（野澤　健）

Effects of the Manner of Articulation of the Syllable-Final Consonant on the Perception of American English Vowels by Native Japanese Speakers:

Divergence Between Japanese Speakers' Image of English Vowels and what English Vowels Really Sound like to them

A Thesis Submitted for the Degree of Doctor of Philosophy,

Studies in Language and Culture,

Graduate School of Language and Culture,

Osaka University

by

NOZAWA, Takeshi

（野澤　健）

November　2018

Table of Contents

Acknowledgment

Firstly, I would like to express my sincere gratitude to my chief advisor Prof. Shiro Kori for the continuous support of my Ph.D study and related research through the past three years. He has always been my guiding light, and has offered me indispensable advice.

Besides my chief advisor, I would like to thank the rest of my dissertation committee: Prof. Takeshi Yamamoto, and Prof. Nobuyuki Hino, for their insightful comments and encouragement.

I thank members of Phonetic Linguistics Project of Graduate School of Language and Culture at Osaka University for sharing their insight.

I also thank my colleagues at Ritsumeikan University for allowing me to concentrate on my Ph.D. study.

Last but not the least, I would like to thank my wife Yuriko. Without her support, I would have given up continuing my Ph.D. study.

Abstract

Researchers generally agree that the perception of non-native phones is strongly affected by individuals' L1 phonology, and previous studies have demonstrated that L1 and non-native phones are related at allophonic level rather than phonemic level.

Syllable-final /n/ and /l/ exercise a strong influence on the quality of the preceding vowel in American English, but little has been known as to how non-native speakers perceive American English vowels in these contexts.

This dissertation attempts to address this issue. Native Japanese speakers performed four tasks: (1) perceptual assimilation task, (2) identification task, (3) discrimination task, and (4) production task. Six American English vowels /i, ɪ, ɛ, æ, ɑ, ʌ/ are chosen. In perceptual assimilation task, Japanese speakers identified American English vowels in terms of Japanese vowel categories. In identification task, they identified American English vowels they heard. In discrimination task, they discriminated six vowel pairs in AXB format. Japanese speakers produced American English vowels in two different conditions. First, they read aloud words on a word list, and then they repeated after native speakers' utterances.

The results revealed that in general Japanese speakers' identification and discrimination accuracy is lower before /n/ and /l/. This is not simply because vowels sound alike in these contexts, but also because English vowels become more distant from the auditory image Japanese speakers generally hold. The image has been formed by loanwords from English. For instance, /æ/ is commonly transcribed as *a* in Japanese, and so Japanese speakers expect the vowel to sound like *a*. In prenasal context, however, /æ/ is raised and fronted, and so it is more distant from Japanese *a* than in preplosive context. As a result, /æ/ is less accurately identified in prenasal context. Likewise, /i/ and /ɪ/ are differentiated by length as *ii* and *i*. Japanese speakers expect /i/ to sound like long *ii*. But before /l/ vowels are retracted and /i/ does not stay in high front position, and as a result /i/ is less accurately identified before /l/. Token-based analysis revealed that English vowels are more accurately identified when they are perceptually assimilated in Japanese vowel categories as commonly transcribed in Japanese orthography.

Also found is individual difference among Japanese speakers. Some of them qualitatively differentiate /i/ /and /ɪ/ in production while others differentiate these two vowels only by duration. But in perception, even those who differentiate the two vowels qualitatively in production are not sensitive to spectral differences between the two vowels.

List of Figures

List of Tables

1. Introduction

1.1 Background of this study

A great many studies have shown that individuals' L1 phonology exercises a strong influence on how they perceive and produce nonnative phones. For example, native speakers of Japanese are well known for their inaccurate perception of English /r/ and /l/. Speakers of Japanese are said to perceive English /r/ and /l/ as instances of the Japanese liquid /ɾ/. English words "right" and "light" are both transcribed as "raito" (/ɾaito/). However, /r/ and /l/ are not equally challenging in all positions to Japanese speakers. Ingram and Park (1998) demonstrated that, for native Japanese speakers, /r/ and /l/ in word-initial consonant clusters (as in "grow" and "glow") are the most difficult to identify. The reasons these particular difficulties with /r/ and /l/ are not completely clear, but one possible explanation may be that /r/ and /l/ are shorter in this position than in both word-initial position (as in "row" and "low") and Figuren (as in "arrive" and "alive"). Another possibility is that, because Japanese syllable structure is simple and consonant clusters are rare, Japanese speakers in general are not adept at correctly perceiving consonant clusters. Whatever accounts for Japanese speakers' low identification accuracy of English /r/ and /l/ in consonant clusters, learning another language requires learning to identify and discriminate phones that are not phonemically contrastive in one's L1, and it also requires learning to identify and discriminate nonnative phones in environments that are not possible in one's L1.

This study examines native Japanese speakers' perceptions of American English vowels. Variations across individuals and regions occur more in vowels than in consonants, but to perceive a vowel categorically requires labeling different tokens of a vowel as instances of the same vowel regardless of talker-specific characteristics. In other words, to discriminate "beat" and "bit" categorically, perceiving /i/ and /ɪ/ as different vowels is not enough, but rather it necessitates constantly labeling /i/ tokens as /i/ even when these tokens are physically and phonetically different. Therefore, an experiment needs to be designed so that a participant must give a response based on a categorical difference of stimuli. To assess the ability to constantly label different tokens of the same vowel as the vowel, multiple talkers' tokens are employed. If a particular token is more (or less) accurate than the other tokens of the same vowel, then how each token of that vowel is processed in terms of participants' L1 vowel categories will be investigated.

Non-native speakers with limited exposure to the target language are unlikely to have established robust vowel categories of that target language. When speakers do not have established vowel categories of their target language, they are likely to rely on categorical distinctions of their native language. This study recruits native speakers of Japanese with different proficiency levels of English, in part because it was difficult to control the language background of participants, but largely because it was believed that participants all have common features or tendencies shared by native speakers of Japanese regardless of their proficiency in English. Most of them have learned English as a school subject and so have had limited exposure to English outside of the school environment. However, among these native Japanese speakers are those who have lived in English-speaking countries, and one of them was born abroad. Therefore, the manner and amount that they use their native Japanese vowel categories to identify and discriminate English vowels may vary. Those with more linguistic experience may be more sensitive to phonetic differences between English vowels even if they have not established robust English vowel categories. If this is the case, such Japanese speakers may rely less on Japanese vowel categories to identify and discriminate English vowels than do less experienced learners of English. Moreover, all the participants may not perceive the same English vowel as an instance of the same Japanese vowel category, and so for some participants discriminating one particular English vowel pair may be within-category discrimination, while, for others, it may be a case of between-category discrimination. This dissertation also attempts to examine individual differences from another angle. American English vowels uttered by multiple talkers are used as stimuli to see whether participants can correctly identify and discriminate American English vowels regardless of talker differences, rather than rely on talker- or token-specific features. If one talker's token is more or less correctly identified than those of other talkers, phonetic features that make an easy vowel token for Japanese listeners to perceive is different from those that are difficult to perceive. Individual differences will be dealt with in detail in Chapter 8.

Phonetic features of English vowels are thought to be affected by the surrounding sounds. Hillenbrand & Clark (2001) measured formant frequencies of eight American English monophthongs uttered by native speakers in /CVC/ context, where the initial consonants are /h,b,d,g,p,t,k/ and the final consonants are /b,d,g,p,t,k/, and the researchers examined the effects of consonantal context on the vowel quality. Hillenbrand & Clark (2001) found that the place of articulation for the initial consonant exercises a stronger influence on the formant frequencies than does that of the final consonant. However, the manner of articulation of the following consonant affects the vowel quality to the extent that the allophone of the vowel is recognized. For instance, English vowels are said to be nasalized before a nasal consonant, and to be retracted before /l/. Even though these postvocalic consonants exercise a strong influence on the quality of the vowel, little has been explored as to the effects of the manner of articulation of the coda consonant on non-native speakers' perceptions of English vowels. On the contrary, to assess non-native speakers' sensitivity in identifying and discriminating English vowels, most experiments have been designed to minimize the effects of the surrounding consonants. This dissertation attempts to address this issue by comparing native Japanese speakers' perception of American English stops before /t/, /n/, and /l/.

The official school curriculum guidelines developed by the Ministry of Education, Sports, Culture, Science and Technology states that "phonetic notation may be employed to supplement phonetic instruction[1]," which means that learning phonetic symbols is not mandatory. Teachers do not proactively teach their students English pronunciation. Teshima (2011) raises four reasons why Japanese students widely use katakana pronunciation, which is the pronunciation of English phones as if English texts were written in katakana orthography. Such cases use Japanese phones, which are not always close to the original English phones, and some of the phonemic contrasts are lost.

[1] http://www.mext.go.jp/a_menu/shotou/new-cs/youryou/eiyaku/__icsFiles/afieldfile/2011/04/11/1298353_9.pdf

Syllable structures and stress patterns change to comply with phonological rules of Japanese. According to Teshima (2011), (1) the number of hours English taught is insufficient, (2) teachers have no choice but to be satisfied when students merely say something in English (correcting their pronunciation can intimidate them, leading them to become reluctant to speak English), (3) there is no strong necessity to teach pronunciation, and (4) teachers are unsure as to how they should teach pronunciation. A survey conducted by Shibata, Yokoyama and Tara (2006) supports Teshima's claim. The survey says that, in general, teachers are aware of the importance of teaching pronunciation, but that their confidence rating in teaching pronunciation is not as high, and, moreover, teachers' rating of whether they practice teaching pronunciation is even lower. Two surveys conducted by Ota (Ota 2012, 2013) report that about 80% of university students responded that they had had little or no instruction in English pronunciation in middle school and high school (See also Wada, 2015). More than four decades ago, Quackenbush (1974) commented that "most Japanese seldom or never hear spoken English; they do not attempt to pronounce English words, and they do not borrow English words. They simply use words of English origin that are borrowed for them by others, mainly writers. They pronounce them the way they hear them pronounced on radio and television, and they spell them the way they see them spelled in the popular press, that is, as fully assimilated Japanese words with a minimum of departures from the sounds and sound sequences and spelling principles that characterize native Japanese words" (p. 64).

In an EFL context, where learners have limited exposure to target model English pronunciation, learners are more likely to rely on their L1 phone categories or transcriptions of English phones in their L1. This is where the significance of the above-mentioned katakana pronunciation or katakana English comes into focus. Japanese has a large number of loanwords from English. While loanwords from other languages are overwhelmingly nouns and verbs, those from English extend to adjectives, adverbs, interjections, prepositions, numbers, pronouns, prefixes, articles, and conjunctions, and because they permeate the everyday life of Japanese people, some of the first words children acquire are loanwords from English (Irwin 2011: 58). A study of children's

English vocabulary knowledge by Kasahara, Machida, Osada, Takahashi and Yoshizawa (2012), reveals that Japanese children are more familiar with katakana English, i.e. loanwords from English transcribed in katakana, than they are with English words learned through English study and practice. Loanwords can facilitate learning English vocabulary, but because the pronunciation of loanwords is different from that of their source words, they can hinder the acquisition of real authentic English pronunciation. This dissertation argues that what Japanese speakers believe to be the sound of English vowels actually is strongly affected by the early adaptations of English vowels. In other words, how English words are transcribed in Japanese orthography helps form the "image" of English vowels, and Japanese speakers expect English vowels to sound like the "image" they hold; therefore, the more distant a token of an English vowel is from this "image," the less accurately the vowel is identified. Japanese adaptation of English vowels is discussed in detail in Section 1.5.

