168
		i	13	115	128
		none	48	6	54
	%	ɪ	88.7	11.3	100.0
		i	10.2	89.8	100.0
		none	88.9	11.1	100.0

b. 89.2% of original grouped cases correctly classified.

c. 89.2% of cross-validated grouped cases correctly classified.

Table 4.16 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of /i/ and /ɪ/ by participant JP03. Model includes only spectral properties as an independent variable.

	Coefficients	
	Unstandardised	Standardised
spec	1.055	1
constant	-3.024	

Centroid Values	
ɪ	i
0.997	-1.309

4.7.3 Mexican Spanish Participants

In the final test, four of the five Spanish participants showed categorical perception of English /i/ and /ɪ/. One had a categorical boundary based on vowel spectral properties and three had a categorical boundary based on vowel duration.

Participant SP02 had a perception pattern which was very similar to that of the native English participants. As can be seen in Figure 4.12, she had a categorical boundary between English /i/ and /ɪ/ based on the spectral properties of the stimuli. Variables entered into the discriminant analysis were vowel spectral properties and vowel duration. The standardised canonical discriminant function coefficients, see Table 4.17, indicate that the relative weighting given by this participant to the different variables were 83% for vowel spectral properties and 17% for vowel duration. The model had a moderate cross-validated correct classification rate of 78.3% (see Table 4.18). $\Lambda = .592$ and $\chi^2(2, N = 350) = 182.133$, $p < .0005$. Spectral values were calculated from the coefficients and function values in Table 4.14. The mean spectral value of the boundary between /i/ and /ɪ/, was 3.05 (F1, F2, F3 = 414, 1569, 1886 mel; 332, 1966, 2670 Hz), very close to the mean value obtained for the English participants of 3.07 (F1, F2, F3 = 415, 1567, 1875 mel; 333, 1964, 2668 Hz). The spectral distance between 0.1 and 0.9 probability of /ɪ/ identification was 1.46 (ΔF1, ΔF2, ΔF3 = 61, -102, -62 mel; 57, -209, -159 Hz), approximately three times greater than that obtained for the English participants. The 0.1 to 0.9 distance was, however, sufficiently small to be within the criterion for categorical perception (i.e. it was less then 2, half the range of spectral values tested).

In summary, apart from the fact that SP02 had a fuzzier categorical boundary between English /i/ and /ɪ/ , her performance on the test was the same as that of native Canadian English listeners.

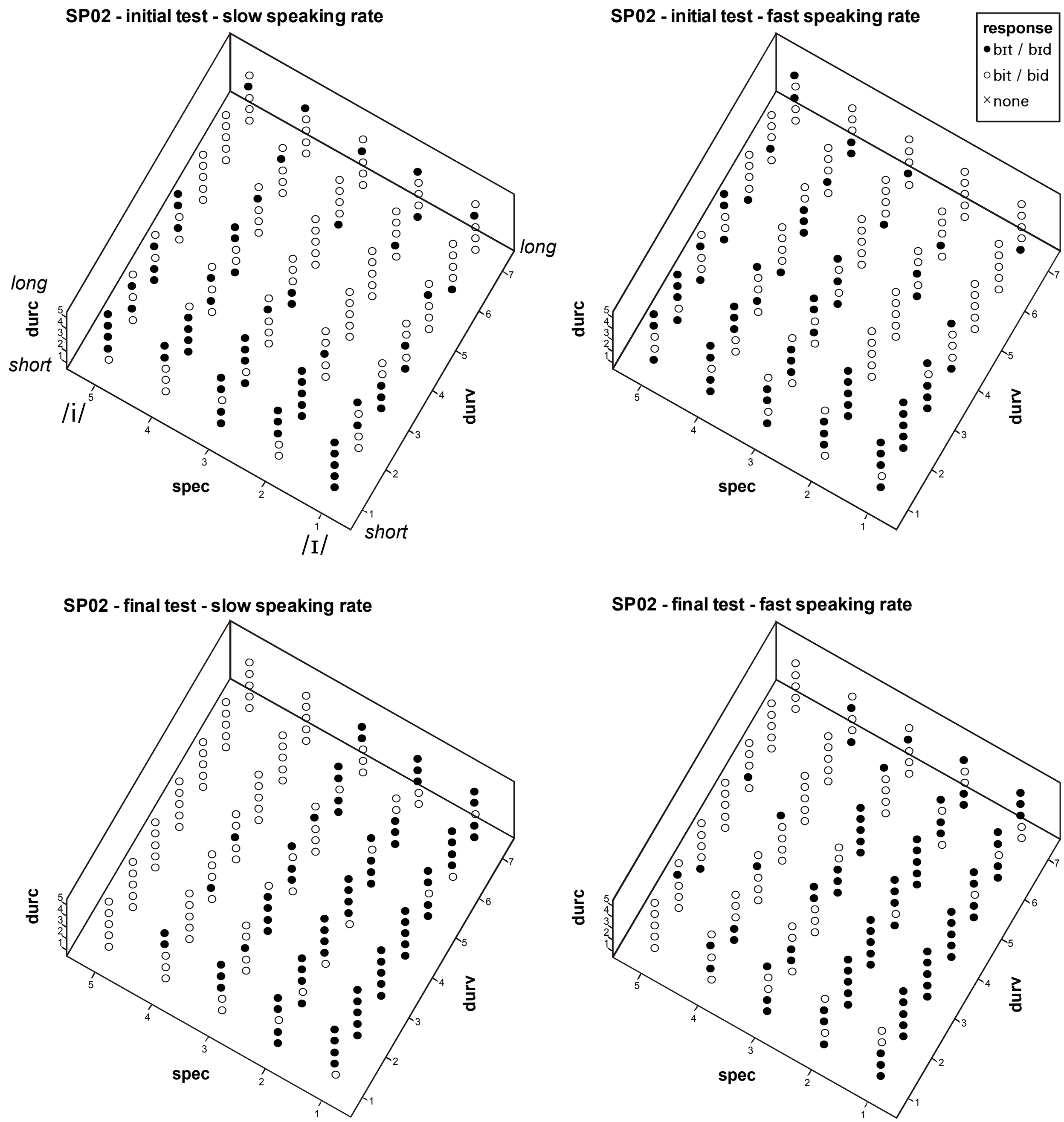


Figure 4.12 The English vowel identification pattern of Spanish participant SP02 in the initial and final English test.

Table 4.17 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of /i/ and /ɪ/ by participant SP02 in the final test.

	Coefficients	
	Unstandardised	Standardised
spec	0.906	0.996
durv	0.099	0.197
constant	-3.113	

Centroid Values	
ɪ	i
-0.882	0.778

Table 4.18 Classification results for discriminant analysis carried out on the identification of /i/ and /ɪ/ by participant SP02 in the final test.

Classification Results[b,c]

		Actual Group Membership	Predicted Group Membership ɪ	i	Total
Original / Cross-validated	Count	ɪ	129	35	164
		i	41	145	186
	%	ɪ	78.7	21.3	100.0
		i	22.0	78.0	100.0

b. 78.3% of original grouped cases correctly classified.

c. 78.3% of cross-validated grouped cases correctly classified.

In the final test, participants SP03, SP04, and SP06 had a categorical boundary between English /i/ and /ɪ/ based on vowel duration. A discriminant analysis was conducted on their results to determine the location of their categorical boundary. The only variable entered into the model was vowel duration and the cross-validated correct classification rate was 79.4% (see Table 4.19). Λ = .609 and $\chi^2(1, N = 1037) = 512.362$, $p < .0005$. The 0.1, 0.5, and 0.9 probability of /ɪ/ identification points were calculated from the coefficients and function values in Table 4.16. The categorical boundary point between these participants' perception of /i/ and /ɪ/ was 4.20 (107 ms). The vowel duration distance between 0.1 and 0.9 probability of /ɪ/ identification was 2.13 (42 ms) indicating a relatively fuzzy boundary. The distance between the 0.1 and 0.9 probability points was much less in the final test than that calculated for the same participants in the initial test, the latter being 6.59 (133 ms) centred on a 0.5 probability point of 4.02 (104 ms). The model included only vowel duration as an independent variable. (Discriminant function coefficients and centroid values for this model are given in Table 4.21, and classification results are given in Table 4.22. $\Lambda = .908$ and $\chi^2(1, N = 1028) = 98.911$, $p < .0005$.) Their boundary in the final test was, therefore, much sharper than it had been in the initial test and met the criterion for categorical perception. Figure 4.13 shows a visual comparison of the three participants' non-categorical perception in the initial test and categorical perception in the final test. Figure 4.14 shows an example of one listener's change in perception from the initial to the final test.

In summary, although the categorical boundary was somewhat fuzzy, three Spanish participants had developed categorical perception of English /i/ and /ɪ/ based on vowel duration.

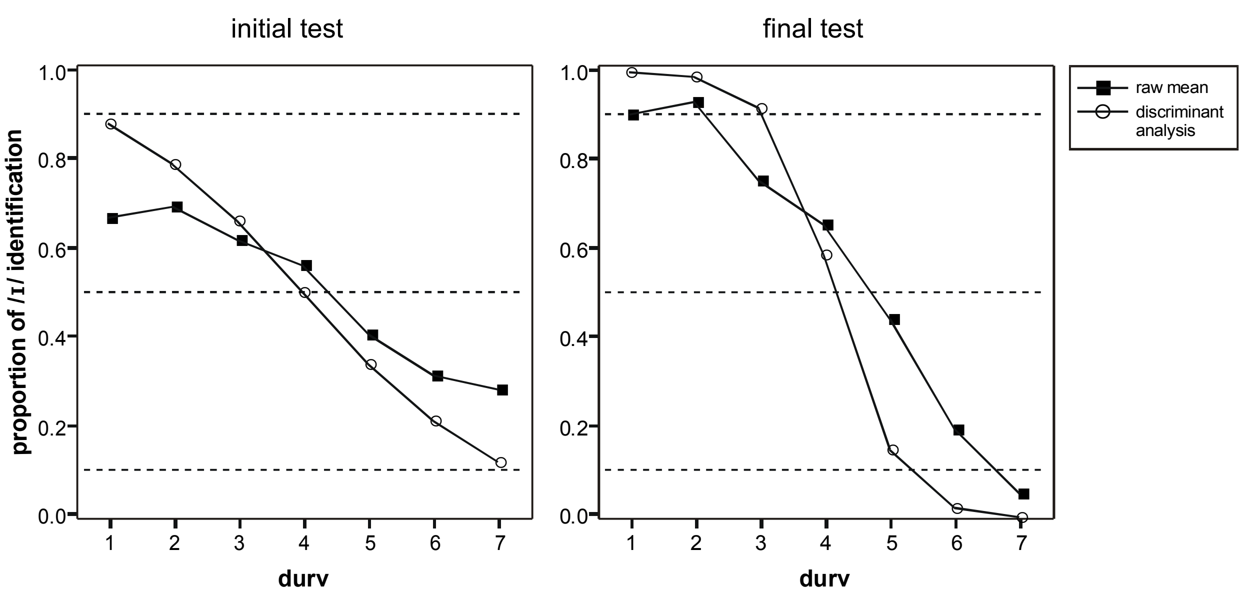


Figure 4.13 Proportion of stimuli identified as /ɪ/ by Spanish participants SP03, SP04, and SP06 in the initial test (left) and the final test (right). Squares represent mean values for given durations. Circles represent probabilities derived from the discriminant analysis. Dashed lines represent 0.1, 0.5 and 0.9 probability thresholds.

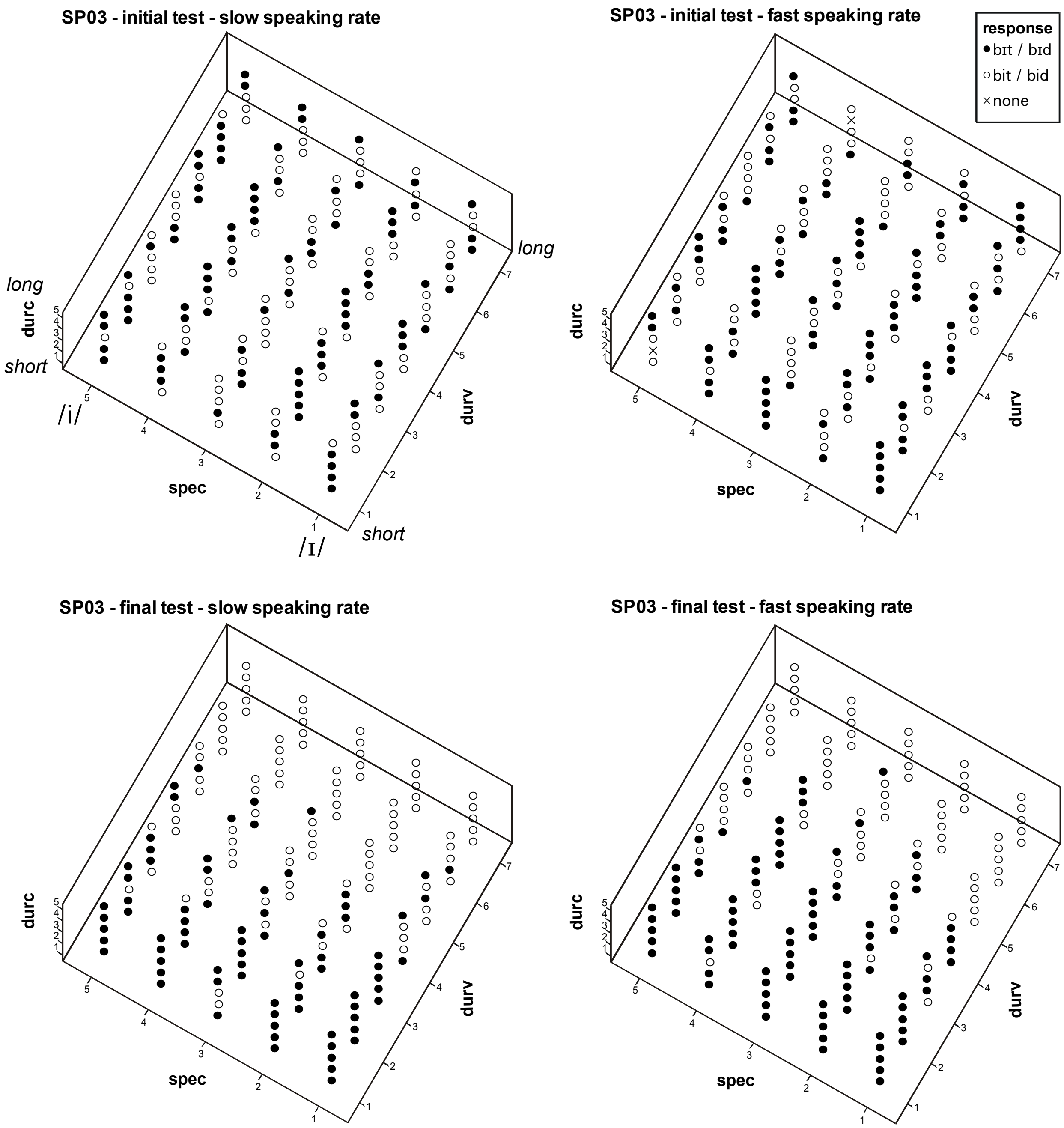


Figure 4.14 The English vowel identification pattern of Spanish participant SP02 in the initial and final English tests.

Table 4.19 Classification results for discriminant analysis carried out on the identification of /i/ and /ɪ/ by participants SP03, SP04, and SP06 in the final test.

Classification Results[b,c]

		Actual Group Membership	Predicted Group Membership ɪ	i	Total
Original / Cross-validated	Count	ɪ	480	103	583
		i	111	343	454
		none	9	4	13
	%	ɪ	82.3	17.7	100.0
		i	24.4	75.6	100.0
		none	69.2	30.8	100.0

b. 79.4% of original grouped cases correctly classified.

c. 79.4% of cross-validated grouped cases correctly classified.

Table 4.20 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of /i/ and /ɪ/ by participant SP03, SP04, and SP06 in the final test.

	Coefficients	
	Unstandardised	Standardised
durv	0.64	1
constant	-2.568	

Centroid Values	
ɪ	i
-0.706	0.906

Table 4.21 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of English /ɪ/ and /i/ by Spanish listeners SP03, SP04, and SP06 in the initial test.

	Coefficients	
	Unstandardised	Standardised
durv	0.524	1
constant	-2.104	

Centroid Values	
ɪ	i
-0.314	0.322

Table 4.22 Classification results for discriminant analyses carried out on the identification of English /ɪ/ and /i/ by Spanish listeners SP03, SP04, and SP06 in the initial test.

Classification Results[b,c]

		Actual Group Membership	Predicted Group Membership ɪ	i	Total
Original / Cross-validated	Count	ɪ	372	148	520
		i	214	294	508
		Ungrouped cases	14	8	22
	%	ɪ	71.5	28.5	100.0
		i	42.1	57.9	100.0
		Ungrouped cases	63.6	36.4	100.0

b. 64.8% of original grouped cases correctly classified.

c. 64.8% of cross-validated grouped cases correctly classified.

The fifth Spanish listener, SP01, had an identification pattern for English /i/ and /ɪ/ in the final test that was essentially unchanged from the initial test: He gave *none* responses to the majority of stimuli with spectral values 1, 2, and 3.

In summary, as predicted above in section 4.7.1, by the final English test the majority of Spanish participants had developed categorical perception of English /i/ and /ɪ/. Three identified the English vowels using vowel duration. One had a native-English-like boundary based on vowel spectral properties. Their categorical boundaries were somewhat fuzzy which may be indicative of an intermediate stage in development.