To sum up, this dissertation attempts to investigate three variables: (1) the effects of talker differences, (2) the effects of postvocalic consonants, and (3) individual differences among Japanese participants. By taking these three factors into account, this dissertation attempts to show that what Japanese speakers expect English vowels to sound like, that is, how their "image" of English vowels strongly affects their perception of English vowels, and that their image or expectation is largely affected by the transcription of loanwords from English in Japanese orthography.

1.2 Description of vowels

In articulatory phonetics, vowels can be defined as "sounds articulated without a complete closure in a mouth or a degree of narrowing which would produce audible friction" (Crystal, 1997:415). Thus, by definition, vowels have no place of articulation. Instead, vowels can be roughly classified by (1) the position of the lips, and (2) the part of the tongue raised, and the height to which it moves (ibid.).

The International Phonetic Association created a vowel chart to describe vowels of languages. By referring to this chart, one can see how one vowel is different from another. For instance, both /i/ and /u/ are close vowels, i.e. the tongue is raised so the oral cavity becomes narrow, but different parts of the tongue are raised. When /i/ is produced, the front part of the tongue is raised, whereas, when /u/ is produced, the back of the tongue is raised. Both /u/ and /ɒ/ are back vowels, but the height of the tongue is different. Both /i/ and /y/ are close (=high) front vowels, but the lips are rounded when producing /y/, but not rounded when producing /i/. Thus, vowels are differentiated horizontally, vertically, and by the shape of the lips.

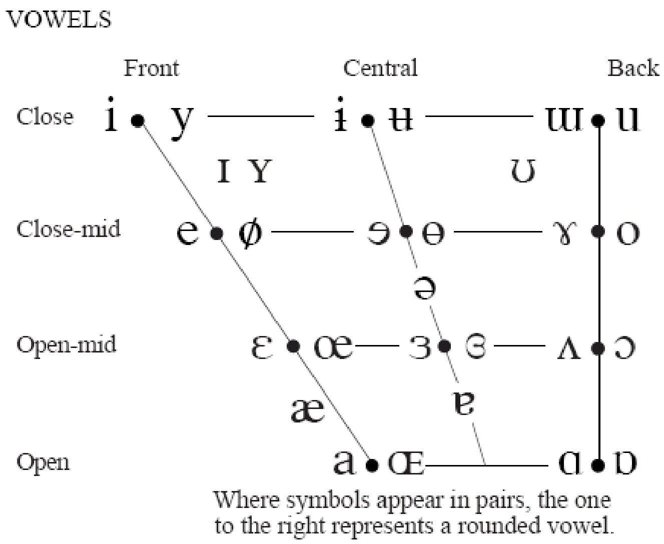

Figure 1. IPA vowel chart (retrieved from the website of the International Phonetic Association)

1.3 American English vowel system

Figure 1 is not language-specific, so it contains more vowels than necessary to describe vowels of one language. But to describe the vowel system of a language, other information that is not mentioned in Figure 1 is necessary. For instance, a language like Japanese uses durational information to differentiate lexical meaning. Figure 1 contains no diphthongs like [aɪ] as in *bite*, or [ɔɪ] as in *toy*. Therefore, to describe the vowel system of English, the English phonological system has to be taken into account.

English full vowels (as opposed to reduced vowels) are roughly classified into three categories: (1) short or lax vowels, (2) long or tense vowels, and (3) diphthongs. Since both /i/ (as in *beat*) and /ɪ/ (as in *bit*) are close (=high) front vowels, what makes them different from each other? *Cambridge English Pronouncing Dictionary* (=*CEPD*) *18th Edition* classifies British English vowels into short vowels, long vowels, and diphthongs as shown in Table 1.

Table 1

Classification of British English vowels according to Cambridge English Pronouncing Dictionary

Short vowels	/ɪ/	/e/	/æ/	/ʌ/	/ɒ/	/ʊ/		
	kit	dress	trap	strut	lot	foot		
Long vowels	/iː/	/ɑː/	/ɔː/	/uː/	/ɜː/			
	fleece	palm	thought	goose	nurse			
Diphthongs	/eɪ/	/aɪ/	/ɔɪ/	/əʊ/	/aʊ/	/ɪə/	/eə/	/ʊə/
	face	price	choice	goat	mouth	near	square	cure

But as for American English, *CEPD* says that "in American English we do not find the difference between long and short vowels described above, and the vowel system is commonly described as having lax vowels, tense vowels, and diphthongs" (viii). Lax vowels correspond to British English short vowels. American English vowels listed in *CEPD* are shown in Table 2.

Table 2

Classification of American English vowels according to Cambridge English Pronouncing Dictionary

Lax vowels	/ɪ/	/e/	/æ/	/ʌ/	/ʊ/
	kit	dress	trap	strut	foot
Tense vowels	/iː/	/ɑː/	/ɔː/	/uː/	/ɝː/
	fleece	palm	thought	goose	nurse
Diphthongs	/eɪ/	/aɪ/	/ɔɪ/	/oʊ/	/aʊ/
	face	price	choice	goat	mouth

CEPD also notes that, in American English, there is no centering diphthong like /ɪə/, and that these are lax vowels followed by /r/ like /ɪr/, so there are fewer vowels in American English.

Cruttenden (2014) uses features long/short to describe British English vowels. He refers to American English vowels but does not say that different features are necessary to describe American English vowels. Other researchers such as Ladefoged & Johnson (2011), Yavas (2006), and Davenport & Hannahs (2005) use tense/lax distinction to account for English vowel system. Edwards (2003) adapts a distinctive feature analysis introduced by Chomsky & Halle (1968) to describe each American English vowel: /i/ (as in *beat*) and /ɪ/ (as in *bit*) are different by the feature [+tense] and [-tense].

Lindsey (1990) argues that while British English retains a long/short vowel distinction, American English developed a tense/lax distinction. As evidence to support his claim, Lindsey (1990) compares the results of Peterson & Lehiste (1960) and Wiik (1965), and concludes that long/tense vowels are longer than lax/short vowels by a greater extent in British English than in American English.

In the present study, these two groups of vowels will be treated as tense and lax vowels rather than long and short vowels, in part because these two groups of vowels are different not just in duration but in quality as well. Vowels are generally longer in stressed syllables than in unstressed syllables. More importantly, lax vowels cannot occur in open syllables, while tense vowels can occur in both open and closed syllables. Jespersen (1926) notes that "ein wichtiger Umstand, der den Bau von Silben betrifft, ist noch nicht besprochen, nämlich die Art und Weise, wie ein Konsonant mit einem Vokal verbunden wird: kommt er schnell und bricht den Vokal in dem Augenblick ab, wo dieser am kräftigsten gesprochen wird, so haben wir 'festen Anschluss' (zwischen Vokal und folgenden Konsonanten); wenn er dagegen erst einige Zeit nach der kräftigsten Aussprache des Vokals kommt, wenn der Vokalklang also schon vor Eintritt des Konsonanten etwa geschwächt ist, so haben wir 'losen Anschluss" (One important circumstance concerning the construction of syllables has not yet been discussed, namely the way a consonant is connected to a vowel: if it comes quickly and aborts the vowel the moment the vowel is the strongest, so we have a 'firm connection' between vowel and the following consonants; if, on the other hand, it comes only sometime after the strongest pronunciation of the vowel, that is, if the vowel sound is weakened even before the consonant enters, we have a "loose connection"). What Jespersen (1926) implies is that the transition from a lax vowel to the following consonant is abrupt, while that from a tense vowel to the following consonant is more gradual. Takebayashi (1998: 287) cites Jespersen (1926) and adds that a lax vowel is more tightly connected with the following consonant than is a tense vowel, and that in a transition from a lax vowel to the following consonant, a glottal stop [ʔ] accompanies, but in a transition from a tense vowel to the following consonant, [ʔ] is rarely observed. Takebayashi (1998: 287) further mentions that the following consonant tends to be longer after a lax vowel. Therefore, no matter how close these two groups of vowels are in duration or in quality, they should be treated differently in English phonology.

Among the other differences between the two vowel systems shown in Tables 2 and 3 is the absence of /ɒ/ in the American English vowel system. The CEPD simply states that "there is no /ɒ/ vowel in GA (=General American)." On this, Deterding (2015) notes that "the majority of people in the United States pronounce words such as *hot* and *shop* with /ɑː/ rather than /ɒ/" (p. 77). Boberg (2015), on the other hand, points out that /ɑː/ as in *palm* is prone to merging, and that in American English it merged with /ɒ/ as in *lot*. Boberg continues, noting that "*lot* began to shift down and forward from its original mid-back position by the seventeenth century, reaching a low-central unrounded position, approximately [ɑ]" (p. 233). The merged vowel can occur in open syllables as in *rah* and *pa*, and so it is treated as a tense vowel. It will be transcribed as /ɑ/.

There are some inconsistencies in phonetic notation of English vowels among researchers and authors. For instance, CEPD, Cruttenden (2014) and Rogen-Revell (2011) all transcribe the vowel in "beat" as /iː/, while Ladefoged (1999), Edwards (2003), Yavas (2006), and Ladefoged and Johnson (2011) all transcribe it as /i/. The transcription /iː/ makes durational difference more explicit, but because there is no phonemic contrast between /i/ and /iː/, /i/ is sufficient and more appropriate to transcribe the vowel in "beat." Similarly, "goose" and "foot" vowels will be transcribed as /u/ and /ʊ/ respectively. The vowel in "dress" can be transcribed as /e/ (CEPD, Cruttenden 2014, and Rogen-Revell

2011) or /ɛ/ (Ladefoged 1999, Edwards 2003, Yavas 2006, Ladefoged and Johnson 2011). Here the vowel is transcribed as /ɛ/ in part because the vowel is closer to cardinal /ɛ/ than to cardinal /e/, and the "face" vowel is sometimes transcribed as /e/ (Ladefoged 1999, Edwards 2003). The vowel in "trap" is sometimes transcribed as /a/ (Giegerich 1992), but it is more commonly transcribed as /æ/, and, in order to make it explicit that it is a front vowel, here it will be transcribed as /æ/.

The vertical and horizontal position of a vowel can be acoustically measured by formant frequencies. Crystal (1997) defines formant as "a concentration of acoustic energy, reflecting the way air from the lungs vibrates in the vocal tract, as it changes its shape. For any vowel, the air vibrates at many different frequencies all at once and the most dominant frequencies combine to produce the distinctive vowel qualities." The first formant frequencies inversely correspond to vowel height, i.e., the higher the frequency is, the lower the vowel is. The second formant frequencies, on the other hand, correspond to the frontedness of a vowel: i.e., the higher the frequency is, the more fronted the vowel is. The third formant frequencies correspond to the roundedness of a vowel, but, because neither English nor Japanese has a rounded/unrounded contrast, the frequencies of the first two formants are enough to describe English and Japanese vowels.

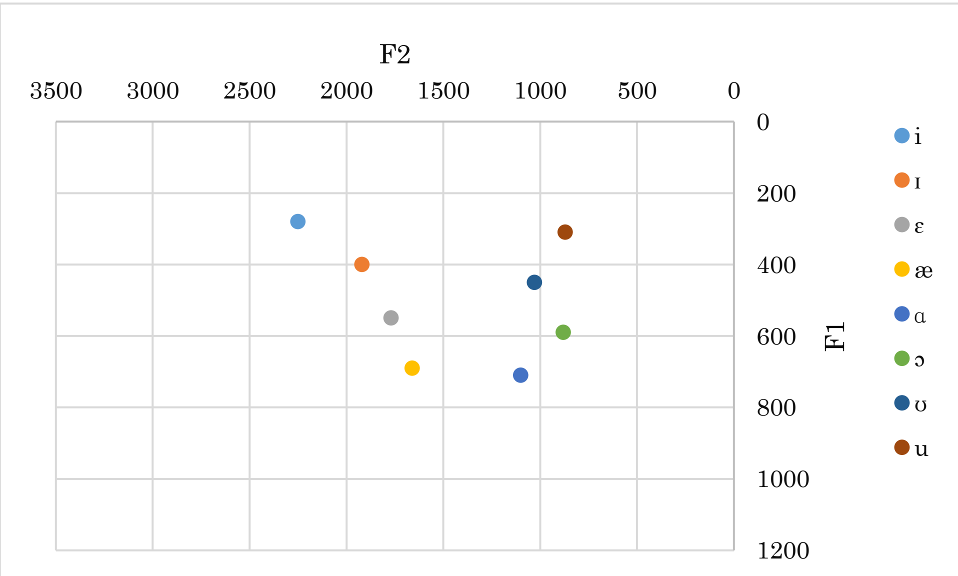

Figure 2. F1 and F2 frequencies of eight American English vowels (Adapted from Ladefoged & Johnson, 2011)

Figure 2 shows the frequencies of the first two formants of eight American English vowels (monophthongs) based on the data shown in Ladefoged and Johnson (2011: 193). Although the authors do not provide information about the speakers' gender, it can be assumed that formant frequencies were extracted from male utterances.

Similarly, Yavas (2006) provides formant frequencies of ten American English vowels, including /ʌ/ and /ɝ/ as shown in Figure 3.

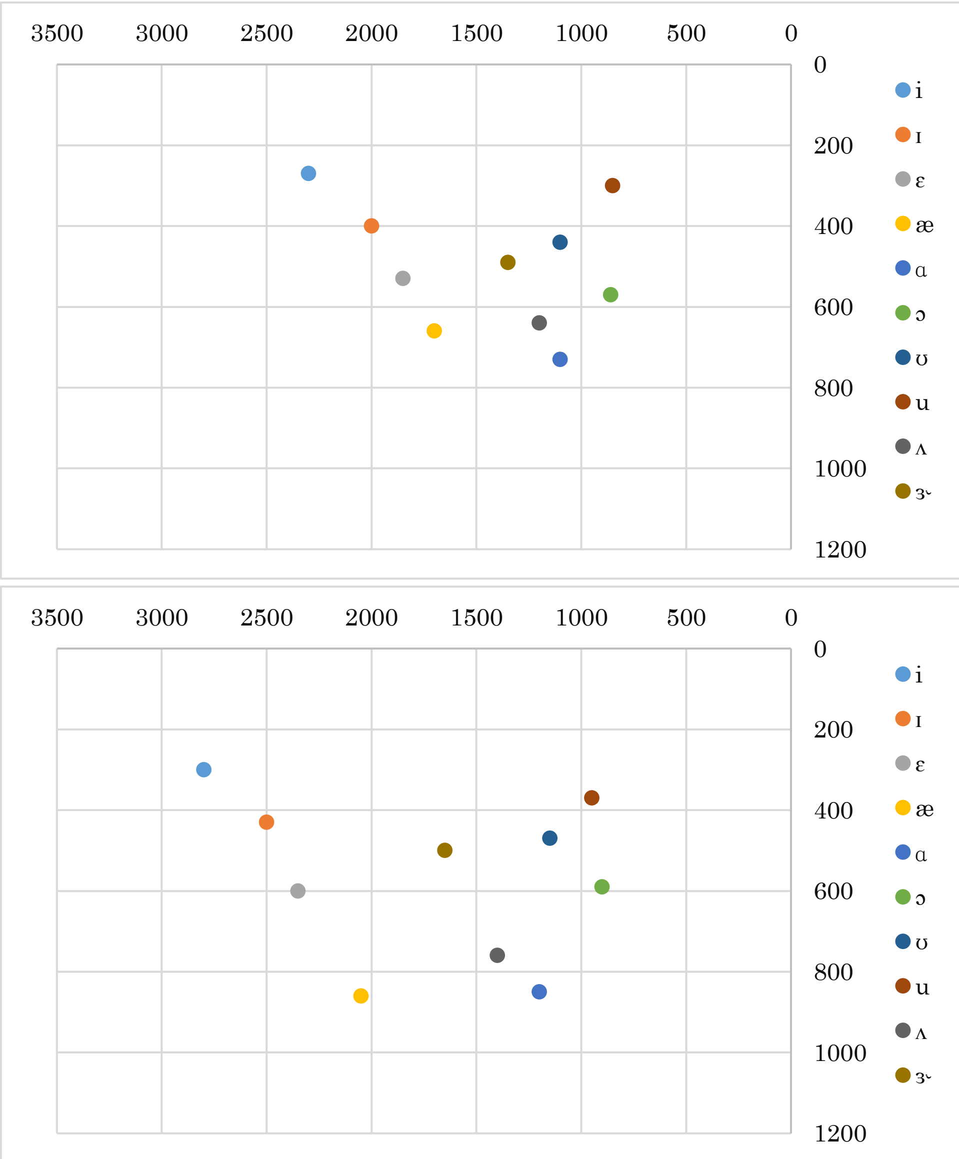

Figure 3: F1 and F2 frequencies of ten American English vowels: male speakers (upper level), female speakers (lower level) (Adapted from Yavas,2006)

Bradlow (1993a) also provides mean F1 and F2 frequencies of 11 American English monophthongs uttered in /CVC/ frame by 4 male speakers.

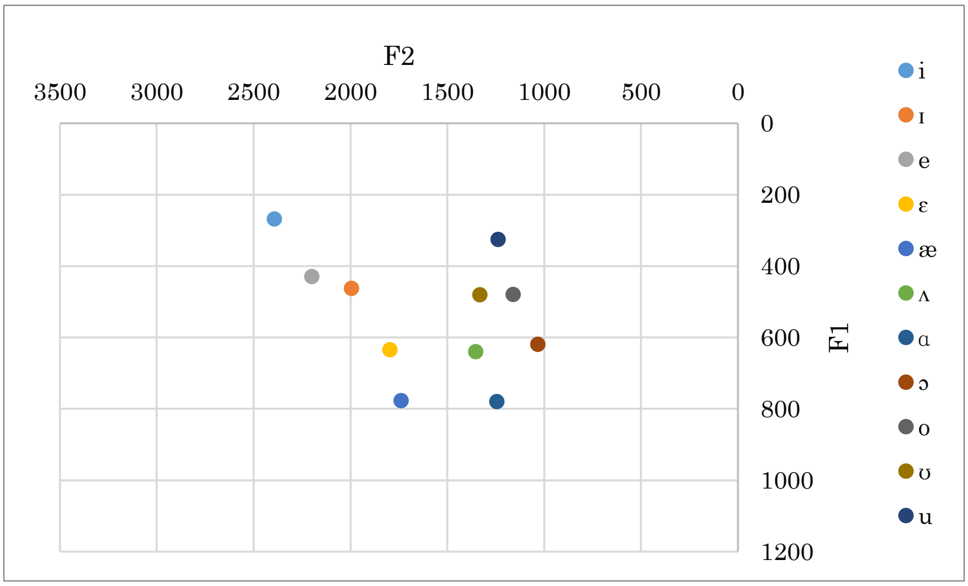

Figure 4. F1 and F2 frequencies of 11 American English vowels: averaged across 20 tokens (4 speakers × 5 repetitions) (Adapted from Bradlow 1993a)

Here "face" and "goat" vowels are treated as monophthongs and transcribed as /e/ and /o/, respectively. These two vowels are usually classified as monophthongs in American English (Bradlow 1993, Ladefoged 1999). In British English, these two vowels are usually treated as diphthongs. Regardless of their status as monophthongs or diphthongs in American English, these two vowels are often accompanied by offglides. In Frieda and Nozawa (2007), Japanese listeners equated these two vowels with the Japanese vowel sequences /ei/ and /ou/, respectively.

In Bradlow's data (Bradlow, 1993a), low vowels /æ/ and /ɑ/ are lower than those in Ladefoged and Johnson (2011) and in Yavas (2006). Moreover, the high back vowel /u/ is somewhat fronted in Bradlow (1993)'s data. Despite these differences, the relative position of each vowel is basically the same.

Thus, American English has the following full vowels (as opposed to reduced vowels):

Tense vowels: /i/, /e/, /ɑ/, /ɔ/, /o/, /u/, /ɝ/

Lax vowels: /ɪ/, /ɛ/, /æ/, /ʊ/, /ʌ/

Diphthongs: /aɪ/, /ɔɪ/, /aʊ/

1.4 Japanese vowel system

Japanese has a much simpler vowel system than that of American English and has only five qualitatively distinctive short vowels /a, e, i, o, u/. The Japanese low vowel /a/ is between Cardinal vowel [a] and [ɑ] (Vance 1987, 2008). The Japanese mid front vowel /e/ is unrounded and is between Cardinal vowel /e/ and /ɛ/ (Vance, 1987, 2008). The Japanese high front vowel /i/ is also unrounded and is in the same tongue position as cardinal vowel [i] (Vance, 2008). The Japanese mid back vowel is weakly rounded and is between cardinal vowel [o] and [ɔ] (Vance 2008). Vance (2008) compares Japanese high back vowel /u/ and French /u/, and comments that, while French /u/ is produced with lip protrusion, Japanese /u/ is produced with lip compression, and that the compression is weaker in a casual speech (pp. 53-54), and thus it is commonly described as unrounded. Vance (2008) also adds that the Japanese /u/ is a bit more forward than cardinal vowel [u]. Saito (1996), on the other hand, describes the Japanese /u/ as close to the cardinal vowel [u]. On this point, Sugito (1996) performed an acoustic analysis of the five vowels uttered by two Japanese speakers (One from Tokyo, and the other from Osaka), and comments that /u/ in the Tokyo area is more forwarded and gets closer to [i], while /u/ in the Osaka area, it is close to [o] (p. 7). Homma (1969) recorded her own speech and commented that the /u/ of Kyoto-accented Japanese may be more back than that of standard Japanese, but that it is still somewhat more forward than cardinal vowel [u]. Imanishi and Miwa (1989) compare the five vowels uttered by native speakers recruited at five different locations across Japan. Figure 2 shows mean frequencies of F1 and F2 of the five vowels uttered by speakers of standard Japanese. Imanishi and Miwa (1989)'s data agree with Vance (1987, 2008) and Sugito's (1996) observation.