Chapter 5

General Discussion and Conclusions

5.1 General Discussion

5.1.1 Japanese listeners' English and Japanese categorical boundaries

In both the initial English test, the Japanese listeners made some use of spectral properties but primarily identified the English stimuli according to vowel duration. They also primarily identified the Japanese stimuli according to vowel duration. The locations of the Japanese listeners' categorical boundaries for Japanese voiceless-consonant stimuli (matching the silent consonant closure of the English stimuli) and the English spectral-value-4 stimuli (matching the spectral value of the Japanese stimuli) were effectively identical: The 6 ms difference between the English /i/-/ɪ/ boundary and the Japanese /iː/-/i/ boundary was within the 0.1 to 0.9 probability range for English lax vowel and Japanese short vowel identification. This result is consistent with the first part of Hypothesis 1: "Japanese listeners with limited exposure to English will identify English /i/ and /ɪ/ according to the same duration criteria they use to identify Japanese /iː/ and /i/."

Additional evidence is needed to support the claim that Japanese listeners have the same categorical boundary in English and Japanese because the Japanese listeners transfer their L1 perception pattern, rather than because of a general L2 learning strategy. One source of evidence comes from a comparison of Japanese listeners' and Spanish listeners' results. Since Spanish lacks any duration contrast, any use of duration by Spanish listeners would be due to a general L2 learning strategy and not to transfer. Although the Spanish listeners used vowel duration to identify English /i/ and /ɪ/ in the initial English test, they did not have a categorical perception pattern like that of the Japanese listeners. Since Japanese and Spanish have similar five-vowel-quality systems but differ in that Japanese, but not Spanish, has a long-short vowel contrast, the difference in perception patterns between the Japanese and Spanish listeners suggests that the Japanese listeners' categorical perception of the English vowels is due to transfer of the Japanese long-short contrast.

Since three of the five Spanish listeners developed a duration-based categorical boundary between English /i/ and /ɪ/ by the time of the final English test, it could be argued that the difference in perceptual patterns between the Japanese and Spanish listeners in the initial test was due to the Japanese listeners being further along a developmental process than the Spanish listeners, i.e. that the Japanese listeners had earlier had non-categorical duration-based perception of English /i/ and /ɪ/, but had developed categorical perception by the time of the initial test. Such an explanation would be tenable if the Japanese listeners had more English experience than the Spanish listeners; however, the opposite was the case: The Spanish listeners had studied English for almost twice as many years as the Japanese listeners (12-15 compared to 7-9 years) and had started studying English at a much younger age (4-6 compared to 13-14).[1] The weight of evidence therefore supports the hypothesis that the Japanese listeners identified English /i/ and /ɪ/ via assimilation to Japanese /iː/ and /i/ according to the same long-short vowel duration criteria used to identify Japanese /iː/ and /i/.[2]

5.1.2 Spanish listeners' initial perception of English /i/ and /ɪ/

In the initial test, the Spanish listeners were able to distinguish the English vowels to a small degree, but their perception was clearly not categorical. This result is compatible with a category-goodness assimilation (PAM) and therefore supports Hypothesis 3: "Spanish listeners will not initially have categorical perception of English /i/ and /ɪ/. They will assimilate English /i/ and /ɪ/ to Spanish /i/ via a category-goodness assimilation pattern, (with English /i/ as a good match for Spanish /i/ and English /ɪ/ a poor match)[3]. Spanish listeners' ability to distinguish English /i/ and /ɪ/ will be poor." What ability the Spanish listeners had to distinguish English /i/ and /ɪ/ was based on vowel duration, consistent with the Desensitisation Hypothesis (Bohn, 2002).

5.1.3 Effects of exposure on "similar" vowel categories

Six months compared to one month of living in an English speaking society had no effect on the Japanese listeners' perception of English /i/ and /ɪ/. There was no difference in the vowel duration or spectral properties of the categorical boundary. The mean duration of the boundary was unchanged at 108 ms, and the relative weighting of the vowel duration and vowel spectral cues was

---

[1]Given the geographical proximity of Mexico to the US, the Mexican participants may also have received a better quality of experience with aural English compared to the Japanese participants.

[2]An alternative explanation for how Japanese listeners could be farther along in a developmental process with less exposure to English is considered in Section 5.1.5.

[3]No evidence was provided in the present vowel perception study relating to the section in parenthesis.

almost identical (63 and 28% respectively compared to 62 and 26%). This result supports the second part of Hypothesis 1, and Hypothesis 2:

> ***Hypothesis 1 second part:*** English /i/ before a voiced consonant is assimilated to Japanese /iː/; and English /i/ before a voiceless consonant, and English /ɪ/ are assimilated to Japanese /i/. English vowels assimilated to Japanese /i/ are identified as English /ɪ/, and English vowels assimilated to Japanese /iː/ are identified as English /i/.
>
> ***Hypothesis 2.*** Experience with English will have little effect on Japanese listeners' perception of English /i/ and /ɪ/. The assimilation pattern in Hypothesis 1 results in two diaphone categories: English /i/ before a voiced consonant plus Japanese /iː/; and English /i/ before a voiceless consonant, plus English /ɪ/, plus Japanese /i/. The properties of the diaphone categories will initially be close to the norms for the Japanese vowels but will eventually be intermediate between the norms for the Japanese and English vowels which contribute to each diaphone. Since English /i/ and /ɪ/ allophones are assimilated to Japanese /iː/ and /i/ according to Japanese vowel duration criteria, English vowel allophones contributing to a diaphone have duration properties within the range for the Japanese vowel contributing to that diaphone. Hence, all vowel allophones contributing to the diaphone categories have the same duration properties as Japanese /iː/ or /i/, and the duration properties of the diaphone categories will continue to be identical to Japanese /iː/ and /i/. One diaphone consists of Japanese /iː/ and an allophone of English /i/; since these vowels have very similar spectral properties there will be no change in the spectral properties of this diaphone. The other diaphone consists of Japanese /i/, an allophone of English /i/, and English /ɪ/; although English /ɪ/ differs spectrally from Japanese /i/, movement of the properties of the diaphone category towards English /ɪ/ will be tempered by English /i/ which is spectrally very similar to Japanese /i/. Hence the spectral properties of this diaphone category will remain close to Japanese /i/ and move only slightly towards English /ɪ/.

Note that although a change in the spectral properties was predicted for the latter diaphone in Hypothesis 2, no change was observed. However, since the size of the change was predicted to be slight, the lack of observed change does not necessarily lead to the falsification of the claim that the Japanese listeners formed two diaphone categories.

A lack of change in the Japanese listeners' perceptual pattern could be indicative of them having reached the final stage that they will ever reach in a developmental process. However, given that evidence has been presented in support of the first part of Hypothesis 1 "Japanese listeners will initially identify English /i/ and /ɪ/ according to the same duration criteria they use to identify Japanese /iː/ and /i/," (see Section 5.1.1) it is argued that this indirectly supports the second part of Hypothesis 1 (see above), since use of Japanese duration criteria would lead to the assimilation of English /ɪ/ to a short Japanese vowel and allophone specific assimilation of English /i/. Therefore, the initial stage in Japanese listeners' perception of English /i/ and /ɪ/ according to Hypothesis 1, is the same as the final stage in the putative developmental process. This suggests that there is no developmental process and that the Japanese listeners' perception of English /i/ and /ɪ/ remains unchanged because of the way they assimilate these vowel to Japanese categories as outlined in Hypotheses 1 and 2. The weight of evidence therefore supports Hypothesis 2.

5.1.4 Effects of exposure on a "new" vowel category

Four out of five Spanish participants developed a categorical boundary between the English vowels (three developed a duration based boundary and one developed a spectrally based boundary). These results therefore generally support Hypothesis 4: "Experience with English will improve Spanish listeners' ability to distinguish English /i/ and /ɪ/. Since English /i/ is very similar to Spanish /i/, these two vowels will form a diaphone category. Since English /ɪ/ is only a poor match for Spanish /i/, a new category will be established for English /ɪ/. The new category will develop first duration and then spectral properties that are close to the English norm for /ɪ/." It is not clear whether the vowel-duration-based categorical boundary developed by three of the Spanish listeners represented an intermediate stage leading to a spectrally-based categorical boundary or a final stage in their development of English /i/-/ɪ/ perception. A more extended longitudinal study with a greater number of tests would be needed to determine whether this developmental pattern, hypothesised by Escudero (2000, 2001a), is viable.

An alternative explanation for the development of the categorical duration-based boundary was suggested by A. M. Schmidt (personal communication, June 4, 2002). She rejected the claim that a "new" category had been formed. Instead she suggested that since in Mexican high schools, students are taught that English /i/ is longer than English /ɪ/, the listeners applied this strategy, and improved in their ability to use it between the initial and final test because they had had five months of practice using the strategy in an English speaking environment.[4] Schmidt's explanation would fail to account for the results of the Spanish listener who developed a spectrally based categorical boundary. Further testing with a larger number of participants would be required to determine whether the results of the latter listener are exceptional or representative of a subgroup of the larger population.

5.1.5 Other issues arising from the results

An issue arising from the results of the current research is whether the Japanese listeners' secondary use of duration cues in their perception of English /i/ and /ɪ/ was due to exposure to English or due to L1 experience. If, as argued in Sections 5.1.1 and 5.1.3, English /i/ and /ɪ/ are assimilated to Japanese /iː/ and /i/ according to Japanese long-short vowel criteria, this would result in one diaphone category that includes examples of English /i/ and another that includes examples of English /i/ and English /ɪ/. The influence of the /ɪ/ on the latter diaphone category may cause it to develop different spectral properties from the former diaphone category.[5] Given that the Japanese listeners' assimilation pattern would be expected to assign examples of the English vowels to two different Japanese vowels from the time of their very first exposure to English, they may have developed the spectral properties of the diaphone category before their arrival in Canada. In contrast the Spanish listeners would be expected to assimilate English /i/ and /ɪ/ to the single Spanish /i/ category at the time of their first exposure to English. This was their assimilation pattern at the time of the initial English test shortly after their arrival in Canada. Despite many years of studying English in Mexico, they did not begin to develop two categories for the perception of English /i/ and /ɪ/ until after extended exposure to English in an English speaking environment. The lack of change in the Japanese listeners' response pattern from the initial to final English tests would then be explained as due to them having already reached the final stage in a developmental process by the time of the initial test. Again, the difference in the Japanese and the Spanish listeners' perception

---

[4] In Japanese high schools, students are also taught that English /i/ is longer than English /ɪ/.

[5] This may also result in an L2 influence on the L1 since the L1 as well as the L2 sound is based on the diaphone category (see Flege, 1995, p 241).

would be due to the existence of the long-short vowel contrast in Japanese but not in Spanish: The Japanese duration-based assimilation would have allowed the establishment of two diaphone categories which may have been spectrally modified by relatively little exposure to English. In contrast a relatively large exposure to English failed to separate the Spanish listeners' single diaphone category into two categories.

An alternative explanation for the Japanese listeners' secondary use of spectral cues in the perception of English /i/ and /ɪ/ is that spectral properties may be a secondary cue to the Japanese long-short vowel contrast: Fitzgerald (1996) found that Japanese speakers produced spectral differences between long and short vowels (long vowels being more peripheral in the vowel space) and that a statistical model was moderately successful at distinguishing long-short pairs using spectral information alone. Future research is planned to determine whether Japanese listeners do indeed make use of these spectral differences when identifying long and short vowels. If this is the case, then the secondary use of vowel spectral properties in the identification of English /i/ and /ɪ/ may be due to L1 experience rather than L2 learning.

Results in the study suggested that although Japanese and Spanish listeners made use of duration to distinguish English /i/ and /ɪ/, speaking rate had no, or minimal, effect on their use of duration, i.e., a faster speaking rate did not result in a reduction in the duration-based boundary. Since Spanish does not make use of duration it is not unreasonable to suppose that Spanish participants may have used absolute duration irrelevant of speaking rate. However, since Japanese does have vowel duration contrasts, some sort of speaking-rate normalisation might have been expected. It may be that Japanese listener apply very little normalisation in terms of shifts in long-short duration boundaries relative to speaking rate, this would be consistent with the findings of Strange et al. (1998), Guion et al. (2002), and Ingram & Park (1997).

5.2 Future Research

Production data were gathered from the participants during the same sessions in which the perception data were gathered. These data will be analysed and compared to the perception data. Given Flege's SLM hypothesis that production is based on perceptual categories and that production will eventually match perception (Flege, 1995), this will potentially shed additional light on the results reported in this thesis. If participants produce significant acoustic distinctions between vowel categories, then this may indicate that their ability to distinguish those vowels in the perception tests is due to categorical perception rather than an artifact of the testing technique.[6] If Japanese participants produce the Japanese English vowel pairs Japanese /i/ - English /ɪ/, and Japanese /iː/ - English /i/ with the same within-pair duration ranges, and with between pair separation, then this will support the claim that Japanese listeners assimilate English vowels to Japanese categories according to duration; and that English vowels assimilated to Japanese /i/ are identified as English /ɪ/, and that English vowels assimilated to Japanese /iː/ are identified as English /i/ (Hypothesis 1). Participants JP09 and SP01 had somewhat unexpected spectrally-based perception patterns for the English vowels. Comparison of their perception results with L1 and L2 production results may indicate the correct choice between the possible explanations for these patterns given in Chapter 4.

Data were also collected on the perception and production of the post vocalic /t/-/d/ voicing contrast. The vowel and consonant data will be analysed together to determine any interactions between the two. Since the speech perception task in real communication involves recognising sequences of segments in which some of the cues to the first segment may be synchronous with the second segment, and some of the cues to the second segment may be synchronous with the first segment, any research that examines the perception of only one of the segments gives only a partial picture of perceptual performance.

As discussed in Section 5.1.5, the results of this thesis also suggest that additional research, involving the collection of new data, should be carried out on whether spectral properties are a secondary perceptual cue to vowel duration in Japanese. Also as discussed in Section 5.1.5, additional longitudinal research is needed on the development of categorical English /i/-/ɪ/ perception by Spanish listeners/speakers. Such a study may need to cover a longer period of time than that in the current research and would need to collect data perhaps monthly or bi-monthly in order to test the developmental patterns hypothesised by Escudero (2000, 2001a, 2001b).

5.3 Summary of Conclusions

The results of the experiments conducted in this thesis were consistent with the claim that Japanese listeners had a duration-based categorical boundary between English /i/ and /ɪ/ that was in the same position as their categorical boundary between Japanese /iː/ and /i/. Six months',

---

[6]An example of such an artifact would be if a listener perceived a continuous set of within-phonemic-category stimulus properties, but, because the test required them to group them into categories, assigned them to different categories using non-phonemic but acoustically perceptible properties.

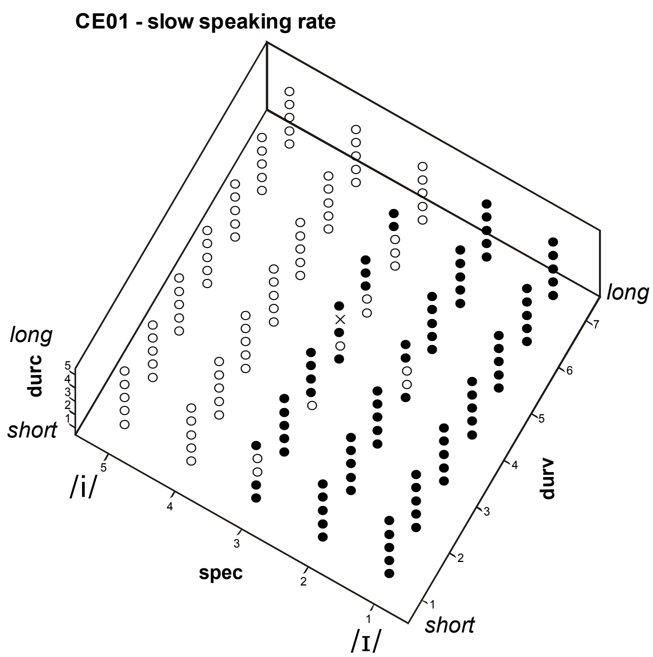
compared to one month's, exposure to an English speaking environment was not found to have any effect on the Japanese listeners' perception of the English vowels. This is consistent with the hypothesis developed by Morrison (2002a) that English /i/ before a voiced consonant is assimilated to Japanese long /iː/ but before a voiceless consonant it is assimilated to Japanese short /i/; that before both a voiced and voiceless consonant English /ɪ/ is assimilated to Japanese short /i/; that English vowels assimilated to Japanese long /iː/ are identified as English /i/, and that English vowels assimilated to Japanese short /i/ are identified as English /ɪ/. Under this hypothesis, since the English vowels are assimilated to Japanese categories according to Japanese duration criteria, they have similar durations to the Japanese members of the Japanese-English diaphone categories which are formed, and therefore do not change the duration criteria of the diaphones, which continue to match the Japanese vowel duration properties. Although some spectral changes were predicted due to the influence of English /ɪ/ on one of the diaphones, no such changes were observed. Since the changes were, however, predicted to be slight, the lack of observed change does not falsify the hypothesis that diaphone categories are formed as outlined above.