Fujimura (1972) also measured the formant frequencies of five vowels uttered by five adult male speakers of Japanese. Figure 5 is based on this result. Tokens that Fujimura (1972) notes as unreliable are excluded. Basically, the location of the five vowels in the vowel space is more or less the same as Imanishi and Miwa's (1989) data, but what is noticeable here is largely individual difference. One speaker's /a/ token is extremely higher than those of the other talkers. Two talkers' /u/ tokens are more fronted, and one talker's /e/ token is lower

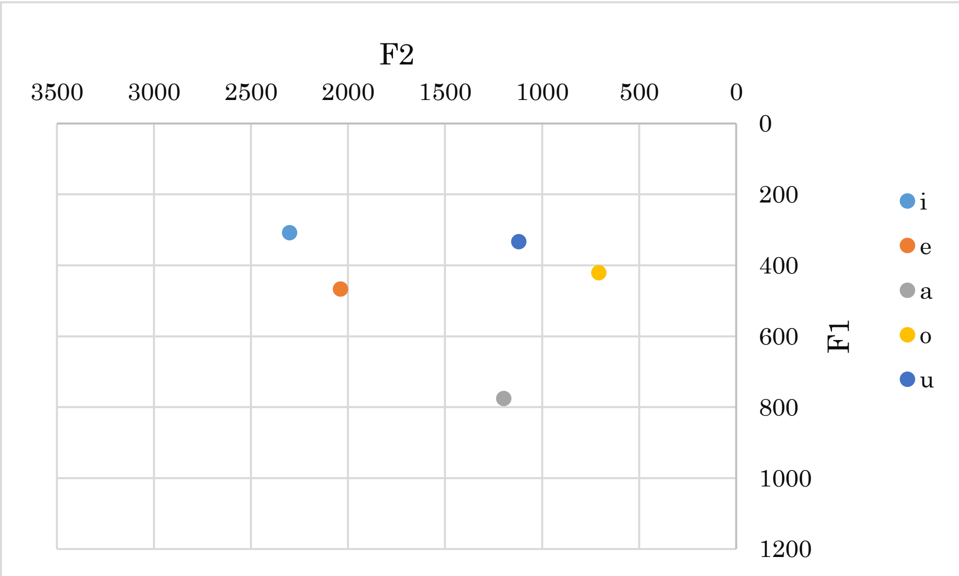

Figure 5. F1 and F2 frequencies of Japanese /a, e, i, o, u/ uttered by a speaker of Standard Japanese (Adapted from Imanishi, 1989: 91)

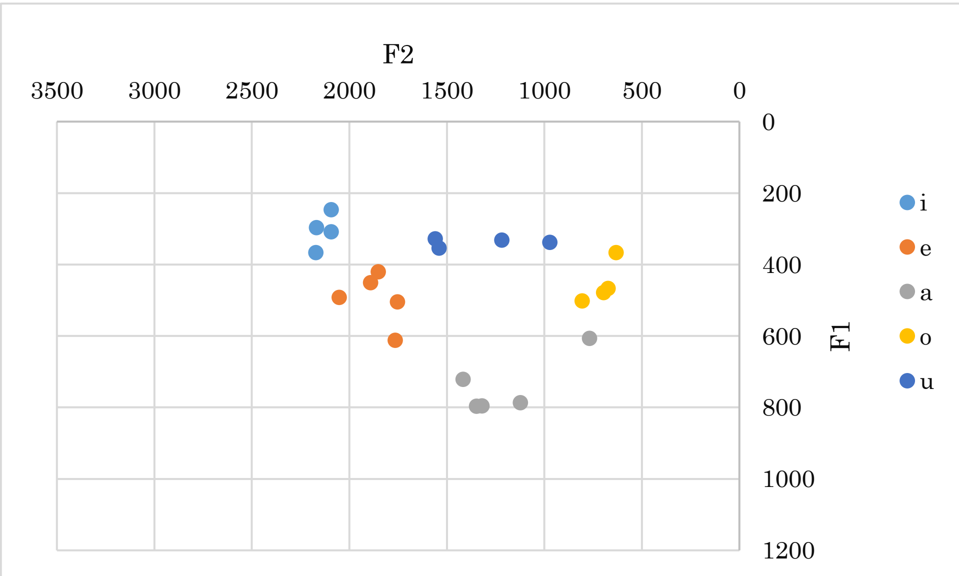

Figure 6. F1 and F2 frequencies of Japanese /a, e, i, o, u/ uttered by five male speakers (Adapted from Fujimura, 1972: 218)

Although Japanese has only five vowels, these five short vowels /a, e, i, o, u/ have spectrally equal, long vowel counterparts /aː, eː, iː, oː, uː/. Nozawa (2018) measured the five Japanese vowels uttered by four female speakers in /hVdo/ context. Two of these speakers are from Nagoya, and the other two are from Osaka and Kobe. Figures 7 and 8 are based on Nozawa's (2018) data. Long vowels are slightly more peripheral than short vowel equivalents. The vowels / iː/ and /eː/ are more fronted than /e/ and /i/, while /oː/ and /uː/ are a little backer than /o/ and /u/, and /aː/ is lower than /a/. This tendency agrees with the results of Hirata & Tsukada (2003, 2009).

Obviously, American English vowel space is more crowded or partitioned than its Japanese counterpart. To take front vowels as an example, in American English there are three vowels between /i/ and /æ/, but, in Japanese, there is only /e/ between /i/ and /a/, and the space between /e/ and /a/ is large. Japanese /i/ and /e/, and /o/ and /u/ are close, and only /a/ is a bit distant from the other vowels. The vowels /u/ and /o/ can be distinguished horizontally rather than vertically.

Despite the difference in the size of vowel inventory, the vowels of American English and Japanese spread widely in the vowel space, rather than occupying a corner of that space. Liljencrants and Lindblom (1972) demonstrated that, regardless of the number of vowels, vowels of a language tend to spread widely over the vowel space to maintain auditory difference among vowels. Bradlow (1993b) measured frequencies of F1 and F2 of /i/, /e/, /o/, /u/ of American English, Spanish, and Greek. Spanish and Greek each have a five-vowel system like Japanese. She found that the English vowel space is expanded acoustically to accommodate more vowel categories than Spanish and Greek. If this is the case, the vowel space of American English may be expanded more than that of Japanese as well.

Bearing in mind these differences in vowel systems, in the next section, this dissertation will explore how borrowings from English are transcribed.

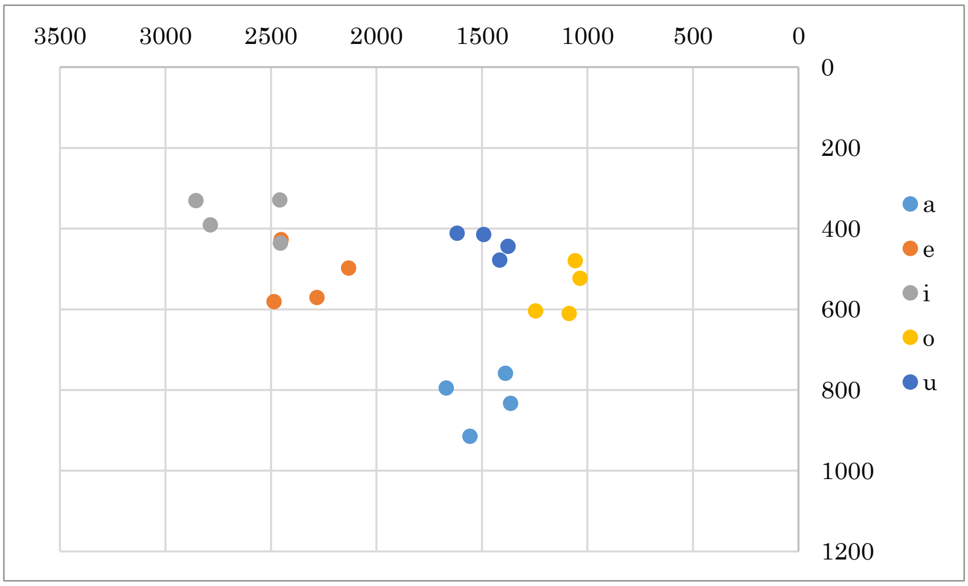

Figure 7. F1 and F2 frequencies of Japanese short /a, e, i, o, u/ uttered by four female speakers (Adapted from Nozawa, 2018)

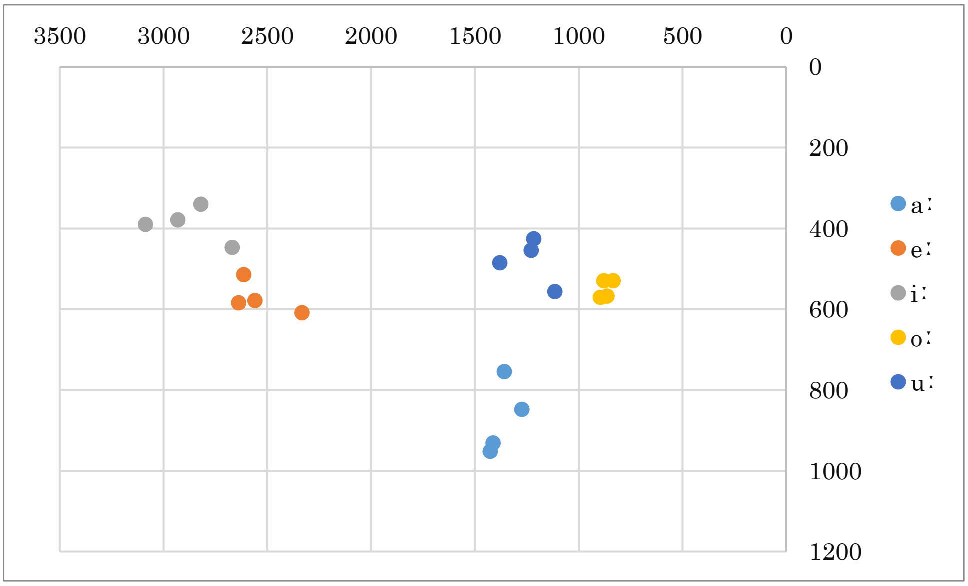

Figure 8. F1 and F2 frequencies of Japanese long /aː, eː, iː, oː, uː/ uttered by 4 female speakers (Adapted from Nozawa, 2018)

1.5 Japanese adaptation of English vowels

In 1872, the Japanese government executed the Education Law (*gakusei*), and English became a compulsory subject in middle and elementary schools, and about 2300 foreign advisors (*oyatoi gaikokujin*) were brought in. Of these foreign advisors, 40% were

British, 16% were American, and 1% was Australian (Irwin 2011: 54). More than half of the advisors were English speakers, and the majority was from Britain.

Boddernberg (2011) notes that "in the second half of [the] 19th century and at the beginning of the 20th century Received Pronunciation was taught at Japanese schools, for Great Britain and her variety of (British) English had an extremely high prestige and even American teachers tried to speak RP in Japanese classrooms."

Quackenbush (1974, cited in Irwin 2011) indicates, "When dictionary traditions were being formulated in the late 19th Century, English language scholars were heavily influenced by foreign advisors from Britain (Irwin 2011, 79). Irwin (2011) attributes the established Japanese adaptation of English phones to the influence of these foreign advisors from Britain. Quackenbush (1974) notes, "The pronunciation of a word is determined by how it is spelled in the katakana writing system rather than the other way around. The katakana spellings are based, with very few exceptions, on a set of conventions that evolved under the influence of Britishers, who have dominated English language teaching for most of its history in Japan. Ultimately this set of conventions derives from the phonological differences between Japanese and cultivated British English" (p.61). Irwin (2011) adds that British influence is typically seen in the adaptation of the vowel in the word "lot" and in the English rhotic vowels. This British influence remains today, even though American English is taught almost exclusively in Japan now.

According to Irwin (2011), a source word of a donor language can be adapted into Japanese either auditorily or orthographically. Moreover, if the source is auditory, then the adaptation is based on auditory input, whereas if the source is orthographic, then the adaptation is based on dictionary traditions (p. 76). Irwin (2011) notes that the majority of loanwords in Japanese have an orthographic source and that these loanwords comply with dictionary traditions. Irwin (2011) adds that "all (dictionary traditions) have in common the fact that their adaptation rules were established and standardized by Japanese scholars of foreign languages, then perpetuated through their pedagogical practices and foreign language textbooks" (p.78). Irwin (2011) calls loanwords based on auditory input auditory "loans," and those are based on dictionary traditions or dictionary loans. One good example of auditory and dictionary loans may be *hebon* and *heppbaan*, both of which derive from "Hepburn."

Irwin (2011) introduces a third type of loanwords. Like dictionary loans, their source is based on orthography, but their adaptation is based on an inaccurate representation of pronunciation. These loans are called spelling loans (p.79). Examples of this category may include *sutajiamu* (stadium), *suponji* (sponge), and *monkii* (monkey).

Mutsukawa (2009) introduces two different models of loanword adaptation. (1) The phonetic-based model: this model assumes that speakers of the host language do not know the phonological representation of the source language, and "the speakers of the host language perceive inputs in accordance with the phonological system of the host language" (p.8). (2) The phonology-based model: this model assumes that "loanwords are introduced by bilinguals who have competencies in both the source and host languages, and that phonological output of the source language, which is based on the phonemic representation, is the input to the host language and the input is incorporated directly into the lexicon of the host language" (p.9). Mutsukawa (2009) is not very specific about which model he supports nor about which model is applied to the Japanese adaptation of English words, but because Mutsukawa (2009) provides a commonly accepted transcription custom of English vowels in Japanese orthography, it can be assumed that a phonology-based rule is applied at least in relation to the Japanese adaptation of English vowels.

Examples that Mutsukawa (2009) provides are not comprehensive, but similar examples are given in Kobayashi (2005) and Irwin (2011) as well. The following are commonly seen correspondences between English vowels and their Japanese adaptations. Examples of such commonly accepted adaptations are shown in Table 2. All of the examples are taken either from, Kobayashi (2005), Mutsukawa (2009), or Irwin (2011).

English has at least 14 vowels, and Japanese has only five qualitatively distinctive vowels, so to adapt English vowels to Japanese, more than one English vowel falls into one Japanese vowel category. Durational difference in Japanese is used to adapting English vowels. As shown in Table 2, when English vowels are adapted, lax vowels are usually transcribed as short vowels, whereas tense vowels and diphthongs are transcribed as long vowels or a combination of two short vowels.

To see the adaptation pattern diagrammatically, Japanese vowels are plotted on American English vowels on Figure 4. As seen Figure 9, American English vowels are adapted to make them close to Japanese categories. These are examples of what Irwin (2011) calls dictionary loans: the low back vowel /ɑ/ is adapted as /o/ rather than /a/. This is largely because, as mentioned above, Japanese scholars of English were strongly influenced by foreign advisors from Britain. The British equivalent of /ɑ/ is /ɒ/, which is a little higher and may be closer to Japanese /o/ than to /a/. Quackenbush (1974) states that "most loanwords, whether derived from speech or from writing, and without regard to dialect of origin, are conventionally written and pronounced in Japanese as if they had been borrowed from a precisely articulated form of British English" (p. 71). Irwin (2011) gives a good example of a British or American origin of loanwords. The word "soccer" is transcribed as *sakkaa* rather than *sokkaa* because the term "soccer" is mostly used in the US and rarely used in Britain. This is explained by the fact that "soccer" is from the US, but the soccer term "offside" is *ofusaido*. Likewise, "volleyball" is *bareebooru*, but the tennis term "volley" is *boree* (pp. 96-7).

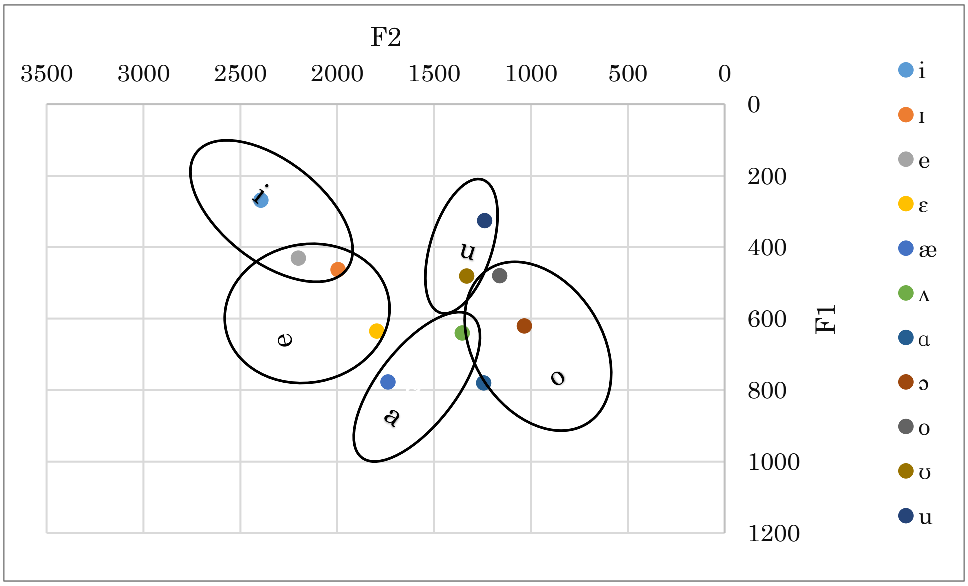

Figure 9. Japanese vowel categories plotted on American English vowel space based on adaptation pattern (F1, F2 values of American English vowel are from Figure 4)

Table 3

Examples of Japanese adaptation of English vowels

English vowels	English examples	Japanese vowels	Japanese adaptation
/i/	key, scene	/iː/	*kii, shiin*
/e/	date, race	/eː/	*deeto, reesu*
/ɔ/	water, call	/oː/	*wootaa, kooru*
/o/	zone, boat	/oː/	*zoon, booto*
/u/	cue, room	/uː/	*kyuu, ruum*
/ɝ/	bird, curtain	/aː/	*baado, kaaten*
/ɪ/	pin, picnic	/i/	*pin, pikunikku*
/ɛ/	pen, net	/e/	*pen, netto*
/æ/	ham, map	/a/	*hamu, mappu*
/ɑ/	top	/o/	*toppu*
/ʊ/	book, looks	/u/	*bukku, rukkusu*
/ʌ/	Sunday, cut	/a/	*sandei, katto*
/aɪ/	ice, line	/ai/	*aisu, rain*
/ɔɪ/	oil, toilet paper	/oi/	*oiru, toirettopeepaa*
/aʊ/	pouch, house	/au/	*pauchi, hausu*

Notes. Mutsukawa (2009) states that /ɑ/ (in his transcription [a]) is adapted as [a], but he gives “top” as an example, so I assume that it is a typographical error.

The Ministry of Education, Sports, Culture, Science and Technology (MEXT)[2] released a guideline to transcribe borrowings from other languages as a notification from the Cabinet #2. The guideline shows loanwords from English are, for the most part, adapted in accordance with adaptation of dictionary loans, but there are some exceptions. Table 3 shows examples of these exceptions.

This guideline does not intend to correct the current use of loanwords, but instead it appears to approve widely accepted usage even when they are inconsistent. For example,

2 http://www.mext.go.jp/b_menu/hakusho/nc/k19910628002/k19910628002.html

/æ/ is usually transcribed as /a/, but when it is preceded by a velar plosive, it is adapted as /ja/ as shown in Table 3. Other examples include *kyabin* (cabin), *gyararii* (gallery), and *gyanburu* (gamble). This is largely because, in the transition from /k/ or /g/ to /æ/, F2 makes a downward movement (Olive, Greenwood & Coleman 1993), and native Japanese speakers hear a /j/-like glide between /k/ or /g/ and the vowel. Thus, the input is at least partially auditory. The vowel /ʌ/ is usually transcribed as /a/, but, as in Table 3, it can be adapted as /o/. These are examples of spelling loans, such as those of *sponge* and *monkey* mentioned above. In all of these examples, /ʌ/ is represented by the letter "o."

Table 4.

Examples of loanwords that are not dictionary loans

English vowels	Japanese vowels	examples	source words
æ	ja	kyanpu/kjaNpu/	camp
		kyandee/kjaNdeː/	candy
ʌ	o	rondon /ɾoNdoN/	London
		botan /botaN/	button
e	ei	eito /eito/	eight
		supein /supeiN/	Spain
		keinzu /keiNzu/	Keines
		peinto /peinto/	paint
	e	epron /epɾoN/	apron
		uehaasu /uehaːsu/	wafers
o	o	posutaa /posutaː/	poster

(Source: MEXT Website)

Similarly, the English /e/, as in game, is usually adapted as a long /eː/ in Japanese. Thus, *game* becomes *geemu*. In Table 3, provided examples in which /e/ is adapted as /ei/. In these examples, the English vowel is represented by the letters *ei* or *ai*; therefore, it can be assumed that the adaptation is a result of analogical inference. Also shown are instances of the English /e/ adapted as a Japanese /e/. These are both disyllable words, and vowels tend to be shorter in multisyllable words than in monosyllable words, so English /e/ may have been adapted to short /e/ rather than to long /eː/. Thus, the input is at least partially auditory.

Nozawa's (2018) findings reveal that this commonly-accepted adaptation of English vowels (as shown in Table 2) strongly affects the identification of American English vowels by native Japanese speakers. In Nozawa (2018), Japanese speakers were asked to choose the English vowels closest in sound to the Japanese vowels they heard. They chose /æ, ɛ, ɪ, ɑ, ʊ/ to correspond to the Japanese short vowels /a, e, i, o, u/, uttered in a frame /hVdo/ (the final /o/ was deleted when used as stimuli). However, when they were asked to choose the Japanese vowel that best represents the American English vowels they heard, /ɑ/ was chosen as the closest vowel to Japanese /a/ more frequently than was /æ/, and /ɑ/ was also a better exemplar of Japanese /a/ than was /æ/ (as shown in higher category goodness rating). In an identification experiment, /ɑ/ was more frequently misidentified as /æ/ (72.9%) than it was identified correctly (10.4%).

1.6 Major hypotheses about the perception of non-native phones

In these sections, I refer to two major hypotheses about the perception of non-native phones.

1.6.1 Perceptual Assimilation Model (Best 1995)

Best's Perceptual Assimilation Model (=PAM) (Best 1995) postulates that non-native phones are recognized in terms of learners' L1 categories as

(1) Assimilated to a native category

(2) Assimilated as an uncategorizable speech sound, or

(3) Not assimilated to speech (nonspeech sound).

PAM states that those non-native phones assimilated to native categories are further divided into (a) good exemplars of that category, (b) acceptable but not ideal exemplars of the category, or (c) notably deviant exemplars of the category.

According to PAM, the discriminability of non-native contrasts is highly predictable as combinations of the above.

Two-Category Assimilation (TC Type): If each non-native phone is assimilated to a different native category, the discrimination is expected to be excellent.

Category-Goodness Difference (CG Type): If both sounds of a non-native contrast are assimilated to one native category, but when they differ in the divergence from the native ideal (one is close to the native prototype, and the other is deviant), then the discrimination is expected to be moderate to very good.

Single-Category Assimilation (SC Type): When both non-native sounds are assimilated to the same category, and they are equally close to or deviant from the native ideal, then the discrimination is expected to be poor.

Both Uncategorizable (UU Type): When both non-native sounds are "outside of any particular native category," then the discrimination is expected to be poor to very good, depending on how close these two non-native phones are.

Uncategorized versus Categorized (UC Type): When one non-native sound is assimilated to a native category, and the other is outside native categories, the discrimination is expected to be very good.

Non-assimilable (NA Type): When both non-native sounds are heard as nonspeech sounds, the discrimination is expected to be good to very good.

According to PAM's prediction, if both members of a non-native contrast are categorized in terms of a learner's native categories, the discrimination should be most accurate in TC type, followed by CG type, and the discrimination is least accurate in SC type.

One problem with PAM is that it is not clear how often a non-native phone has to be classified as an exemplar of the same L1 category in order to be categorized. There seems to be no definite consensus among researchers about this point. Non-native segments can be classified as exemplars of multiple native categories. Lengeris (2009) required a non-native phone to be labeled as an exemplar of the same native category in 60% of instances in order to be considered "categorized," and Bundagaard-Nielsen et al. (2010) adopted 50% of the criteria. Harnsberger (2001), on the other hand, required a non-native phone to be labeled as an exemplar of the same native categories in 90% of instances in order to be "categorized." If the criteria are different, the assimilation pattern will also be different. Furthermore, it is unclear how different a category goodness rating has to be in order to call the non-native contrast a "Category Goodness" Type.