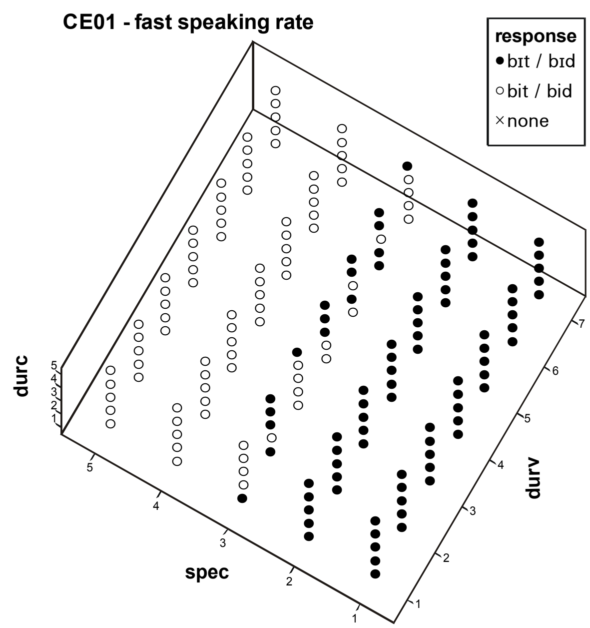
Spanish listeners did not have categorical perception of English /i/ and /ɪ/ after one month's residence in an English speaking environment, but the majority developed categorical perception after six months. It has been argued that the difference between the longitudinal results for the Japanese and Spanish listeners was because the Japanese listeners perceived the English vowels as similar to Japanese /iː/ and /i/, resulting in a form of two-category assimilation and the formation of diaphone categories; whereas, the Spanish listeners perceived English /i/ as a good match for Spanish /i/, and English /ɪ/ as a poor match, resulting in a category-goodness assimilation and the formation of a new category for English /ɪ/.

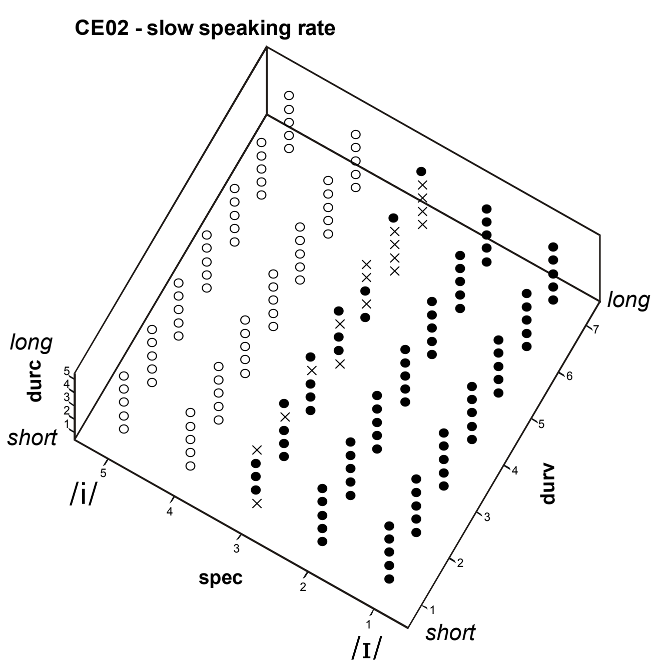
Appendix

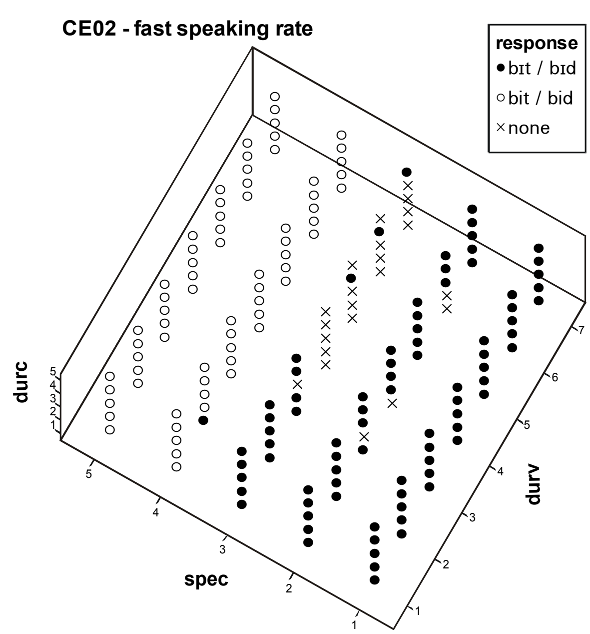
Individual Results

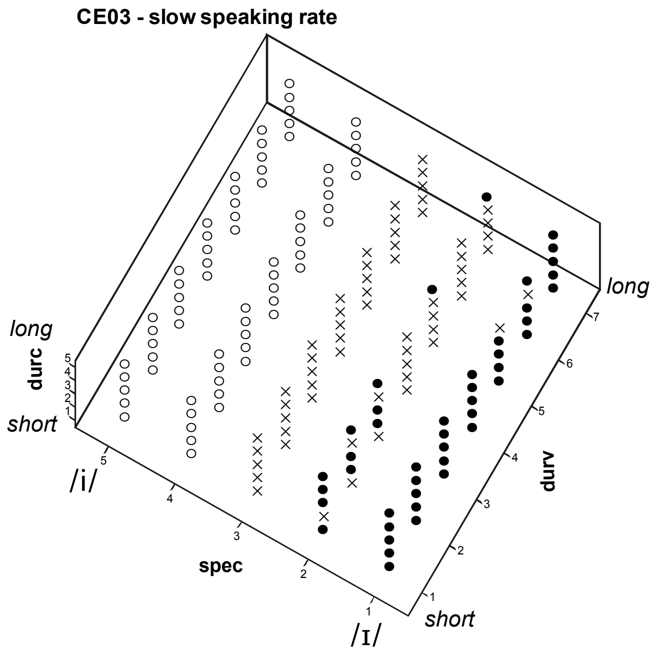
Canadian English listeners' perception of English vowels

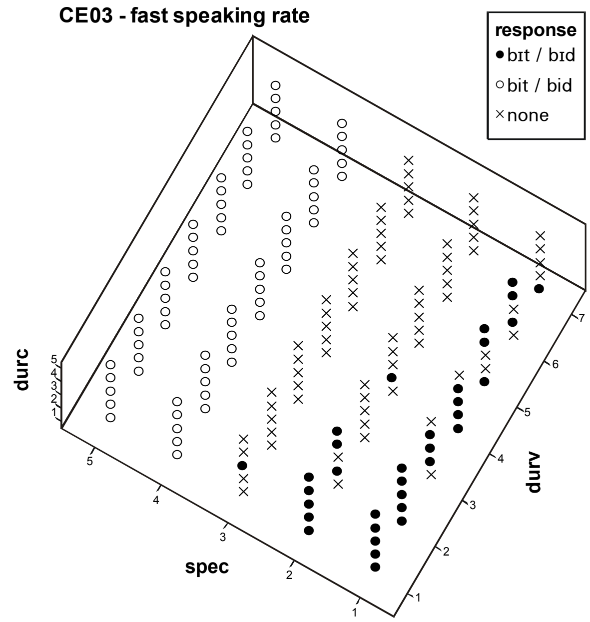


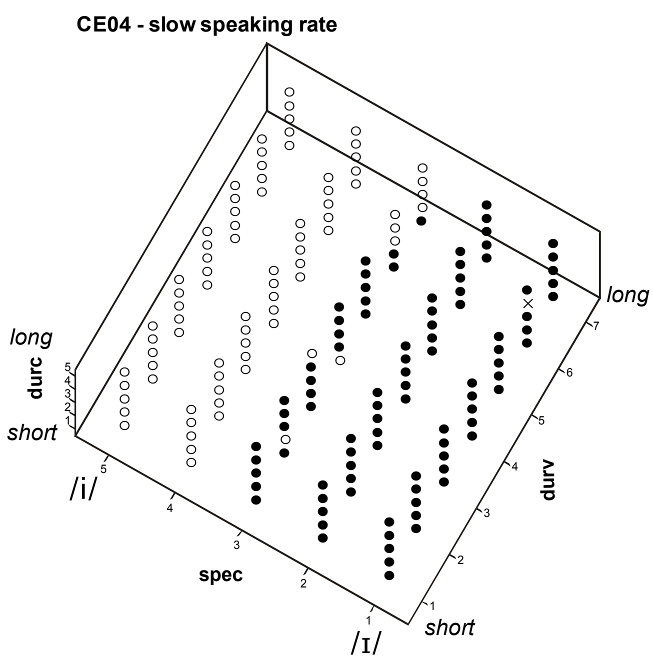


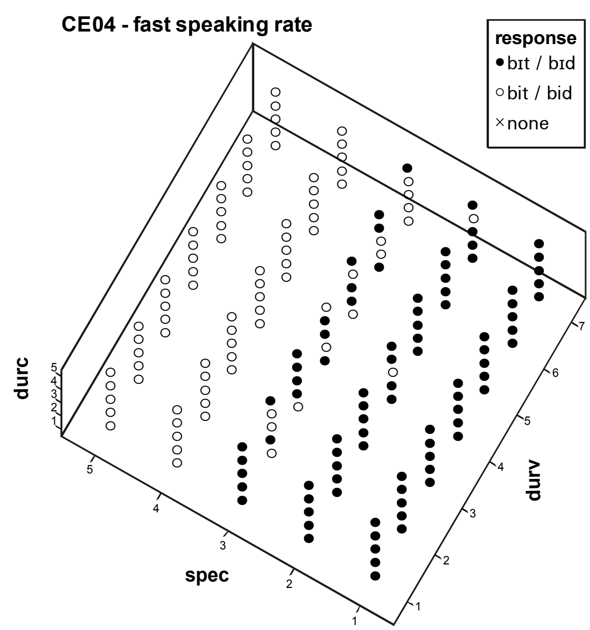


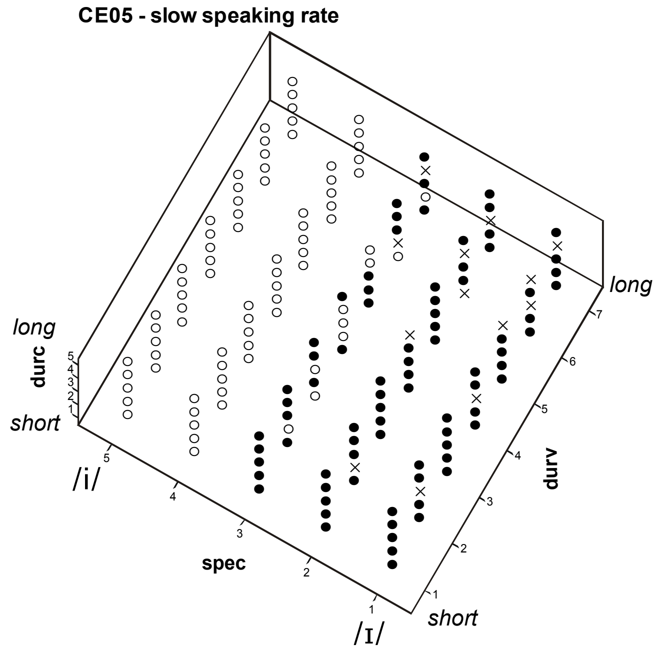


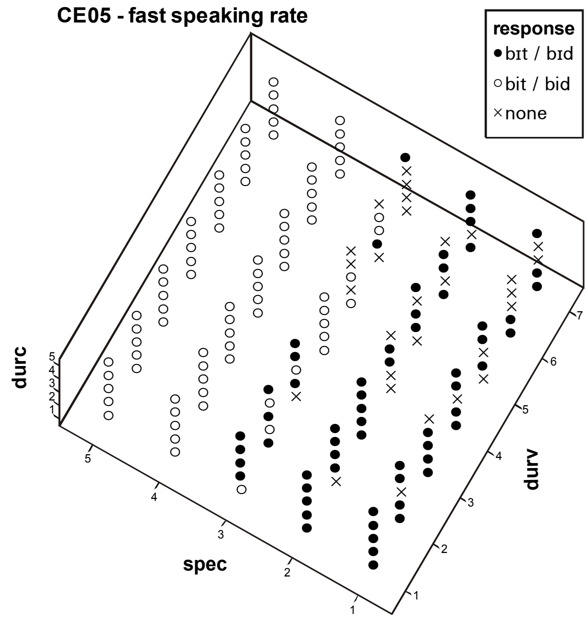

CE03 - slow speaking rate
long
durc
short
/i/
spec
/ɪ/
short
long
durv

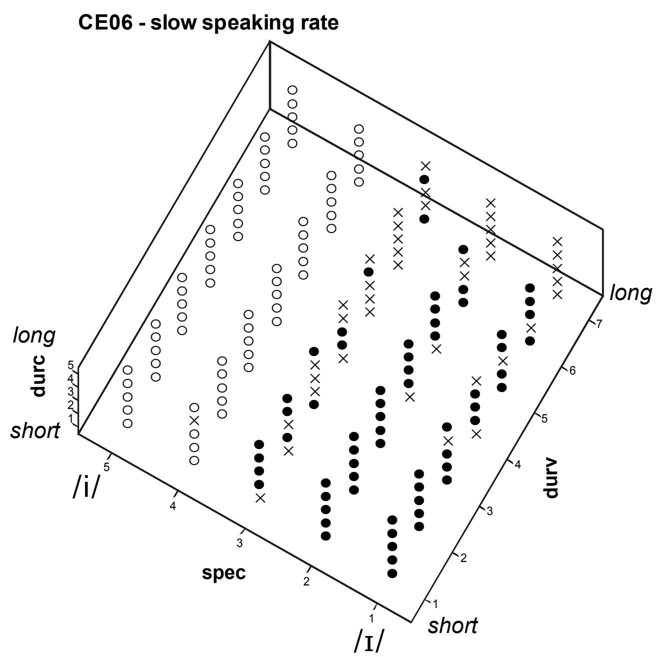

CE03 - fast speaking rate
response
bɪt / bɪd
bit / bid
none
durc
spec
durv

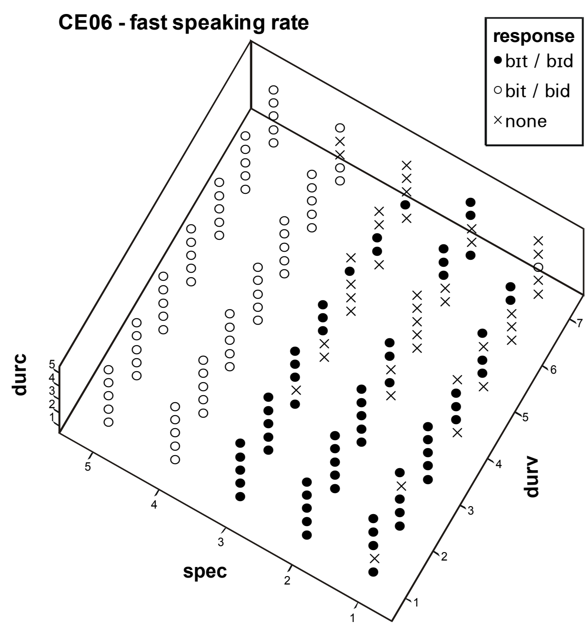

CE04 - slow speaking rate
long
durc
short
/i/
spec
/ɪ/
short
long
durv

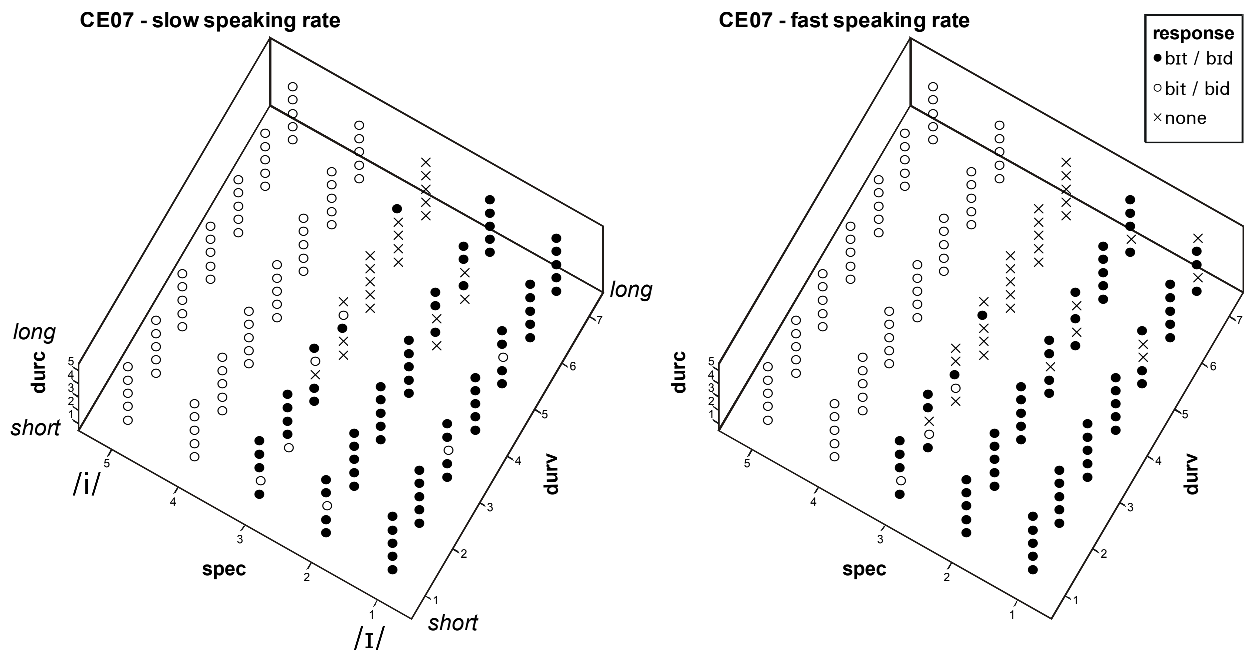

CE04 - fast speaking rate
response
bɪt / bɪd
bit / bid
none
durc
spec
durv


CE05 - slow speaking rate
long
durc
5
4
3
2
1
short
/i/
5
4
3
2
1
spec
/ɪ/
short
1
2
3
4
5
6
7
durv
long

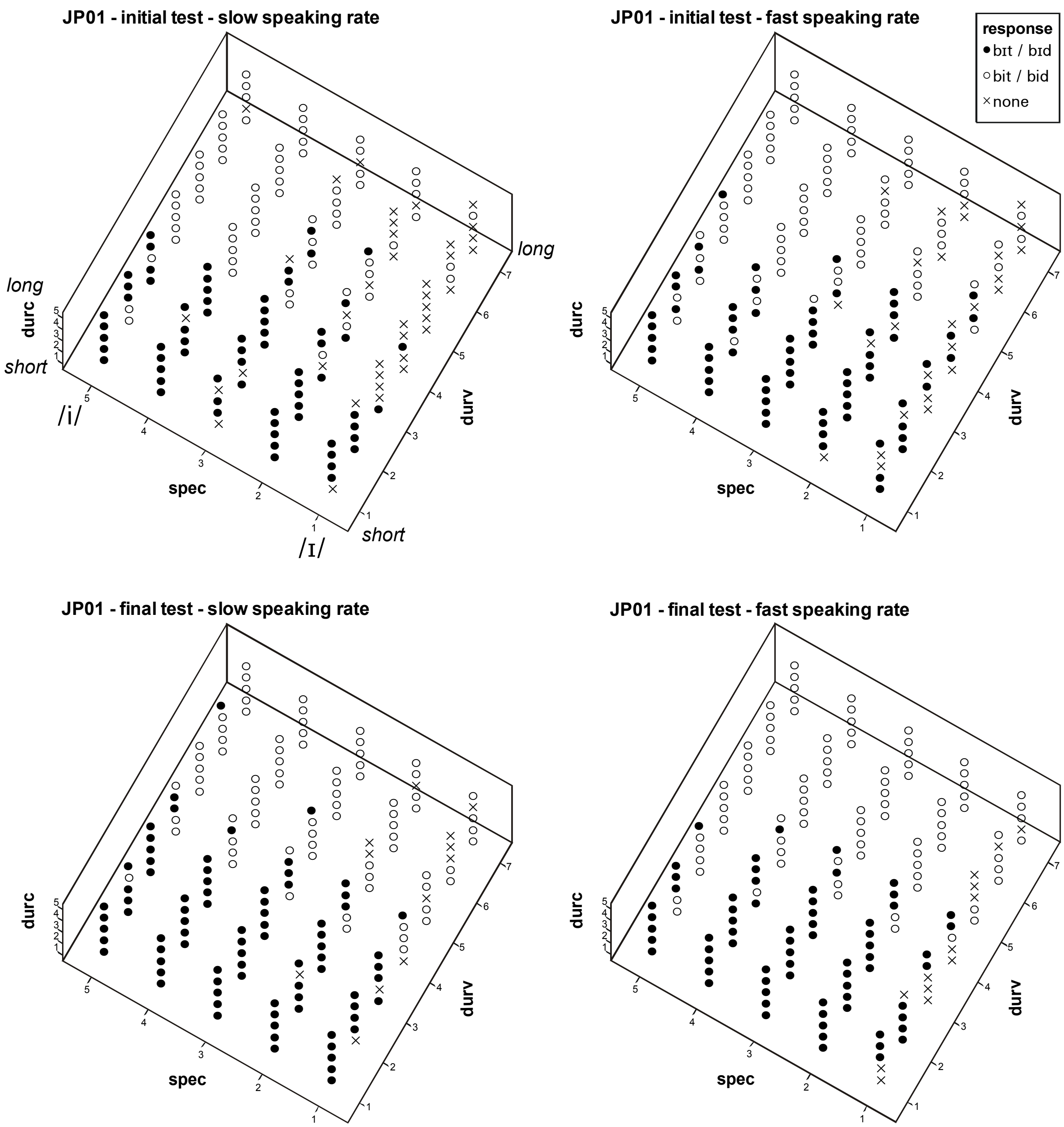

CE05 - fast speaking rate
response
bɪt / bɪd
bit / bid
none
durc
5
4
3
2
1
5
4
3
2
1
spec
1
2
3
4
5
6
7
durv

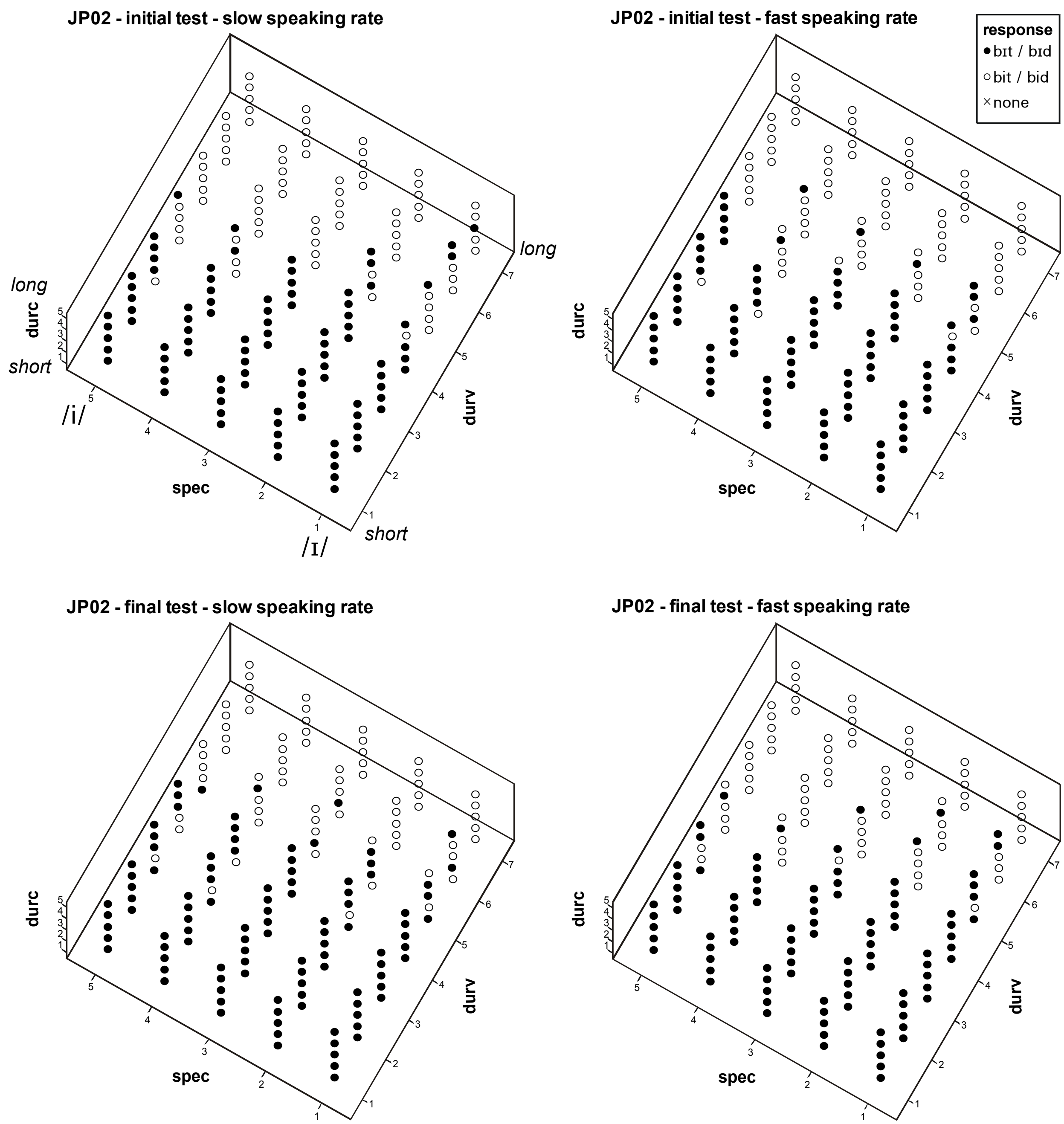

CE06 - slow speaking rate
long
durc
5
4
3
2
1
short
/i/
5
4
3
2
1
spec
/ɪ/
short
1
2
3
4
5
6
7
durv
long

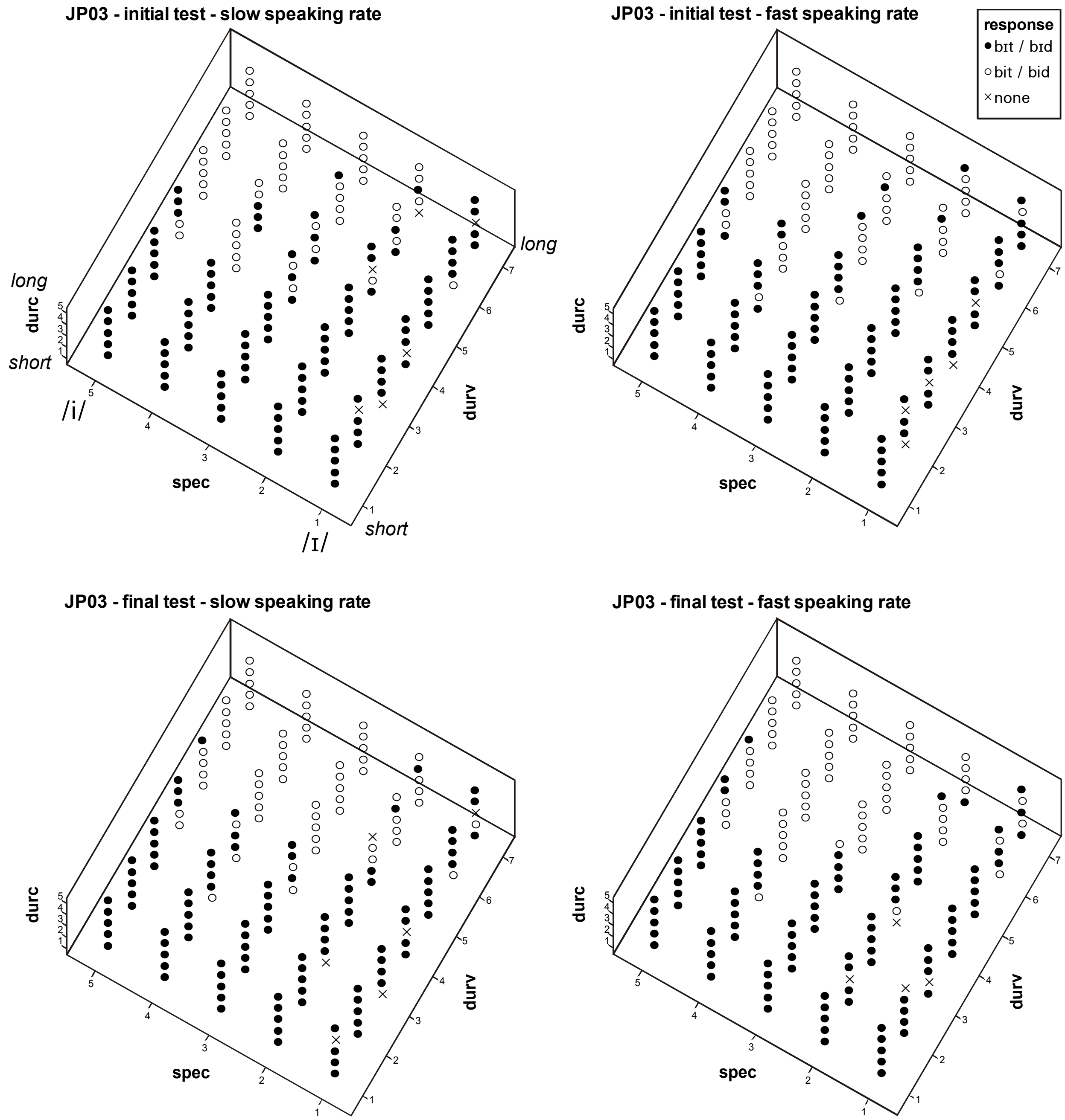

CE06 - fast speaking rate
response
bɪt / bɪd
bit / bid
none
durc
5
4
3
2
1
5
4
3
2
1
spec
1
2
3
4
5
6
7
durv

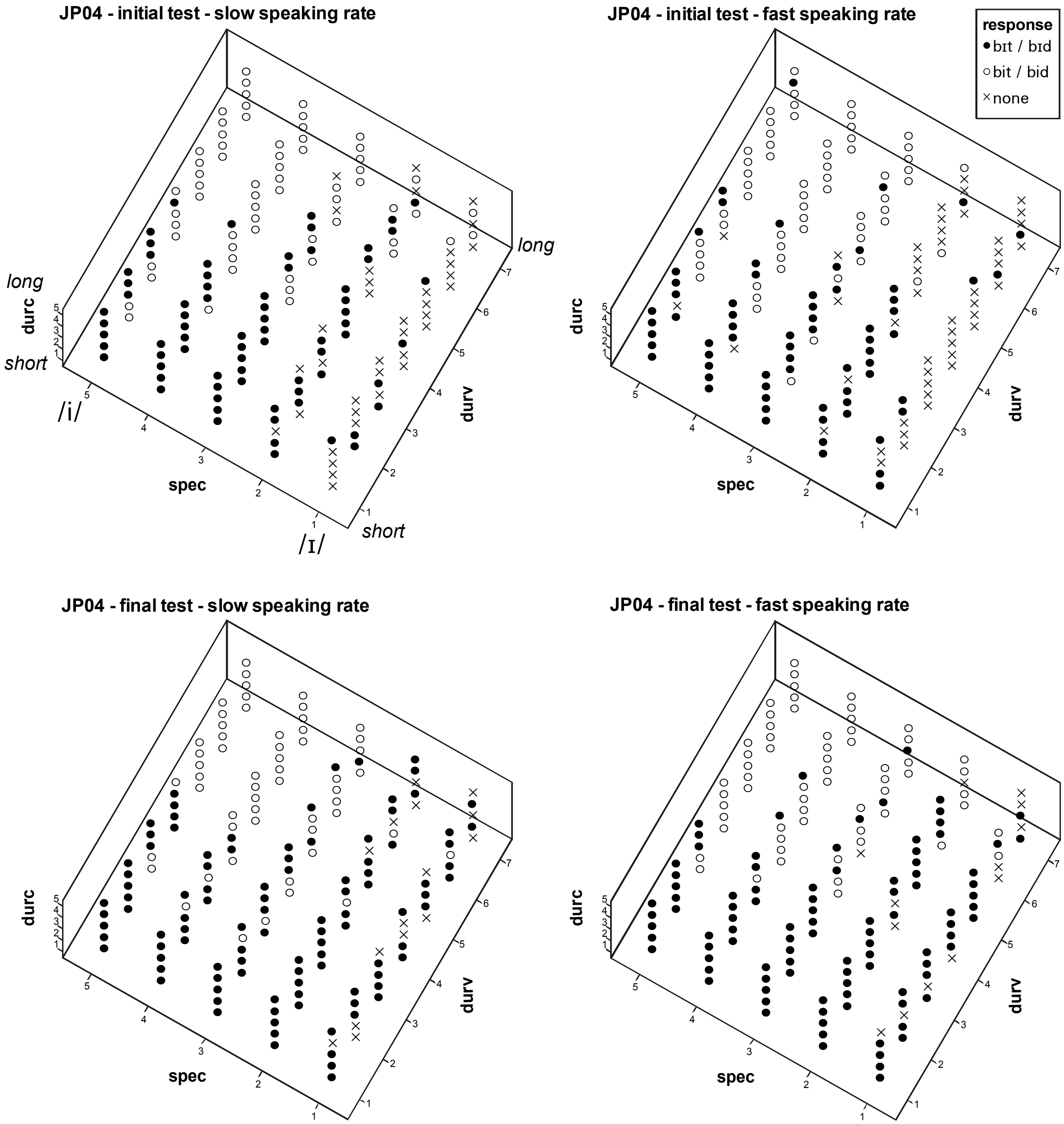

CE07 - slow speaking rate
CE07 - fast speaking rate
response
bɪt / bɪd
bit / bid
none
durc
durv
spec
long
short
/i/
/ɪ/