1.6.2 Speech Learning Model (=SLM) (Flege 1995)

Flege (1995: 239) summarizes the postulates and hypotheses of SLM as follows:

Postulates

P1 The mechanism and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning.

P2 Language-specific aspect of speech sounds are specified in long-term memory representations called phonetic categories.

P3 Phonetic categories established in childhood for L1 sounds evolve over the life span to reflect the properties of all L1 or L2 phones identified as a realization of each category.

P4 Bilinguals strive to maintain contrast between L1 and L2 phonetic categories, which exist in a common phonological space.

Hypotheses

H1 Sounds in the L1 and L2 are related perceptually to one another at position-sensitive allophonic level, rather than at a more abstract phonemic level.

H2 A new phonetic category can be established for an L2 sound that differs phonetically from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds.

H3 The greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned.

H4 The likelihood of phonetic differences between L1 and L2 sounds, and between L2 sounds that are noncontrastive in the L1, being discerned decreases as the AOL increases.

H5 Category formation for an L2 sound may be blocked by the mechanism of equivalence classification. When this happens, a single phonetic category will be used to process perceptually-linked L1 and L2 sounds (diaphones). Eventually, the diaphones resemble one another in production.

H6 The phonetic category established for L2 sounds by a bilingual speaker may differ from a monolingual speaker's if 1) the bilingual's category is "deflected" away from an L1 category to maintain phonetic contrast between categories in a common L1-L2 phonological space; or 2) the bilingual's representation is based on different features, or feature weights, than a monolingual's.

H7 The production of a sound eventually corresponds to the properties represented in its phonetic category representation.

Aoyama et al. (2004) measured the perception and production of English /r/ and /l/ among native Japanese adults and children and demonstrated that native Japanese speakers showed greater improvement for their perception and production of /r/ than for /l/. Significantly, Japanese liquid /ɾ/ is said to be perceptually closer to English /l/ than to /r/. Aoyama et al. (2004) concluded that these results can be taken as a support for a hypothesis made by SLM that "a new phonetic category can be established for an L2 sound that differs phonetically from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds (H2)." In other words, these Japanese children improved their sensitivity to discern phonetic difference between English /r/ and Japanese /ɾ/ during this course of the study, so, in Best's term (Best, 1995), English /r/-/l/ contrast changed from Single-Category Type to Category-Goodness Type.

SLM does not make an explicit prediction as to whether a given non-native contrast is easy or difficult, but it can be assumed that if one member of the contrast is remarkably distant from a learner's L1 ideal, while the other member is very close, then at least some phonetic differences between the two members would be discerned.

SLM makes an explicit hypothesis about category formation in L2 learning; therefore, the model is appropriate for a longitudinal study that aims to investigate learners' change or development. The present study does not intend to determine learners' growth, but instead to measure Japanese speakers' perception and production accuracy of American English at a given time.

SLM also presupposes that learners continue being exposed to their target language, and, through this exposure, their perception changes. This is possible when learners are in an environment where the target language is spoken, but, in an EFL context, the exposure to English is limited or sporadic, and it is unlikely that a learner can receive enough input to form a new category. Furthermore, in an EFL context, unlike in an ESL context, there is no local dialect or variety of the target language that can serve as learners' input. Native speakers of English whom a learner encounters may be American, or British, or Australian, for example.

1.7 Previous Studies on vowel perception

Citing results of several previous studies, Rosner and Pickering (1994) suggest that "the center frequencies of the first two formants or their auditory transformation are sufficient to separate the different vowels of a language" (p. 96). So the vertical and horizontal positions of a vowel in a vowel space can also account for the auditory perception of a vowel. Findings by Picket (1957) support this argument from Rosner & Pickering Picket performed a masking experiment, in which listeners heard English vowels in noises of various spectra. He found that when a noise masks one formant, the unmasked formant is correctly perceived. Findings by Pols., van der Kamp and Plomp (1969) demonstrate that F1 and F2 frequencies are the most important in vowel perception, indicating that perceptual and physical properties are correlated. Schouten and van Hessen (1992) report that the perception of vowels is much less categorical than that of the place of consonant articulation. Perceptual boundaries between vowels may not be as solid as those between the places of articulation of consonants, but native speakers of American English can not only choose the categorically different stimulus in triads, but they can also label three different tokens of categorically same stimuli as the same (Nozawa & Flege 2001, Frieda & Nozawa 2007). In these studies, participants' sensitivity to discriminate 14 American English vowel pairs was assessed, and no significant difference was observed between native speakers' A' scores for all of the 14 vowel pairs, whereas nonnative speakers A' scores were not just lower than those of native speakers, but also some vowel pairs were more difficult for them to discriminate. This suggests that native speakers have little difficulty perceiving naturally-spoken vowels categorically. More recently, Zhang, Chen, Yan, Wang, Shi & Ng (2016) compared the categorical perception of vowels by native Korean and Mandarin listeners, and found that Korean listeners' perception is more categorical along the /a/-/ɜ/-/u/ vowel continuum, and that Mandarin listeners more often label stimuli as /a/ than Korean listeners.

It is because native vowels are perceived categorically that the L1 vowel categories exercise influence on the perception of non-native vowels. Polka (1995) compared the discrimination accuracy of German two-vowel pairs /y/ vs /u/ and /Y/ vs /U/ by monolingual English listeners and found that the tense vowel pair /y/ vs /u/ was more accurately discriminated. Polka observed that a larger difference in category goodness was perceived more often in tense vowel pairs than in lax vowel pairs.

Bohn and Flege (1990) examined the perception of English vowels /i, ɪ, ɛ, æ/ by native German speakers. Prior to the experiment, the authors assumed that English /i, ɪ, ɛ/ were close to German /i, ɪ, ɛ/, whereas no German vowel was perceptually close to /æ/. First, the German participants chose a German vowel that is closest to each English vowel, and the results revealed that the English /i/ and /ɪ/ were consistently equated to the German /i/ (BIET) and /ɪ/ (BITT) respectively. The responses to /ɛ/ varied according to talkers, but /ɛ/ (BETT) was the preferred response. As for /æ/, /ɛː/ (BÄHT), or /ɛɐ̯/ (BERT) were preferred, but the participants' confidence was much lower. Then the authors had the participants hear two synthesized continua beat-bit, bet-bat, and found that, while the German learners of English identified the endpoint of beat-bit predominantly categorically, the German participants' identification of the bet-bat continuum was more gradient with much larger standard deviations.

Flege & McKay (2004) examined the perception of Canadian English vowels by Italian university students living in Canada, and found that, of the 9 vowel pairs, the Italian listeners' sensitivity to discriminate two vowel pairs /ɛ/-/æ/ and /ɒ/-/ʌ/ did not reach the chance level. The authors then performed a perceptual assimilation task and found that, for most of the vowel pairs, a perceptual assimilation pattern can account for discrimination accuracy. The larger the overlap, the less accurate the discrimination is. But for two particular vowel pairs, the authors found that perceptual assimilation and discrimination accuracy did not agree. For /eɪ/-/ɛ/, high overlap is observed (87%), but discrimination is relatively better; on the contrary, /ɪ/-/ɛ/ is relatively poorly discriminated even though the overlap is low. To investigate the reason for this discrepancy, the authors recalculated the overlap scores on an individual basis, and they reached a low overlap for /eɪ/-/ɛ/ (40%). To account for the low discrimination accuracy of /ɪ/-/ɛ/, the authors also added that "/ɪ/ and /ɛ/ differ relatively little in terms of their midpoint formant frequencies and duration and do not show a different pattern of formant movement" (p.13).

As for studies using native Japanese speakers, Morrison (2002) compared native Japanese listeners' discrimination accuracy of the English vowel pair /i/-/ɪ/ uttered before voiceless and voiced consonants and found that the vowel pair was more accurately discriminated in the context of the voiced consonant. He also performed a perceptual assimilation task and found that, in the context of voiceless consonant, both /i/ and /ɪ/ were equated to the Japanese high short vowel /i/, whereas in the context of the voiced consonant, /i/ was equated to Japanese high long vowel /iː/, and /ɪ/ was equated to short vowel /i/. The English /i/ is intrinsically longer than /ɪ/, but, in the context of the voiceless consonant, the vowel is not long enough to equate to the Japanese long vowel /iː/. Consequently, both /i/ and /ɪ/ equated to the short vowel /i/. This is an example that demonstrates how native Japanese speakers rely on durational differences to differentiate English /i/ and /ɪ/.

Similarly, Nozawa and Wayland (2012) demonstrated that native Japanese listeners discriminate /i/-/ɪ/ better in the context of voiced consonants and also that /i/ is identified more accurately.

Frieda and Nozawa (2007) examined native Japanese and Korean listeners' discrimination accuracy of 14 American English vowel pairs (/i/-/ɛ/, /i/-/ɪ/, /ɛ/-/e/, /ɛ/-/ɪ/, /ɛ/-/æ/, /æ/-/ɑ/, /æ/-/aɪ/, /ʊ/-/u/, /ʊ/-/ʌ/, /oʊ/-/u/, /oʊ/-/aʊ/, /ɑ/-/ʊ/, /ɑ/-/ʌ/, /ɑ/-/oʊ/) and found that listeners' linguistic experience exercises a strong influence on their discrimination accuracy for these vowel pairs. For instance, native Korean listeners' sensitivity to discriminate /i/-/ɪ/ pair (as measured by the A′ values in signal detection theory[3]) was significantly lower than that of native English listeners, whereas even inexperienced native Japanese listeners' A' scores for this vowel pair were not significantly lower than those of native English listeners. Native Korean listeners equated both /i/ and /ɪ/ with the Korean high front vowel /i/, while inexperienced Japanese listeners equated these two vowels to Japanese /iː/ and /eː/, respectively.

Strange, Akahane-Yamada, Kubo, Trent, Nishi and Jenkins (1998) examined perceptual assimilation of 11 non-rhotic American English vowels by native Japanese listeners in two different conditions, /hVbɑ/ disyllables in citation form and /hVb/ monosyllables in a carrier sentence, finding that American English long vowels were more likely to be labeled as Japanese long vowels when they were uttered and presented in a sentence. Based on their results, the authors concluded that no vowel pairs were Single Category type based on Best's model (Best 1995), and that in disyllable condition, eight vowel pairs [iː-ɪ, uː-ʊ, ɑː-ʌ, ɔː-ʌ, æː-ɛ, ɔː-oʊ, ɑː-ɔː] were Category Goodness or Categorizable/Noncategorizable type, and so they predicted that these pairs should be intermediate perceptual difficulty while other pairs were Two-Category type, and so the discrimination is expected to be easy. However, the authors did not perform a discrimination task to verify their prediction.

On the contrary, Nishi, Strange, Akahane-Yamada, Kubo and Trent-Brown (2008) tested the perceptual assimilation of Japanese vowels by native listeners of American English vowels and found that native English listeners predominantly equated Japanese short and long vowels to English long vowels /i, eɪ, ɑ (or ɔ), oʊ, u/. The authors explained that "under some stimulus and task conditions, the listeners may be able compare phonetically detailed aspects of non-native segments, while in others, they may resort to a phonological level of analysis in making cross-language similarity judgments" (p.587).

The studies discussed above, except that of Morrison (2002), examine the perception of non-native vowels in a single consonantal context. Bohn & Steinlen (2003) examined the perceptual assimilation of Southern British English by native Danish listeners in /hVt/, /dVt/, /gVk/ frames, and found that the perceptual assimilation of /ɪ, ɛ, ʊ, ʌ/ was

---

[3] See also Snodgrass, J. G., Levy-Berger, G. & Haydon, M. (1985).

strongly affected by consonantal context. Similarly, Strange, Akahane-Yamada, Kubo, Trent and Nishi (2001) investigated the perceptual assimilation of American English vowels by native Japanese listeners in /b-b, b-p, d-d, d-t, g-g, g-k/ syllabic context. Strange et al. (2001) found that American English vowels before a voiced consonant were more likely to be perceptually assimilated to two-mora Japanese vowels than those before a voiceless consonant, and they demonstrated that the spectral assimilation pattern varied based on the consonantal context. Strange et al. (2001) predicted that, as the perceptual assimilation patterns varied across consonantal context, the discrimination accuracy should vary accordingly, but the authors did not perform a discrimination experiment to verify their claim.