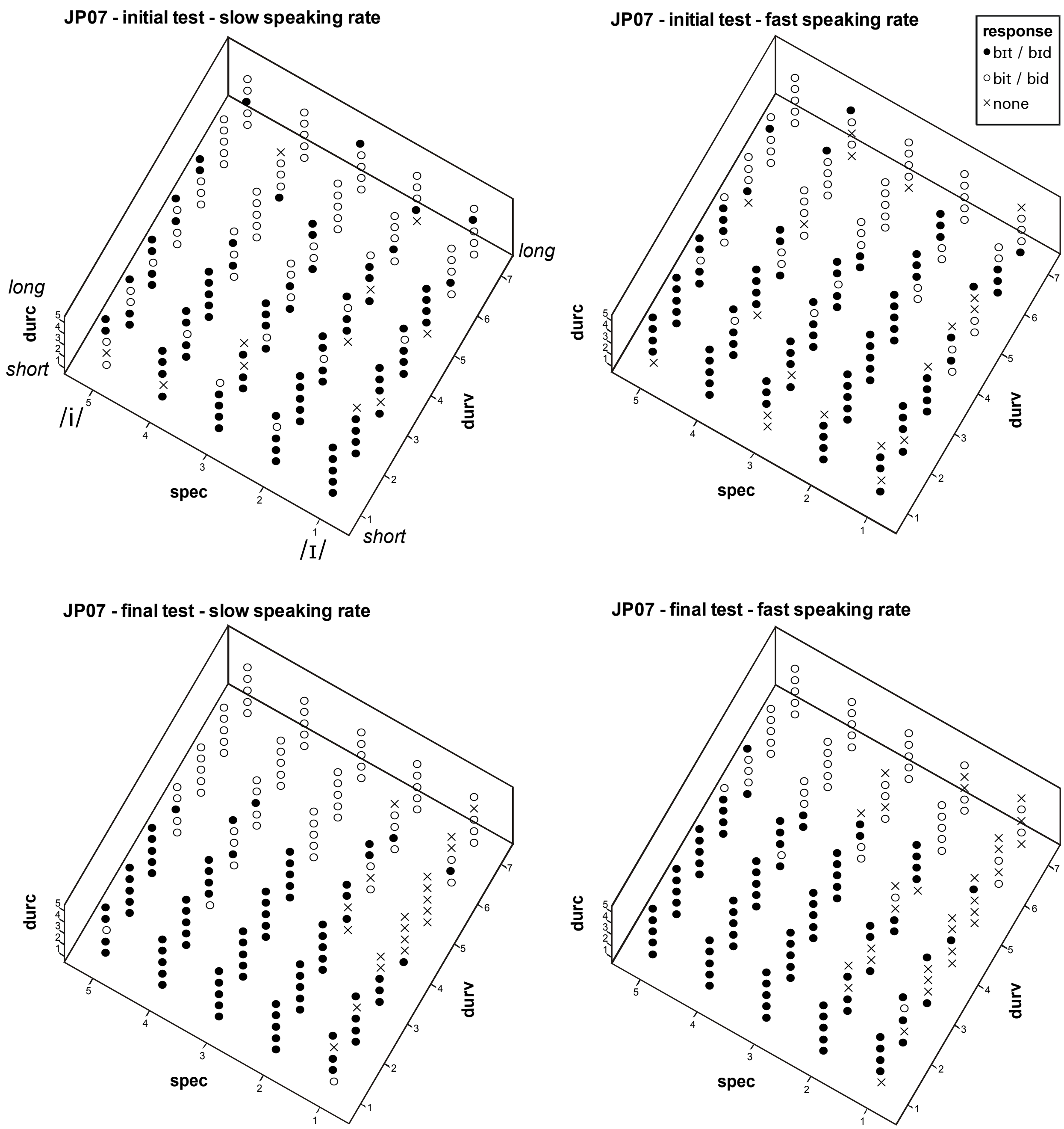
Japanese listeners' perception of English vowels

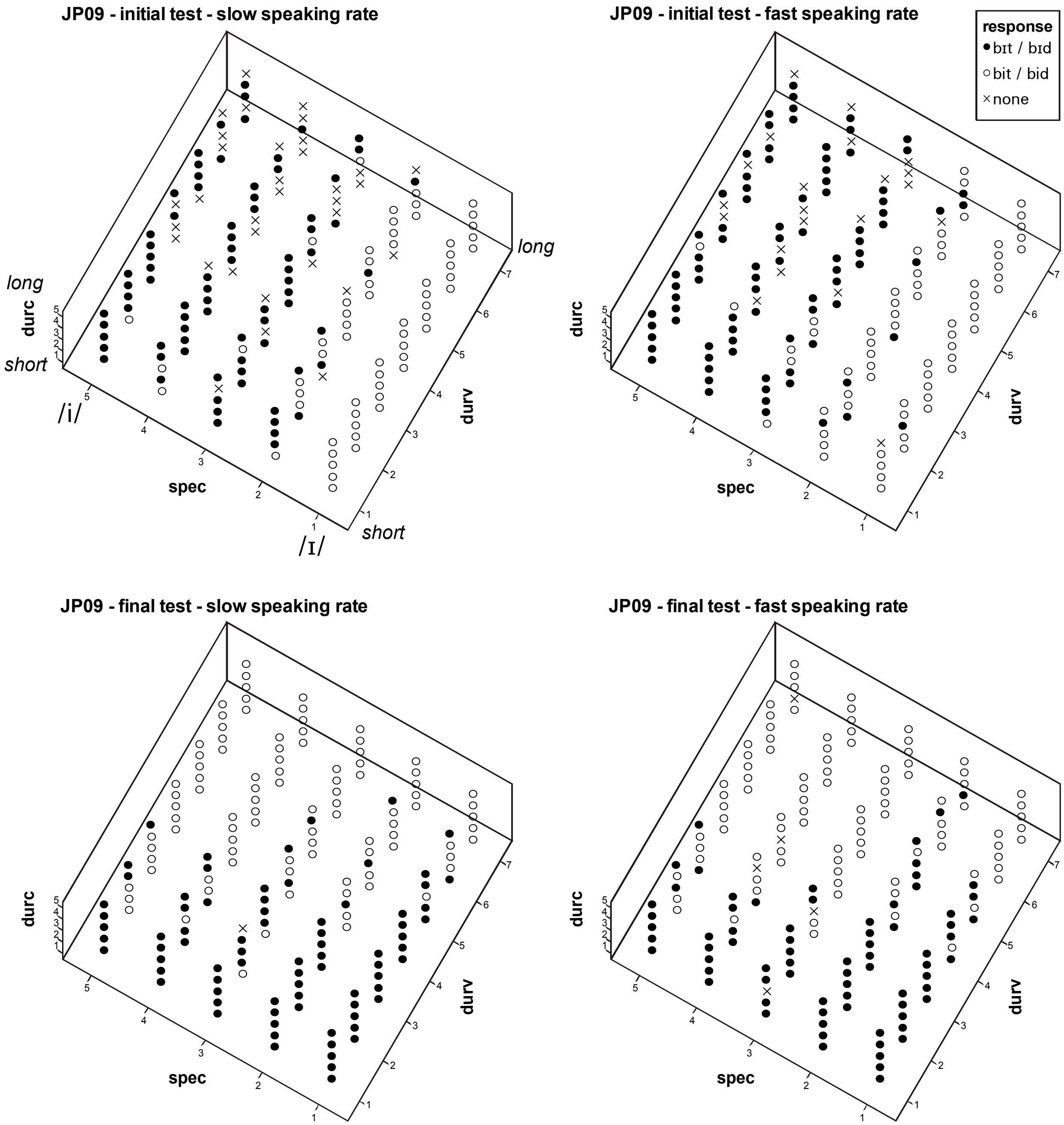



JP02 - initial test - slow speaking rate
JP02 - initial test - fast speaking rate
response
bɪt / bɪd
bit / bid
none
long
short
durc
/i/
/ɪ/
spec
durv
JP02 - final test - slow speaking rate
JP02 - final test - fast speaking rate

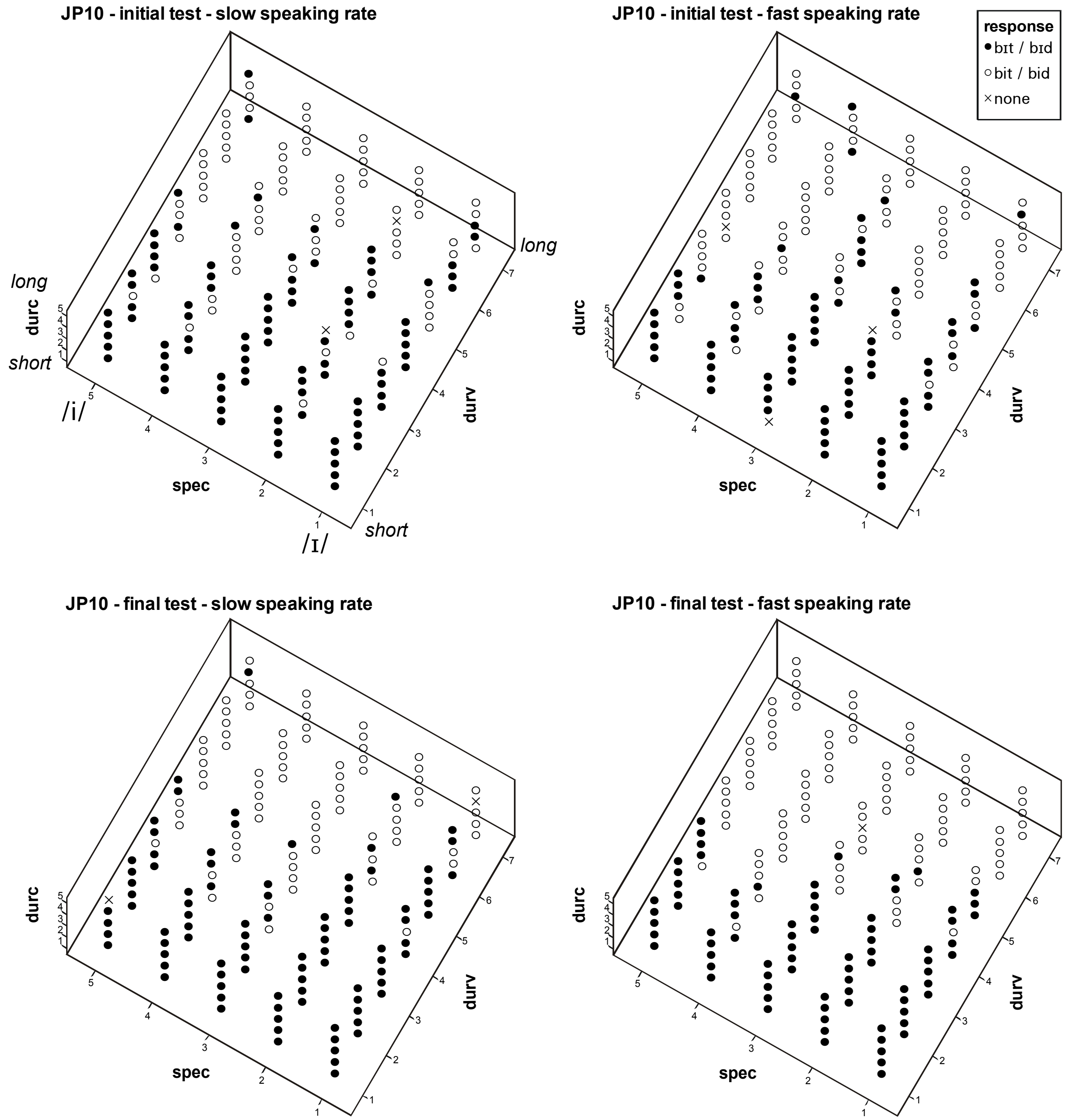

JP03 - initial test - slow speaking rate
JP03 - initial test - fast speaking rate
JP03 - final test - slow speaking rate
JP03 - final test - fast speaking rate
response
bɪt / bɪd
bit / bid
none
durc
durv
spec
long
short
/i/
/ɪ/


JP04 - initial test - slow speaking rate
JP04 - initial test - fast speaking rate
JP04 - final test - slow speaking rate
JP04 - final test - fast speaking rate
response
bɪt / bɪd
bit / bid
none
durc
long
short
spec
/i/
/ɪ/
durv
long
short


JP07 - initial test - slow speaking rate
JP07 - initial test - fast speaking rate
response
bɪt / bɪd
bit / bid
none
long
short
durc
/i/
/ɪ/
spec
durv
JP07 - final test - slow speaking rate
JP07 - final test - fast speaking rate

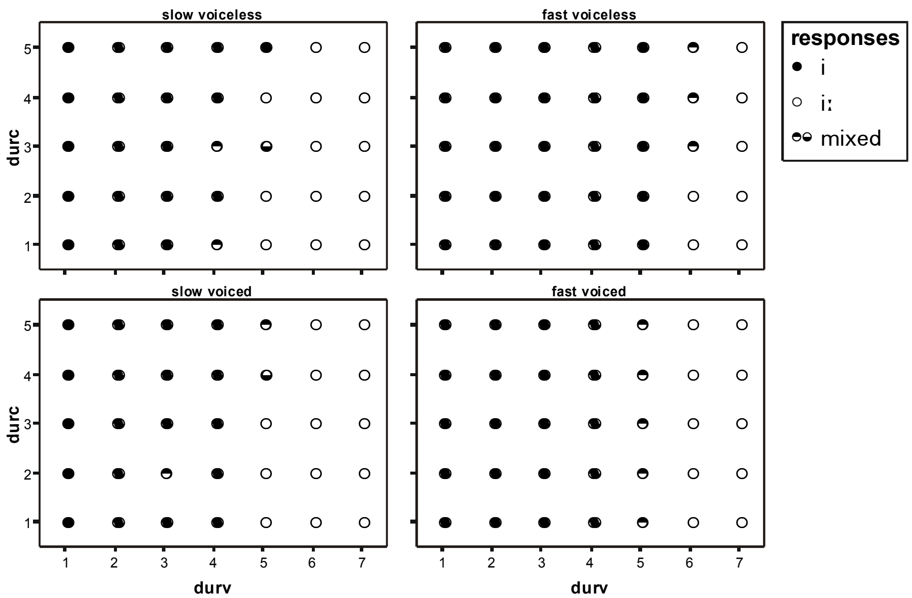


JP09 - initial test - slow speaking rate
JP09 - initial test - fast speaking rate
response
bɪt / bɪd
bit / bid
none
long
short
durc
/i/
/ɪ/
spec
durv
JP09 - final test - slow speaking rate
JP09 - final test - fast speaking rate

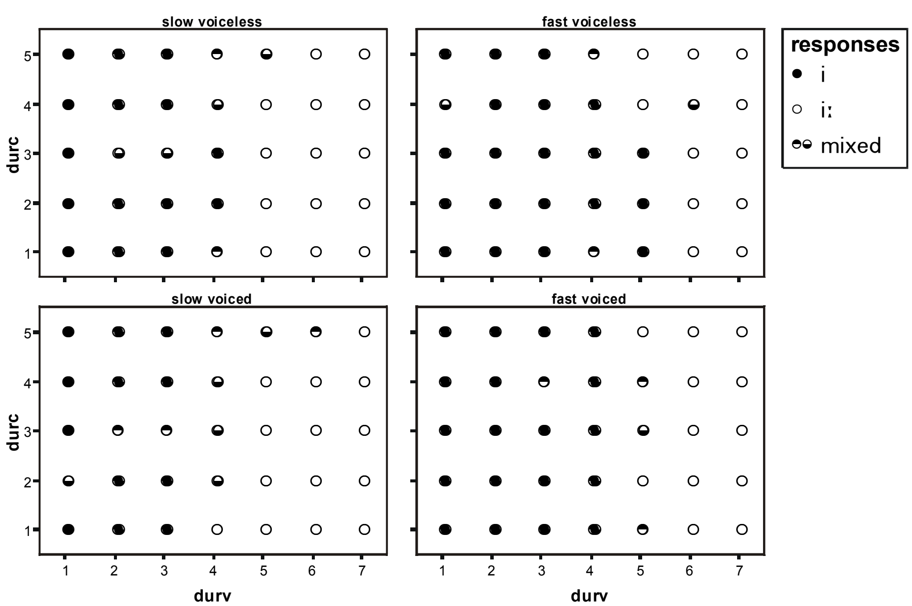

JP10 - initial test - slow speaking rate
JP10 - initial test - fast speaking rate
response
bɪt / bɪd
bit / bid
none
long
short
durc
/i/
/ɪ/
spec
durv
JP10 - final test - slow speaking rate
JP10 - final test - fast speaking rate

Japanese listeners' perception of Japanese vowels

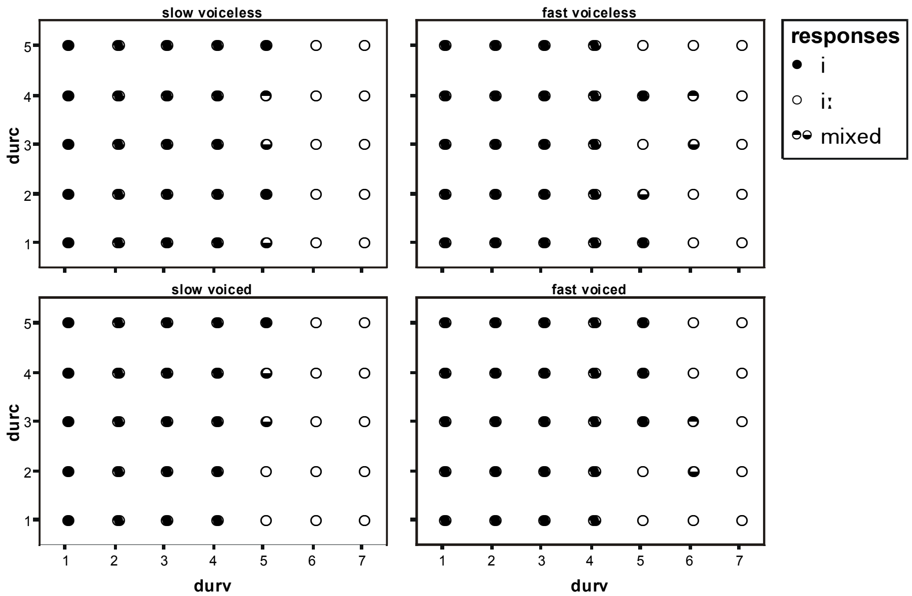
JP01



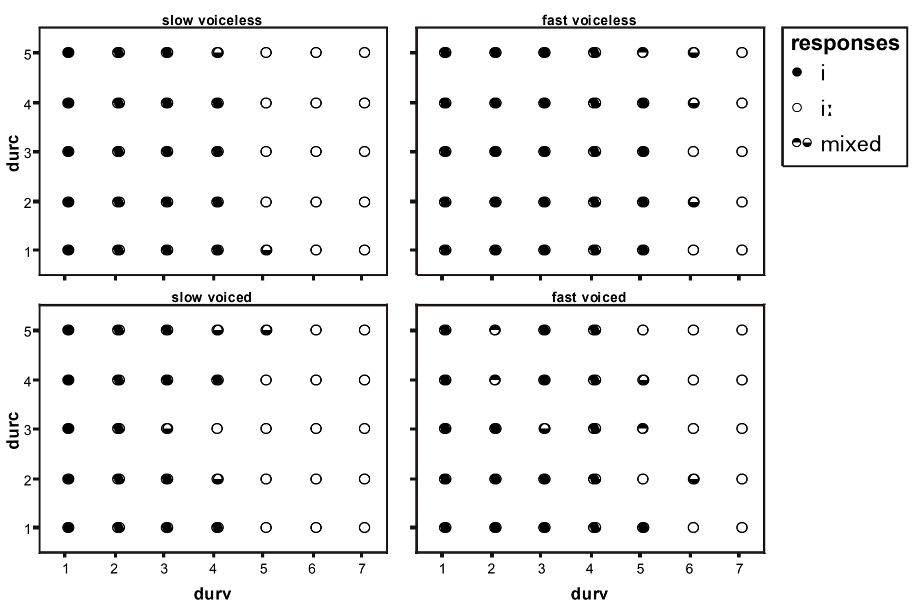
JP02

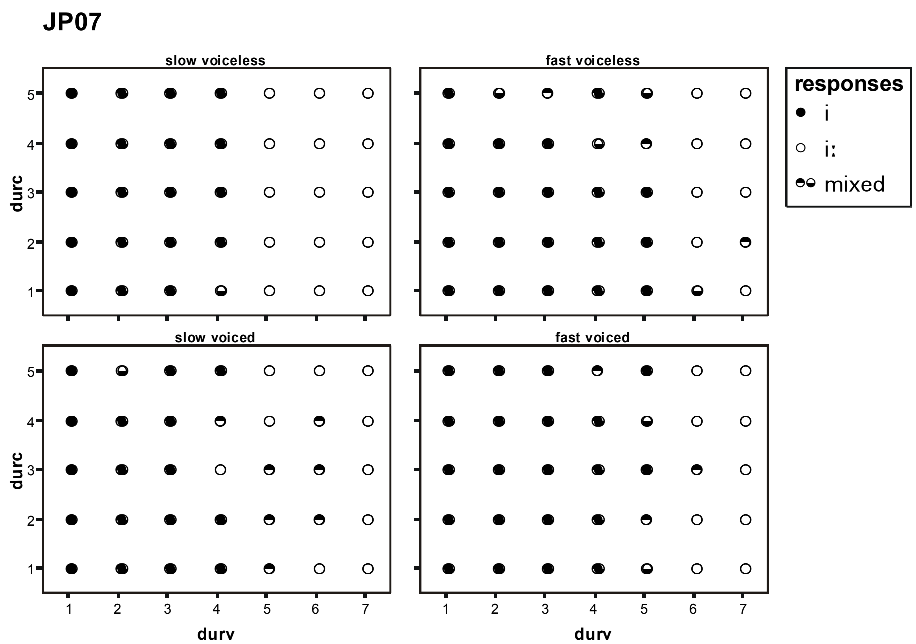


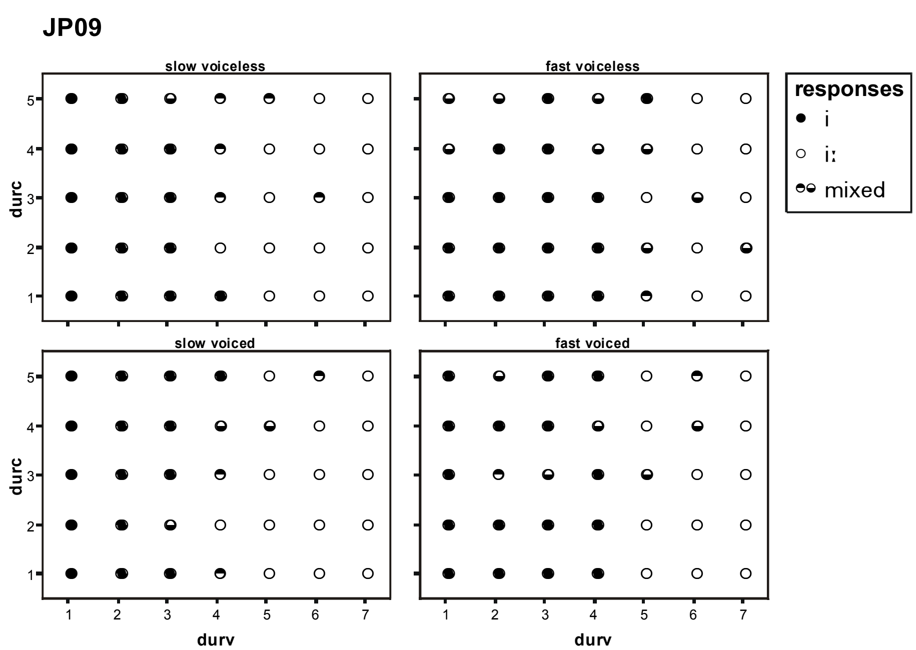
JP03

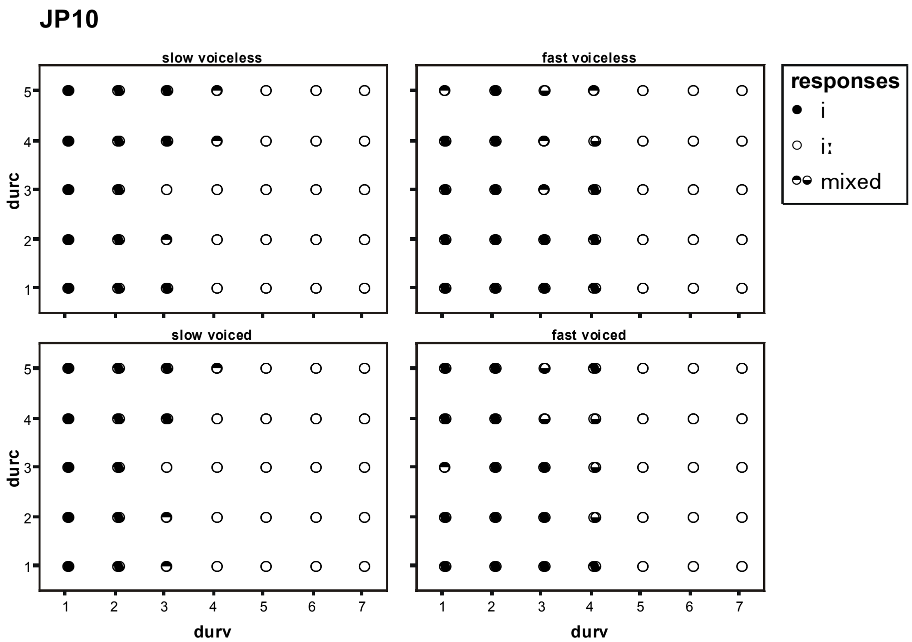

slow voiceless
fast voiceless
slow voiced
fast voiced
durc
durv
responses
i
iː
mixed

JP04

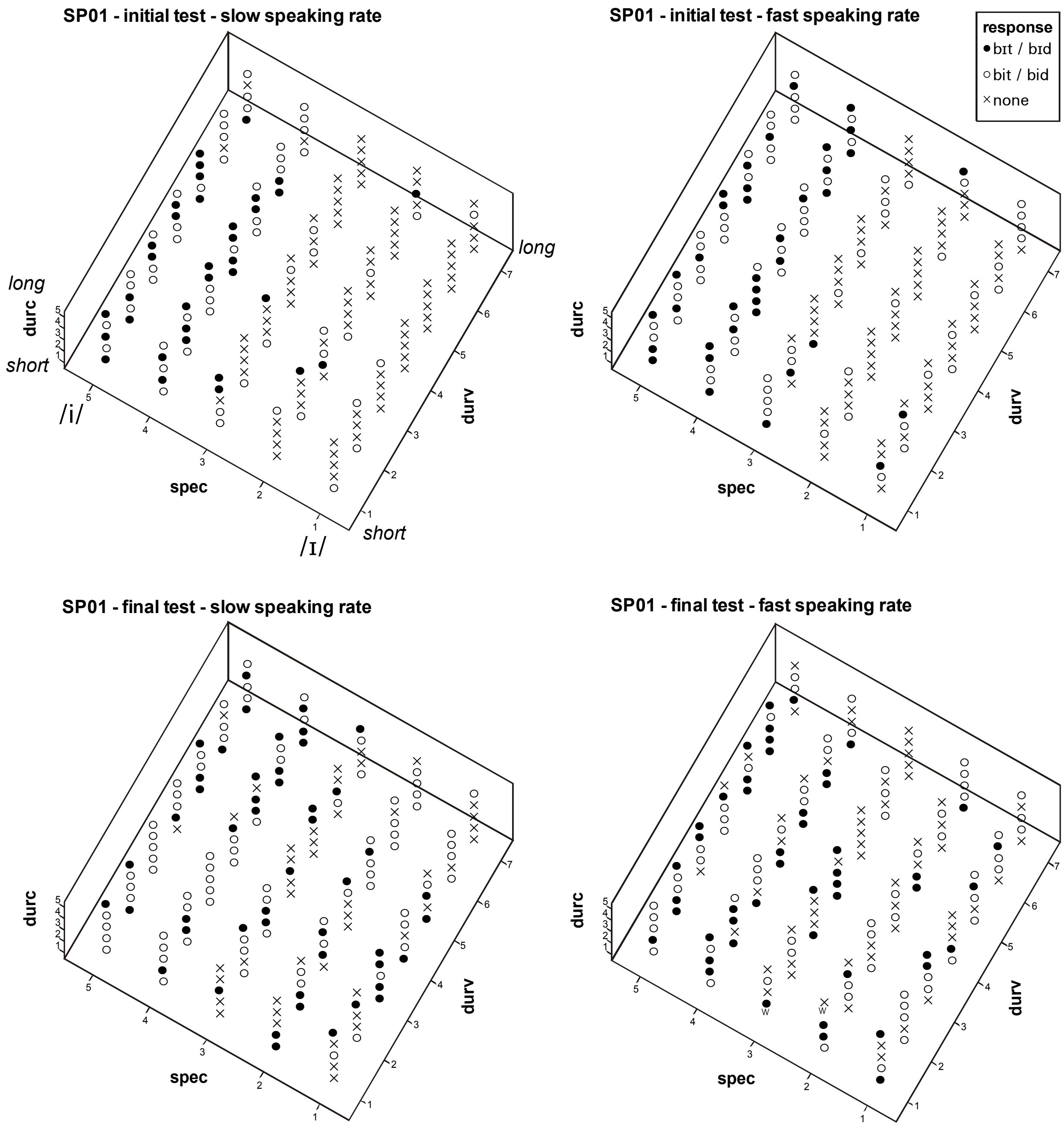

slow voiceless
fast voiceless
slow voiced
fast voiced
durc
durv
responses
i
iː
mixed


JP07
slow voiceless
fast voiceless
slow voiced
fast voiced
responses
i
iː
mixed
durc
durv

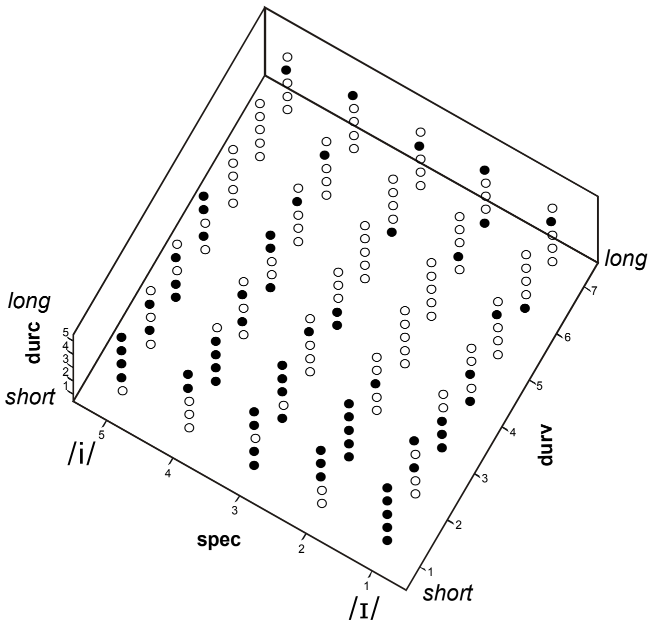

JP09
slow voiceless
fast voiceless
slow voiced
fast voiced
responses
i
iː
mixed
durc
durv


JP10
slow voiceless
fast voiceless
slow voiced
fast voiced
responses
i
iː
mixed
durc
durv
1
2
3
4
5
6
7

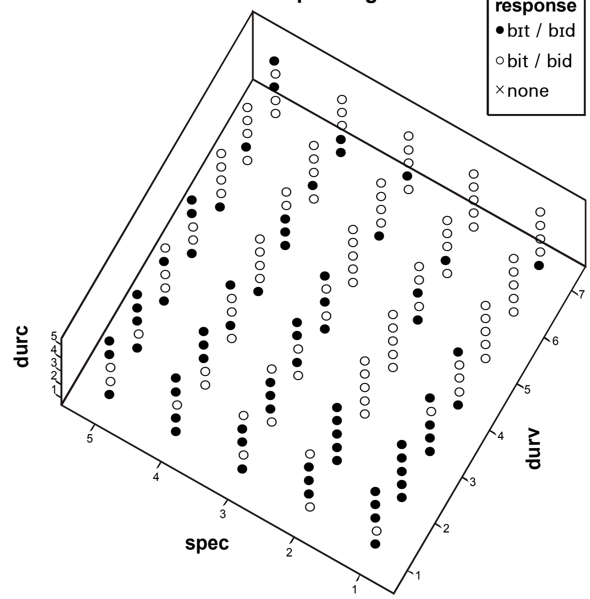
Mexican Spanish listeners' perception of English vowels



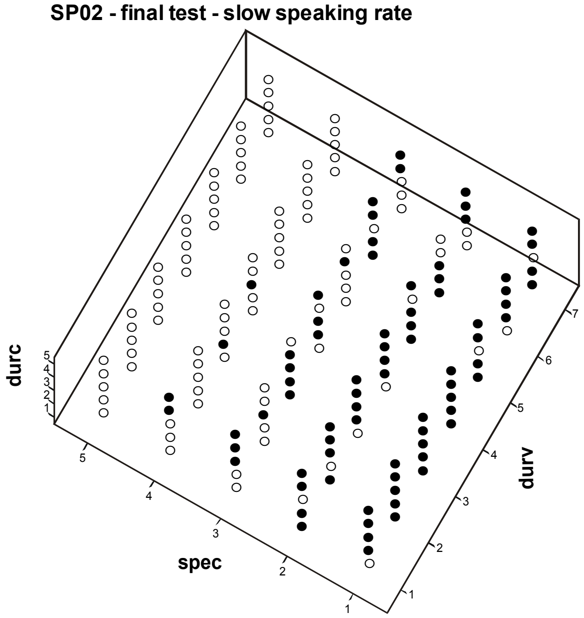
SP02 - initial test - slow speaking rate



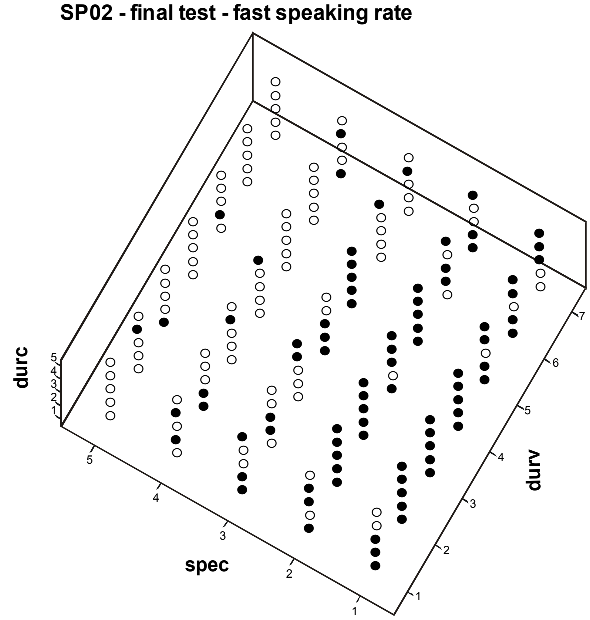
SP02 - initial test - fast speaking rate