Nozawa & Wayland (2012) performed identification, discrimination, and perceptual assimilation experiments of American English /i, ɪ, ɛ, æ, ɑ, ʌ/ in /pVt/, /bVd/, /tVt/, /dVd/, /kVt/ and /gVd/ frames by native Japanese listeners. Nozawa and Wayland (2012) found that /i/-/ɪ/was better discriminated in the context of voiced consonants and that /i/ was more often equated with Japanese two-mora /iː/ in a voiced consonant context, while /ɪ/ was equated with one-mora /i/ in both voiceless and voiced consonant contexts. The results agree with the Morrison (2002) findings. Nozawa & Wayland also found that /æ/ was equated with /ja/ after a velar consonant, while in the other contexts, it was more often equated with /a/ or /aː/, and they found both that /æ/ was better identified after a velar consonant and that /æ/-/ɑ/ and /æ/-/ʌ/ were discriminated better in a velar consonant context.

All of these studies that take consonantal context into account consider the place of articulation of the proceeding consonant, and all the consonants in vowel perception studies are obstruents. To the best of my knowledge, no systematic studies have been done on the effects of the manner of articulation of surrounding consonants on the perception of non-native vowels. In this dissertation, the perception of American English vowels before /n/ and /l/ by native Japanese listeners is compared with that before /t/. With regard to allophonic differences of English vowels, Ladefoged and Johnson (2011) specify that “vowels are nasalized in syllables closed by a nasal consonant,” and, as for vowels before /l/, “you should be able to hear a noticeably different vowel quality before a velarized [ɫ]" (p.101).

In the following section, the characteristics of vowels before a nasal and /l/ (or [ɫ]) will be discussed.

1.8 Vowels before /n/ and /l/

1.8.1 Vowels before /n/

Olive, Greenwood and Coleman (1993) describe the characteristics of American English vowels before a nasal consonant as "in the F1 region of the vowel the nasal spectrograms show the nasal formant as a double formant for vowels with a high or mid F1, and as a broadening of F1 for vowels with low F1" Ladefoged (2003) refers to the same phenomenon: "the most obvious fact about nasalized vowels is that the first formant tends to disappear (p. 135)". Ladefoged (2001) describes acoustic characteristics of nasalized vowels as having wider first formant bandwidth (p. 165), meaning that the frequency of the first formant of non-nasalized vowels is better identified. Ladefoged (2001, 2003) does not use the term "nasal formant," but what he implies is that the first formant is obscured by the presence of additional resonance.

What causes this additional resonance is the opening of the nasal cavity. Johnson (2012) comments that the nasalized vowels have two resonant systems operating simultaneously. The air goes through the nasal cavity as well as the oral cavity. The simultaneous use of both airflows creates formants, all of which are present in a nasalized vowel.

Because the nasal formant appears in the F1 region, nasalization can affect the perception of vowel height. It is commonly acknowledged that in nasal vowel systems, the number of phonemic nasal vowels is smaller than or the same as that of oral vowel systems (Beddor 1993), which implies that it is difficult to maintain vowel distinctiveness in a nasal vowel system to the same degree as in an oral vowel system. Beddor (1993) surveyed 75 languages that exhibit allophonic or morphophonemic nasal vowel raising or lowering. American English is one of these languages, with the raising of prenasal /æ/. Beddor (1993) summarizes general effects of nasalization on allophonic and morphophonemic processes of vowels as high nasal vowels are lowered while low nasal vowels are raised. Beddor, Karakow and Goldstein (1986) demonstrated that the frequency of the first spectral peak in a nasal vowel is higher than in the corresponding oral vowel when the vowel is high (i.e. F1 is low) and lower than in the corresponding oral vowel when the vowel is low (i.e. F1 is high).

Previous studies demonstrated that a vowel preceding a nasal consonant is produced with considerably larger velic lowering than the one following a nasal consonant Krakow (1993). Thus, the /æ/ in "pan" is more strongly nasalized than in "nap."

As Beddor (1993) points out, raised /æ/ in a prenasal position is widely recognized as an allophone of the vowel in American English. Labov (2010) demonstrated that the F2 frequencies of /æ/ tokens in prenasal position uttered by a female speaker in Detroit, MI, are significantly higher than /æ/ tokens in other positions. Detroit is in an area where Northern Cities Shift is taking place, leading to /æ/ upward shifts. Thus, the difference in F2 frequencies between prenasal /æ/ and /æ/ is relatively small, but, nonetheless, statistical significance was observed. Labov (2010) observed that the difference is "maximized in speakers from New England, the Midland, and the West" (p.293).

1.8.2 Vowels before /l/

Vowels before /l/ are characterized by the continuously descending F2 (Olive, et al. 1993). Ladefoged and Johnson (2011) observe that the allophones of vowels before the postvocalic /l/ are retracted (p.101). The authors note that front vowels are retracted to the extent that the vowels sound like diphthongs. In a narrow transcription, "peel," "pail," and "pal" could be transcribed as [pʰiʊɫ, pʰeɪʊɫ, pdʰæʊɫ]. The retraction is realized acoustically as the descending F2. Back vowels are less affected by the syllable-final /l/ because these vowels have lower F2, but some contrasts tend to be obscured by the following /l/. Labov, Ash, and Boberg (2006) point out that, in some parts of the United States, /i/-/ɪ/, /ʊ/-/u/ contrasts are lost before /l/. Labov (1994) observes that the "syllable-final /l/ becomes a glide that is sometimes heard as a back rounded [o] or [u], in *goal*, *people*, etc., sometimes heard as [ə] in *call*, *sale*, and sometimes confused with nasality" (p. 275). The coda /l/ makes the preceding vowel less distinctive, but how the allophonic change affects the perception of American English vowels by native Japanese listeners has not yet been fully investigated.

2. The present study

2.1 Overview

Based on the findings of previous studies, the present study attempts to investigate the effects of the manner of articulation of the following consonant on the perception of American English vowels by native Japanese speakers. Most studies on nasalized vowels focus on the issues of perception by native listeners. Nasalized vowels are acoustically less distinctive, but, despite the reduced distinctiveness, native speakers of English seem to perceive English vowels as intended in prenasal context. Less is known about non-native speakers' perception of nasalized vowels. If non-native speakers rely on native vowel categories to identify and discriminate non-native vowels, are non-native vowels mapped into native categories differently in prenasal context than in preplosive context?

Even less is known about the perception of vowels before /l/ or velarized dark [ɫ]. Acoustic characteristics of vowels before [ɫ] are known, but, to the best of my knowledge, no study has been done on the perception of vowels in this context.

It is impractical to include all the vowels in this study. For the following reasons, six monophthongs /i, ɪ, ɛ, æ, ɑ, ʌ/ are chosen for this study. Diphthongs including /e/ and /o/ are predominantly equated with two Japanese vowel sequences (Frieda and Nozawa 2007). Thus, /e/, /o/, /aɪ/, /ɔɪ/, /aʊ/ are mapped into Japanese vowel categories as /ei/, /ou/, /ai/, /oi/ and /au/, respectively. These vowels are expected to be easy to identify for native Japanese speakers. In Frieda and Nozawa (2007), challenging vowel pairs for Japanese listeners were /ɛ/-/ɪ/, /æ/-/ɛ/, /ɑ/-/ʌ/, /ʊ/-/u/. Of these, /ʊ/-/u/ are excluded from this study largely because /ʊ/ is a rather rare vowel in English. It does not occur before a nasal consonant, so /ʊ/-/u/ pair is impossible in this context. In some varieties of American English, the distinction between /ɑ/ and /ɔ/ is lost (i.e. cot-caught merger). Ladefoged and Johnson (2011) note that many Midwestern and Californian speakers do not distinguish the contrast and so employ an intermediate vowel. Labov et al. (2006) attribute the merger to the unbalanced distribution of the two vowels. The /ɑ/ (or short-

o) occurs before all the consonants but /v/, and /ʒ/ but /ɔ/ (or open-*o*) has "a highly skewed distribution" (p. 58), and so the absence of /ɔ/ in which /ɑ/ can occur has obscured the contrast. According to the survey by Labov et al. (2006), the merger is more advanced in Canada, the American West, Eastern New England, and Western Pennsylvania, and the merger is more advanced in syllables closed by a nasal consonant. In Nozawa and Frieda (2002), native English listeners predominantly misidentified /ɔ/ tokens as /ɑ/. It was expected that native speakers may feel uncomfortable producing /ɑ/ and /ɔ/ distinctively, especially before /n/. The vowel /i/ is included because it has the highest F2 frequency, and so it may be most strongly affected by the following /l/; moreover, the F2 descends more drastically in the /i/ than in other vowels. Besides, if nasalization affects the perception of vowel height, the highest vowel has to be included. Thus, the following six vowels have been chosen for this study: /i/, /ɪ/, /ɛ/, /æ/, /ɑ/, /ʌ/.

This study consists of four major experiments. (1) Perceptual assimilation experiment: Japanese speakers identify American English vowels in terms of Japanese vowel categories, (2) Identification experiment: Japanese speakers identify American English vowels as intended, (3) Discrimination experiment: Japanese speakers discriminate six American English vowel pairs, and (4) Production experiment: Japanese speakers produce these six American English vowels.

All of the participants came for two sessions. In the first session, they completed the perceptual assimilation task after signing the consent form and answering a language background questionnaire. In the second session, odd numbered participants worked on the identification task first, and, after finishing the identification task, they moved on to the discrimination task. Even numbered participants, on the other hand, worked on the discrimination task first, and, after that, they worked on the identification task. All the participants took part in the production experiment after completing both the identification and discrimination tasks. Each session lasted no longer than 90 minutes.

2.2 Stimuli

The stimuli were extracted from utterances by four female native speakers of American English. Their utterances were digitally recorded at the sampling rate of 44.1kHz. in the psychology lab at Auburn University in Auburn, Alabama, USA.[4] Five talkers were recorded, but one of them, Talker 2, was the only male, so his utterances were not used as stimuli. The remaining four talkers were from New York (Talker 1), California (Talker 3), Wisconsin (Talker 4), and New York (Talker 5). They uttered 6 American English vowels /i, ɪ, ɛ, æ, ɑ, ʌ/ in /hVt/, /pVt/, /pVn/, /pVl/ frames. The talkers were given a word list and were asked to read aloud each word on the list in isolation. Prior to the recording, an instruction was given that the list contains some non-words. For instance, the talkers were told that "pul" (/pʌl/) is a syllable "pulse" without the final /s/, and it must be be differentiated from "pull." The recorded utterances were digitally edited on Cool Edit 2000[5].

The frequencies of the first two formants were measured at the midpoint of each token of each vowel, using Praat.[6] Tables 3-5 show the frequencies of F1 and F2, and show the vowel duration of each token in each frame.