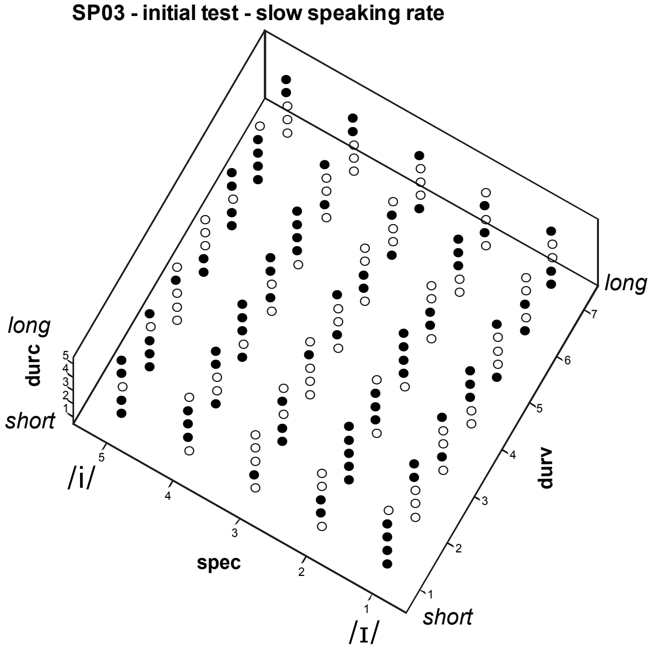
SP02 - final test - slow speaking rate



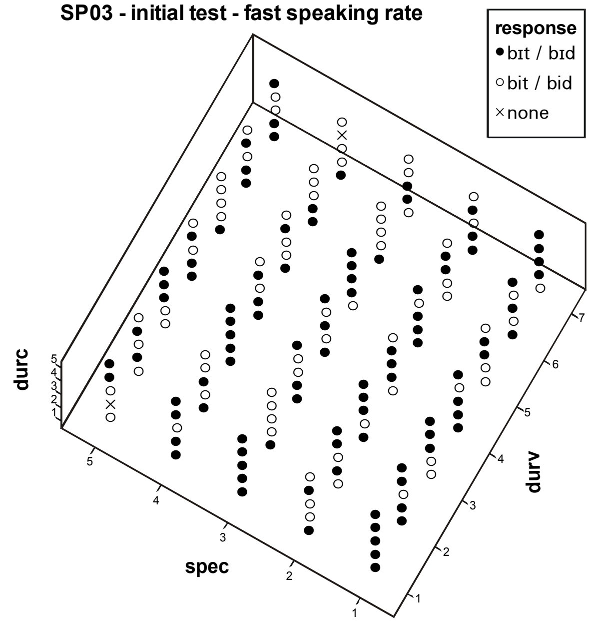
SP02 - final test - fast speaking rate



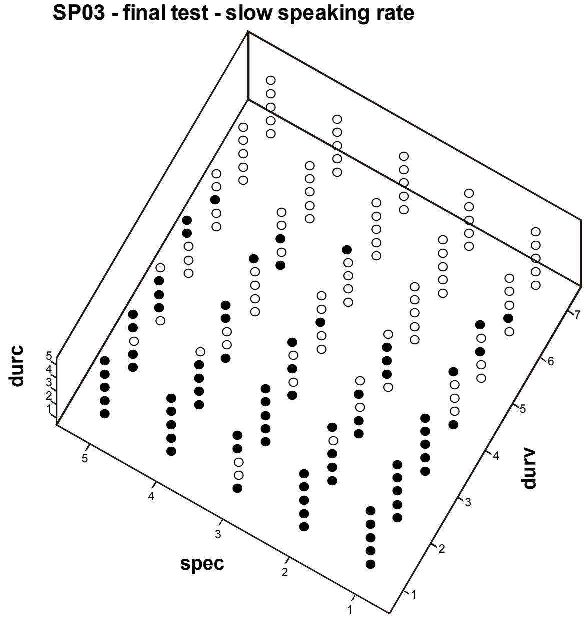
SP03 - initial test - slow speaking rate



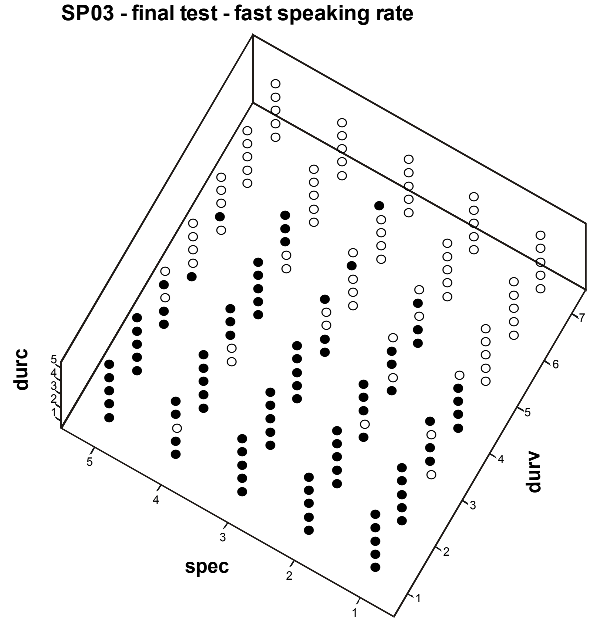
SP03 - initial test - fast speaking rate



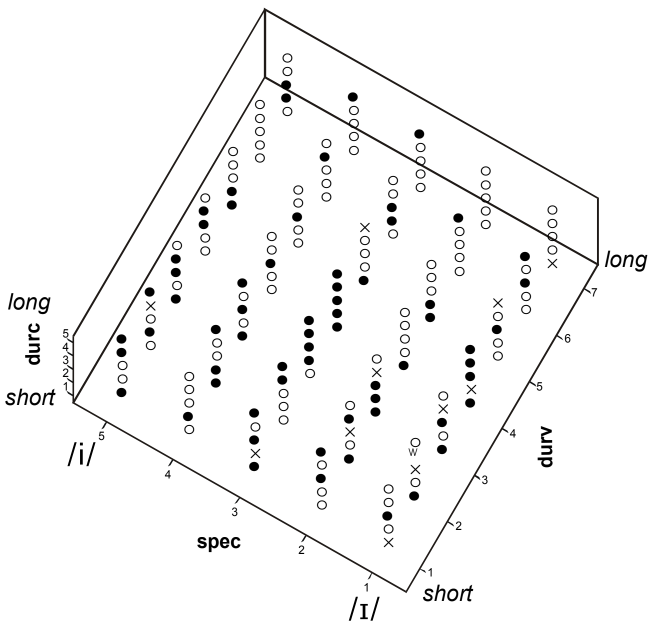
SP03 - final test - slow speaking rate



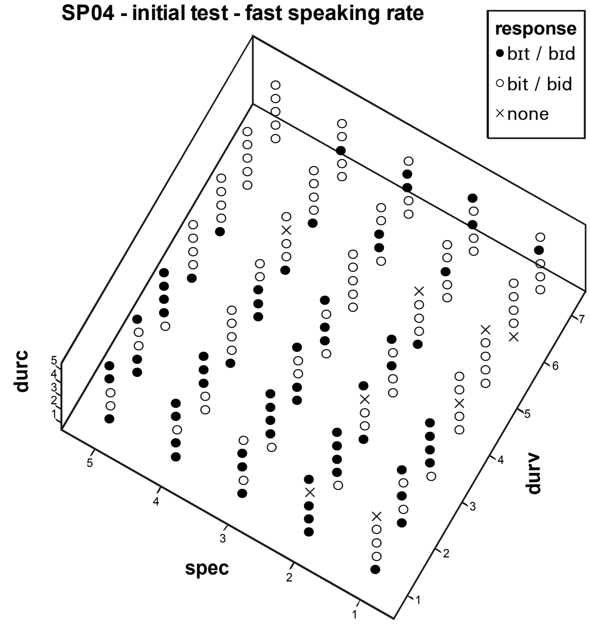
SP03 - final test - fast speaking rate



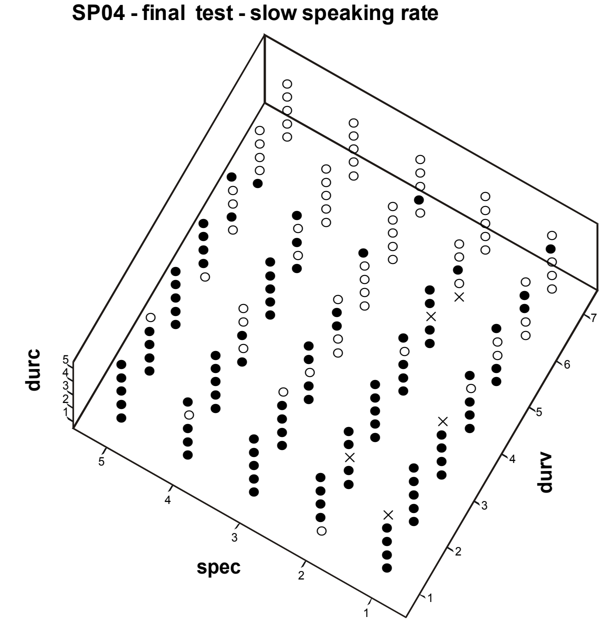
SP04 - initial test - slow speaking rate



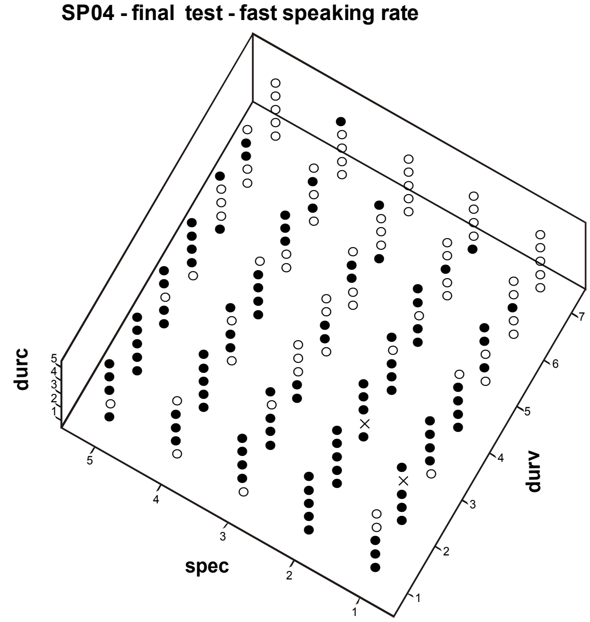
SP04 - initial test - fast speaking rate



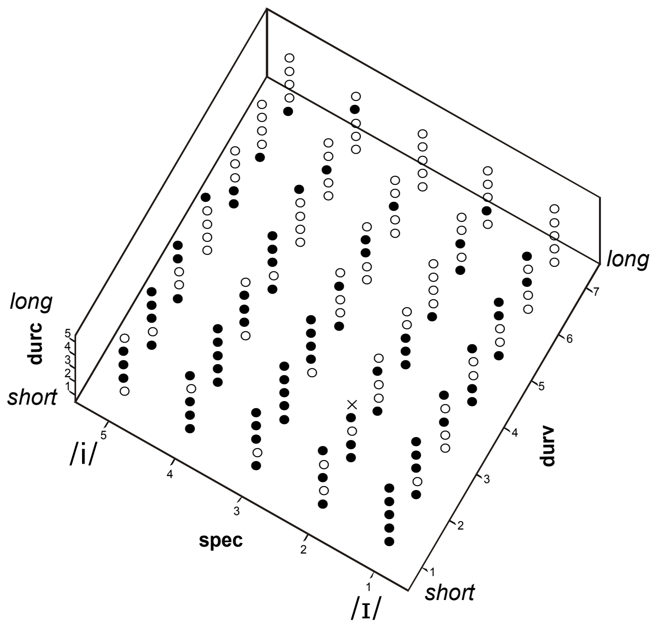
SP04 - final test - slow speaking rate



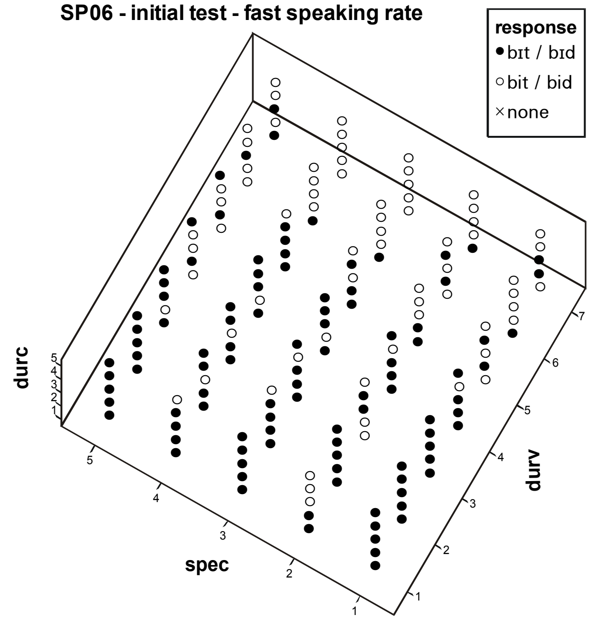
SP04 - final test - fast speaking rate



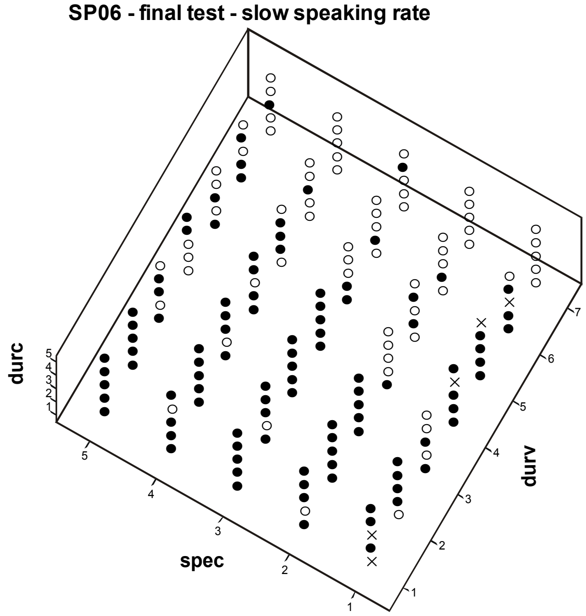
SP06 - initial test - slow speaking rate



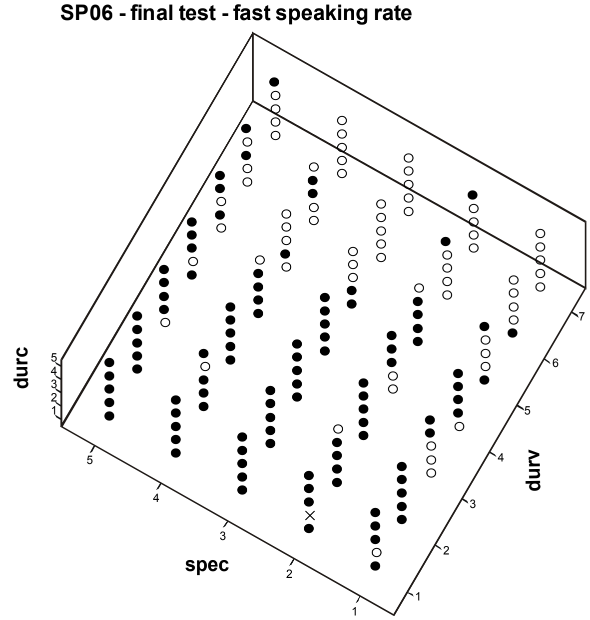
SP06 - initial test - fast speaking rate



SP06 - final test - slow speaking rate



SP06 - final test - fast speaking rate



References

Akamatsu, T. (1997). *Japanese phonetics: Theory and practice*. München & Newcastle: Lincom Europa.

Álvarez González, J. A. (1980). *Vocalismo español y vocalismo inglés* [Spanish and English vowels]. Doctoral thesis, Universidad Computense de Madrid.

Assman, P., Nearey, T., & Hogan, J. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America, 71*, 975-989.

Avis, W. S. (1975). The phonemic segments of an Edmonton idiolect. In J. K. Chambers (Ed.), *Canadian English: Origins and structures* (pp. 118-128). Toronto: Methuen.

Baba, R., Omuro, K., Miyazono, H., Usagawa, T., & Higuchi, M. (1996). The Perception of morae in long vowels: Comparison among Japanese, Korean and English speakers. *Proceedings of ICSLP 96,* Philadelphia, PA. 2474-2477.

Baroni, M, & Vanelli, L. (2000). The relationship between vowel length and consonantal voicing in Friulian. In L. Repetti (Ed.), *Phonological theory and the dialects of Italy* (pp. 13-44). Amsterdam, Philadelphia: John Benjamins.

Bennett, D. C. (1968). Spectral form and duration as cues in the recognition of English and German vowels. *Language and Speech, 11*, 65-85.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. Goodman, & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA: MIT Press.

Best, C. T. (1995a). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Baltimore: York Press.

Best, C. T. (1995b). Learning to perceive the sound pattern of English. In C. Rovee-Collier, & L. Lipsitt (Eds.), *Advances in infancy research* (pp. 217-304). Hillsdale, NJ: Ablex.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*, 305-330.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganisation for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 345-360.

Boersma, P., & Weenink, D. (2001). Praat (version 4.0) [Computer software]. Available: http//:www.praat.org

Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279-304). Baltimore: York Press.

Bohn, O.-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics, 11*, 303-328.

Bohn, O.-S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition, 14*, 131-58.

Brière, E. (1966). An investigation of phonological interference. *Language, 42*, 769-796.

Brown, M. T., & Wicker, L. R. (2000). Discriminant analysis. In H. Tinsley & S. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 209-235). San Diego: Academic Press.

Burnham, D. H. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics, 7*. 207-239.

Catford, J. C. (1965). *A linguistic theory of translation: An essay in applied linguistics*. London : Oxford University Press.

Chen, M. (1970). Vowel length variation as a function of the voicing of the following consonant environment. *Phonetica, 22*, 129-159.

Cochrane, G. R. (1970). Some vowel durations in Australian English. *Phonetica, 22*, 240-250.

Contreras Oller, C. M. (1997). Características acústicas de las vocales altas anteriores del inglés [Acoustic characteristics of high front vowels in English]. In E. Martínez Celdran (Ed.), *Estudios de fonética experimental, VIII* [Studies in experimental phonetics] (pp. 127-152). Barcelona: PPU.

Coren, S., Ward, L. M., & Enns, J. T. (1994). *Sensation and perception* (4th ed.). Forth Worth: Harcourt Brace

Crowther, C., & Mann, V. (1992). Native language factors affecting use of vocalic cues to final consonant voicing in English. *Journal of the Acoustical Society of America, 92*, 711-722.

Crystal, T. H., & House, A. S. (1982). Segmental duration in connected speech: A preliminary analysis. *Journal of the Acoustical Society of America, 72*, 705-716.

D'Introno, F., Del Teso, E., & Weston, R. (1995). *Fonética y fonología actual del español* [Current phonetics and phonology of Spanish]. Madrid: Cátedra.

Denes, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America, 27*, 761-764.

Enomoto, Kayoko. (1992). Interlanguage phonology: the perceptual development of durational contrasts by English-speaking learners of Japanese. *Edinburgh Working Papers in Linguisitics, 3*, 25-35. [ERIC: ED353 837]

Escudero, P. (2000): *Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tense/lax vowels in Spanish speakers*. Unpublished M.Sc. thesis, University of Edinburgh, Edinburgh, Scotland.

Escudero, P. (2001a). The perception of English vowel contrasts: acoustic cue reliance in the development of new contrasts. In J. Leather & A James (eds.), *Proceedings of the 4th International Symposium on the Acquisition of Second-Language Speech, New Sounds 2000, University of Amsterdam*.

Escudero P. (2001b). The role of the input in the development of L1 and L2 sound contrasts: language-specific cue weighting for vowels. In *Proceedings of the 25th Annual Boston University Conference on Language Development, November 2000*. Cascadilla.

Escudero, P., & Boersma, P. (2001). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In *Proceedings of the 25th Penn Linguistics Colloquium. Penn Working Papers in Linguistics*.

Fischer, R. M., & Ohde, R. N. (1990). Spectral and duration properties of front vowels as cues to final stop-consonant voicing. *Journal of the Acoustical Society of America, 88*, 1250-1259.

Fitzgerald, B. H. (1996). *Acoustic analysis of Japanese vowels produced in multiple consonantal contexts*. Unpublished Master's thesis, University of South Florida.

Flege, J. E. (1981). The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly, 15*, 443-455.

Flege, J. E. (1984a). The detection of French accent by American listeners. *Journal of the Acoustical Society of America, 76*, 692-707.

Flege, J. E. (1984b). The effect of linguistic experience on Arab's perception of the English /s/ vs. /z/ contrast. *Folia Linguistica, 18*, 117-138.

Flege, J. E. (1987). The production of "new" and "similar" phonemes in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47-65.

Flege, J. E. (1988). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders: A review* (pp 224-401). Norwood, NJ: Ablex Publishing.

Flege, J. E. (1989). Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training. *Journal of the Acoustical Society of America, 86*, 1684-1697.

Flege, J. E. (1991a). Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Heubner, & C. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theory*. (pp. 249-284). Philadelphia: John Benjamins.

Flege, J. E. (1991b). The interlingual identification of Spanish and English vowels: Orthographic evidence. *Quarterly Journal of Experimental Psychology*, 43, 701-731.

Flege, J. E. (1992). Speech learning in a second language. In C. Ferguson, L. Menn, & C. Stoel-Gammon (eds.), *Phonological development: Models, research, and application* (pp. 565-604). Timonium, MD: York Press.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. Journal of the *Acoustical Society of America, 93*, 1589-1608.

Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp 233-277). Baltimore: York Press.

Flege, J. E., & Hammond, R. (1982). Mimicry of non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition, 5*, 1-17.

Flege, J. E., & Hillenbrand, J. (1986). Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. *Journal of The Acoustical Society of America, 79*, 508-517.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*, 527-552.

Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech, 24*, 125-143.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*, 437-470.

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of The Acoustical Society of America, 95*, 3623-3641.

Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish. *Journal of The Acoustical Society of America, 92*, 128-143.

Flege, J. E., Schirru, C., & MacKay, I. R. A. (in press). Interaction between the native and second language phonetic subsystems. *Speech Communication*.

Fox, R. A., & Terbeek, D. (1977). Dental flaps, vowel duration and rule ordering in American English. *Journal of Phonetics, 5*, 27-34.

Fox, R. A., Flege, J. E., & Munro, M. J. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America, 97*, 2540-2551.

Godínez, M. Jr. (1978). A comparative study of some romance vowels. *UCLA Working Papers in Phonetics, 41*, 3-19.

Gottfried, T., & Beddor, P. (1988). Perception of temporal and spectral information in French vowels. *Language and Speech, 31*, 57-75.

Gregg, R. J. (1975). The phonology of Canadian English as spoken in the area of Vancouver, British Columbia. In J. K. Chambers (Ed.), *Canadian English: Origins and structures* (pp. 133-144). Toronto: Methuen.

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology, 25*, 577-588.

Guion, S. G., Flege, J. E., Akahane-Yamada R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America, 107*, 2711-2724.

Guion, S., Flege, J., Akahane-Yamada, R., & Pruitt, J. C. (2002). *Predicting Japanese adults' discrimination of English and Japanese vowels*. Manuscript in preparation.

Halle, M., Hughes, G., & Radley, J.-P. (1957). Acoustic properties of stop consonants. *Journal of the Acoustical Society of America, 29*, 107-116.

Hammond, R. M. (1986). Error analysis and the natural approach to teaching languages. *Lenguas Modernas* [Modern Languages]*, 13*, 129-139.

Han, M. S. (1994). Acoustic manifestation of mora timing in Japanese. *Journal of the Acoustical Society of America, 96*, 73-82.

Harrington, J., & Cassidy, S. (1999). *Techniques in speech acoustics*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., & Flege, J. E. (1984). Perception of the voiced-voiceless contrast in syllable-final stops. *Journal of the Acoustical Society of America, 76*, 18-26.

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America, 108*, 3013-3022.

Hogan, J. T., & Rozsypal, A. J. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *Journal of the Acoustical Society of America, 67*, 1764-1771.

Homma, Y. (1973). An acoustic study of Japanese vowels: Their quality, pitch, amplitude and duration. *Onsei no kenkyu - Study of sounds, 16*, 347-368.

Homma, Y. (1981). Durational relationship between Japanese stops and vowels. *Journal of Phonetics, 9*, 273-281.

House, A. S. (1961). On vowel duration in English. *Journal of the Acoustical Society of America, 33*, 1174-1178.

Ingram, J. C. L., & Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics, 25*, 343-370.

Imai, S., Flege, J. E., Wayland, R. (2002). Perception of cross-language vowel differences: A longitudinal study of native Spanish learners of English. *Journal of the Acoustical Society of America, 111*, 2364.

Itô, J., & Mester, R. A. (1995). Japanese Phonology. In J. Goldsmith (ed.), *Handbook of phonological theory* (pp. 817-838). Cambridge, MA and Oxford, UK: Blackwell.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using multidimensional scaling. *Journal of the Acoustical Society of America, 97*, 553-562.

Johnston, D. (1999). Cool Edit Pro LE (version 1.2) [Computer software]. Phoenix: Syntrillium.

Keating, P. (1985). Universal phonetics and the organization of grammars. In V. A. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 115-132). New York: Academic Press

Kendall, R. A. (2001). Music experiment development system: MEDS (version 2001-B) [Computer software]. Los Angeles. [On-line]. Available: http://www.ethnomusic.ucla.edu/systematic/Faculty/Kendall/meds.htm

Kewley-Port, D., Akahane-Yamada, R., & Aikawa, K. (1996). Intelligibility and acoustic correlates of Japanese accented English vowels. *Proceedings of ICSLP 96,* Philadelphia, PA. 450-453.

Kingston, J. (in press). Keeping and losing contrasts. *Proceedings of the 28th annual meeting of the Berkeley Linguistics Society*.

Kluender, K. R., Diel, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics, 16*, 153-169.

Kozasa, T. (2002). Duration and F0 cues for vowel length in Japanese. *Journal of the Acoustical Society of America, 111*, 2365.

Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics, 50*, 93-107.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the “perceptual magnet effect”. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp 121-154). Baltimore: York Press.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606-608.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinctions in English. *Language, 33*, 42-49.

Lotto, A. J. (2002). Animal models of speech sound categorization. *Journal of the Acoustical Society of America, 111*, 2384.

Luce, P. A., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of the Acoustical Society of America, 78*, 1949-1957.

Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *Journal of the Acoustical Society of America, 71*, 173-178.

Marín Gálvez, R. (1995). La duración vocálica [Vowel duration]. *Estudios de Lingüística Universidad de Alicante, 10*, 213-226.

Martínez Celdrán, E. (1994). *Fonética* [Phonetics] (4th ed.). Barcelona: Editorial Teide.

McAllister, R., Flege, J. E., & MacKay, I. R. A. (1999). The acquistition of Swedish long vs short vowel contrasts by native speakers of English, Spanish, and Estonian. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 751-754). Department of Linguistics, University of California at Berkeley.

Minnick Fox, M., & Maeda, K. (1999). Categorization of American English vowels by Japanese speakers. *Proceedings of the International Conference of Phonetic Sciences, San Francisco* (pp. 1437-1140).

Mitleb, F. (1984). Voicing effect on vowel duration is not an absolute universal. *Journal of Phonetics, 12*, 23-27.

Morrison, G. S. (2002a). Perception of English /i/ and /ɪ/ by Japanese listeners. *University of British Columbia Working Papers in Linguistics: Vol. 8.* Vancouver, BC: University of British Columbia, Department of Linguistics.

Morrison, G. S. (2002b). *An acoustic and statistical study of Spanish vowel allophones*. Manuscript in preparation.

Morrison, G. S. (2002c). *Cross-dialect and cross-language perception of English vowels: Scottish, Northern English, Canadian, Japanese, and Spanish listeners*. Manuscript in preparation.

Munro, M. J. (1993). Production of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech, 36*, 39-66.

Munro, M. J., Wang, X., & Li, C. H. (2000). Asymmetries in English high vowel production by Cantonese speakers: Findings and methodological issues. Manuscript submitted for publication.

Nakai, S. (1997). Perceptual factors in the acquisition of FL vowels: The locations of vowel prototypes. *Proceedings of the 1997 Edinburgh Linguistics Department Postgraduate Conference.* [On-line]. Available: http://www.ling.ed.ac.uk/~pgc

Nakai, S. (1999). *The perceptual distance between vowels: Effects of prototypicality and extremity*. Unpublished doctoral dissertation, Edinburgh University.

Navarro Tomás, T. (1916). Cantidad de las vocales accentuadas [The quantity of stressed vowels]. *Revista de Filología Española, 3*, 387-408.

Navarro Tomás, T. (1965). *Manual de pronunciación española* [Spanish pronunciation manual] (12th ed. [1st ed. published 1918]). Madrid: Consejo Superior de Investigaciones Científicas.

Nearey, T. M., & Assmann, P. F. (1986). Modelling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America, 80*, 1297-1308.

Nozawa, T., & Flege, J. E. (2000). Nihonjin eigo gakushyuusshya no eigo no boin chikaku [Perception of English vowels by Japanese learners of English]. *Kotoba no kagaku kenkyuu - Journal of the Japan Society for Speech Sciences, 1,* 17-27.

Parker, F. (1974). The coarticulation of vowels and stop consonant. *Journal of Phonetics, 2*, 211-221.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America, 32*, 693-703.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America, 89*, 2961-2977.

Polka, L. (1992). Characterizing the influence of native experience on adult speech perception. *Perception and Psychophysics, 52*, 37-52.

Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America, 97*, 1286-1296.

Port, R. (1981). Linguistics timing factors in combination. *Journal of the Acoustical Society of America, 69*, 262-274.

Port, R., & Mitleb, F. (1983). Segmental features and implementation in acquisition of English by Arabic speakers. *Journal of Phonetics, 11*, 219-229.

Port, R. F., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception and Psychophysics, 32*, 141-152.

Port. R., Al-Ani, S., & Maeda, S. (1980). Temporal compensation and universal phonetics. *Phonetica 37*, 235-252.

Quilis, A., & Esgueva, M. (1983). Realización de los fonemas vocálicos españoles en posición fonética normal [Realization of Spanish vowel phonemes in normal phonetic position]. In *Estudios de fonética I, Colectanea Phonetica VII* (pp. 159-252). Madrid: Consejo Superior de Investigaciones Científicas.

Raphael, J. L. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America, 51*, 1296-1303.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and Language: Advances in Basic Research and Practice: Vol. 10* (pp. 243-335). New York: Academic Press.

Repp, B. H., & Williams, D. R. (1985). Influence of following context on perception of the voiced-voiceless distinction in syllable-final consonants. *Journal of the Acoustical Society of America, 78*, 445-457.

Revoile, S., Picket, J. M., Holden, L. D., & Talkin, D. (1982). Acoustic cues to final stop voicing for impaired- and normal-hearing listeners. *Journal of the Acoustical Society of America, 72*, 1145-1154.

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Baltimore: York Press.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics, 10*, 209-231.

Shibatani, M. (1990). *The languages of Japan*. Cambridge, UK: Cambridge University Press

SPSS Inc. (1999). SPSS base 10.0 applications guide [Computer software manual]. Chicago: Author.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Strange, W., Akahane-Yamada, R., Fitzgerald, B.H., & Kubo R. (1996). Perceptual Assimilation of American English Vowels by Japanese Listeners. *Proceedings of ICSLP 96,* Philadelphia, PA. 2458-2461.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., & Jenkins, J. J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics, 26*, 311-344.

Suen, C., & Beddoes, M. (1974). The silent interval of stop consonants. *Language and Speech, 17*, 126-134.

Summers, W. V. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *Journal of the Acoustical Society of America, 82*, 847-863

Summers, W. V. (1988). F1 structure provides information for final-consonant voicing. *Journal of the Acoustical Society of America, 84*, 485-492.

Takahashi, Y. (1987). *Language universals and transfer effects: Acquisition of English vowel durations by Japanese learners*. Unpublished doctoral dissertation, Stanford University. [UMI: AAT 8801045]

Toda, T. K. (1999). Development of speech discrimination by learners of Japanese as a second language: Remarks from a longitudinal study. In P. Robinson (Ed.), *Representation and Process: Proceedings of the 3rd Pacific Second Language Research Forum: Vol. 1* (pp 207-233). Tokyo: Pacific Second Language Research Forum.

Tsukada, K. (1996). Acoustic analysis of Japanese-accented vowels in English. In P. McCormick & A. Russell (Eds.), *Proceedings of the 6th Australian International Conference on Speech Science and Technology* (pp. 373-378). Canberra: Australian Speech Science and Technology Association.

Tsukada, K. (1998). Japanese-accented English vowels: A perception study. *Asia Pacific Journal of Speech, Language and Hearing, 3*, 43-65.

Tsukada, K. (1999). *An acoustic phonetic analysis of Japanese-accented English*. Unpublished Doctoral dissertation, Macquarie University, Sydney, Australia.

Tsukada, K. (2001) Native vs non-native production of English vowels in spontaneous speech: An acoustic phonetic study. In *Proceedings of Eurospeech 2001* (pp. 305-308). Aalborg, Denmark.

Vance, T. (1987). *An introduction to Japanese phonology*. Albany, NY: State University of New York Press.

Vaquero de Ramírez, M. (1996). El español de América I: Pronunciación [American Spanish I: Pronunciation]. Madrid: Arco Libros.

Walker, D. C. (1975). Another idiolect: Comments on an aricle by Professor Avis. In J. K. Chambers (Ed.), *Canadian English: Origins and structures* (pp. 129-132). Toronto: Methuen.

Walsh, T., & Parker, F. (1983). Vowel length and vowel transition cues to [±voice] in post-vocalic stops. *Journal of Phonetics, 11*, 407-412.

Walsh, T., Parker, F., & Miller, C. J. (1987). The contribution of F1 decline to the perception of [±voice]. *Journal of Phonetics, 15*, 101-103.

Wang, X. (1997). *The acquisition of English Vowels by Mandarin ESL learners: A study of production and perception*. Unpublished master's thesis, Simon Fraser University, Burnaby, British Columbia.

Wardrip-Fruin, C. (1982). On the status of temporal cues to phonetic categories: Preceding vowel duration as a cue to voicing in final consonants. *Journal of the Acoustical Society of America, 71*, 187-195.

Weiss, R. (1976). *The perception of vowel length and quality in German: An experimental-phonetic investigation*. Hamburg: Buske.

Wells, J. C. (1982). *Accents of English*. Cambridge, UK: Cambridge University Press.

Wolf, C. (1978). Voicing cues in English final stops. *Journal of Phonetics, 6*, 299-309.

Yamada, T., Akahane-Yamada, R., & Strange, W. (1995). Perception of English vowels and consonants by Japanese learners of English. [in Japanese] *Proceedings of the Acoustical Society of Japan*, 379-380.

Zimmerman, S. A., & Sapon, S. M. (1958). Note on vowel duration seen cross-linguistically. *Journal of the Acoustical Society of America, 30*, 152-153.