Table 5

F1 and F2 frequencies of at the midpoint of /i, ɪ, ɛ, æ, a, ʌ/ uttered in /hVt/ frame by four talkers

	Talkers	F1	F2	duration (in msec)
/i/	T1	283	2975	143
	T3	355	2902	166
	T4	341	2908	84
	T5	310	2859	95
/ɪ/	T1	524	2747	107
	T3	561	2246	117
	T4	518	2360	97
	T5	486	2143	82
/ɛ/	T1	743	1866	99

[4] These stimuli were originally recorded for my previous project Grant-in-Aid for Scientific Research (C) 14510635.
[5] A software originally developed by Syntrillium Software
[6] A free software downloaded from http://www.fon.hum.uva.nl/praat/

	T3	779	2067	130
	T4	688	2103	80
	T5	608	210	69
/æ/	T1	978	1788	184
	T3	955	1730	146
	T4	973	1798	108
	T5	839	1880	110
/ɑ/	T1	1020	1375	179
	T3	1008	1522	118
	T4	932	1410	114
	T5	854	1357	99
/ʌ/	T1	765	1678	10
	T3	778	1699	130
	T4	780	1700	78
	T5	765	1792	74

Table 6

F1 and F2 frequencies of at the midpoint of /i, ɪ, ɛ, æ, a, ʌ/ uttered in /pVt/ frame by four talkers

	Talkers	F1	F2	duration (in msec)
/i/	T1	321	2786	117
	T3	352	3033	118
	T4	313	2848	88
	T5	344	2710	75
/ɪ/	T1	498	2229	102
	T3	511	2410	93
	T4	520	2280	79
	T5	461	2269	74
/ɛ/	T1	679	1788	114

	T3	745	2072	120
	T4	737	1860	97
	T5	579	1884	66
/æ/	T1	955	1891	159
	T3	936	1947	135
	T4	893	1888	138
	T5	743	1864	76
/ɑ/	T1	1098	1357	158
	T3	965	1517	139
	T4	875	1404	119
	T5	772	1407	98
/ʌ/	T1	707	1517	107
	T3	755	1697	111
	T4	714	1659	97
	T5	621	1757	81

Table 7

F1 and F2 frequencies of at the midpoint of /i, ɪ, ɛ, æ, a, ʌ/ uttered in /pVn/ frame by four talkers

	Talkers	F1	F2	duration (in msec)
/i/	T1	290	3287	211
	T3	370	2946	156
	T4	375	3001	104
	T5	418	2869	79
/ɪ/	T1	477	2539	171
	T3	636	2397	114
	T4	472	2334	61
	T5	550	2247	82
/ɛ/	T1	704	2030	175

	T3	719	2100	123
	T4	630	2024	97
	T5	569	2255	78
/æ/	T1	629	2364	291
	T3	748	2617	238
	T4	569	2373	145
	T5	633	2166	127
/ɑ/	T1	884	1240	263
	T3	722	1204	226
	T4	766	1288	167
	T5	700	1311	123
/ʌ/	T1	748	1424	141
	T3	773	1657	158
	T4	716	1440	107
	T5	643	1595	88

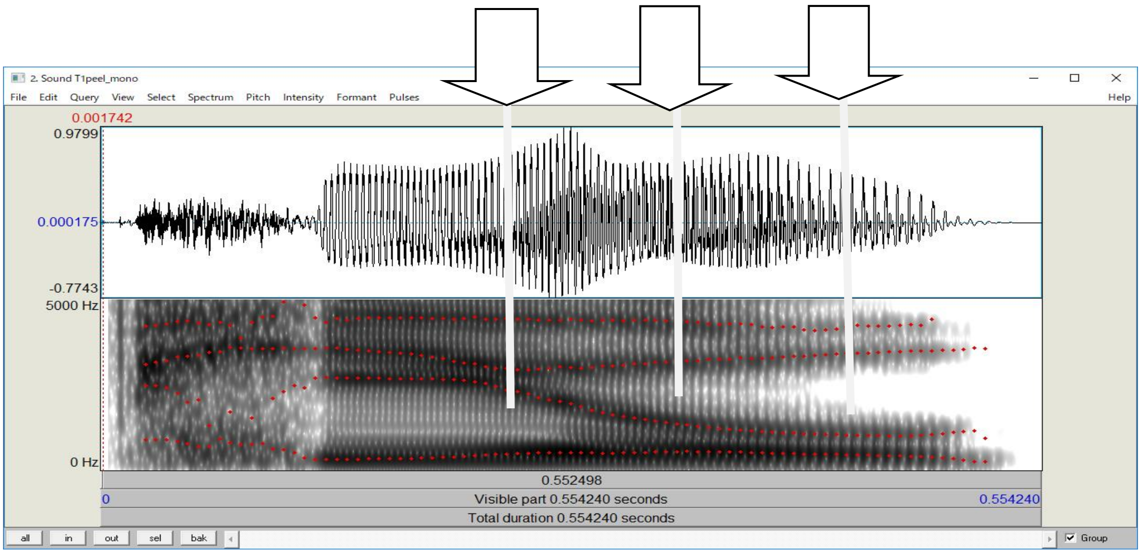

Figure 10. Three locations where F1 and F2 frequencies are measured (from left to right, 25%, 50%, 75%)

In /pVl/ frame, the boundary between the vowel and the following /l/ is difficult to determine, as is is the midpoint of the vowel. This is even more difficult in the utterances of native Japanese speakers, some of whom cannot produce [ɫ] authentically, and therefore F2 movement is small. Thus, frequencies of the first two formants were measured at 25%, 50% and 75% of /Vl/ continuum as shown in Figure 10. The measured F1, F2 frequencies and the duration of /Vl/ continua are shown in Table 8.

Table 8

F1 and F2 frequencies of at 25%, 50%, 75% of continua from /i, ɪ, ɛ, æ, a, ʌ/ to /l/ by four talkers

	Talkers	25%		50%		75%		
		F1	F2	F1	F2	F1	F2	duration (in msec)
/i/	T1	432	2639	518	1507	461	1047	386
	T3	386	2588	540	1422	516	1031	291
	T4	339	2640	475	2099	578	1414	209
	T5	467	2321	527	1869	546	1491	122
/ɪ/	T1	516	1804	530	1168	469	950	337
	T3	573	1670	591	1240	554	1066	266
	T4	596	1665	558	1224	517	1080	177
	T5	595	1473	575	1254	532	1134	108
/ɛ/	T1	735	1719	620	1204	478	981	394
	T3	693	1312	556	1133	530	1095	269
	T4	738	1530	690	1319	588	1148	151
	T5	669	1434	645	1315	573	1190	113
/æ/	T1	906	1827	780	1363	591	1018	418
	T3	770	1118	713	1015	663	1039	262
	T4	785	1966	825	1521	623	1108	208
	T5	796	1564	767	1419	701	1278	125
/ɑ/	T1	825	1254	776	1124	607	1061	369
	T3	704	1054	677	1075	584	1055	306

	T4	886	1201	843	1139	684	1081	168
	T5	844	1242	799	1188	707	1144	168
/ʌ/	T1	661	1086	547	1009	496	1013	369
	T3	537	962	520	961	493	956	274
	T4	761	1205	716	1110	592	1043	143
	T5	643	1075	597	997	526	947	143

Figures 11-13 show the frequencies of the first two formants at the midpoint. In each Figure, vowels are differentiated by color: light blue /i/, orange /ɪ/, gray /ɛ/, yellow /æ/, dark blue /ɑ/, green /ʌ/. The four talkers are differentiated by the marker, ■ Talker 1, ● Talker 3, ▲ Talker 4, * Talker 5.

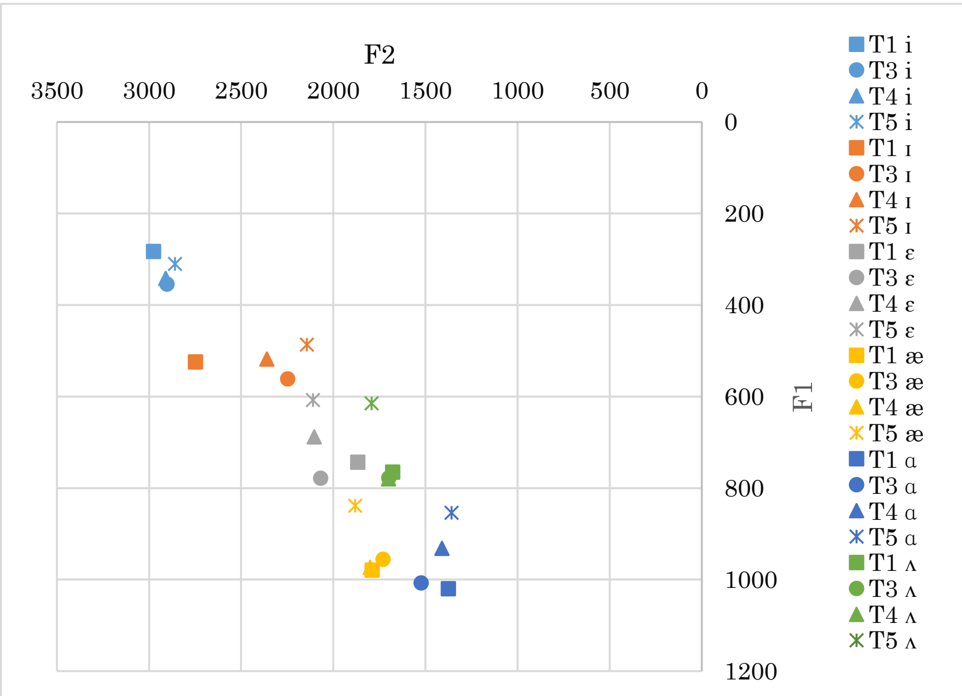

Figure 11. F1, F2 frequencies of six vowels uttered by four talkers in /hVt/ frame

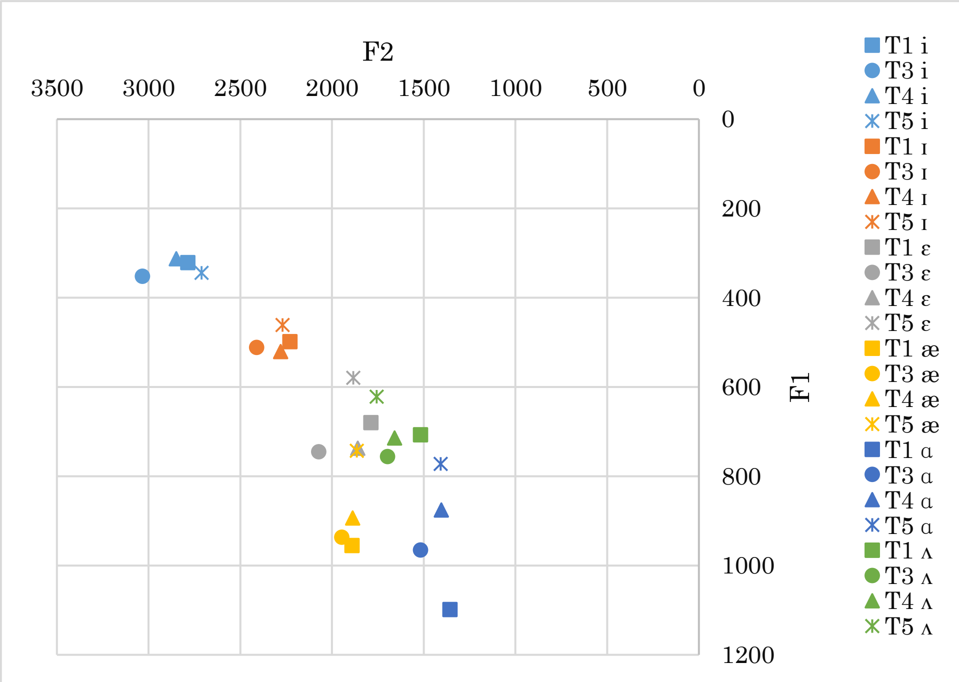

Figure 12. F1, F2 frequencies of six vowels uttered by four talkers in /pVt/ frame

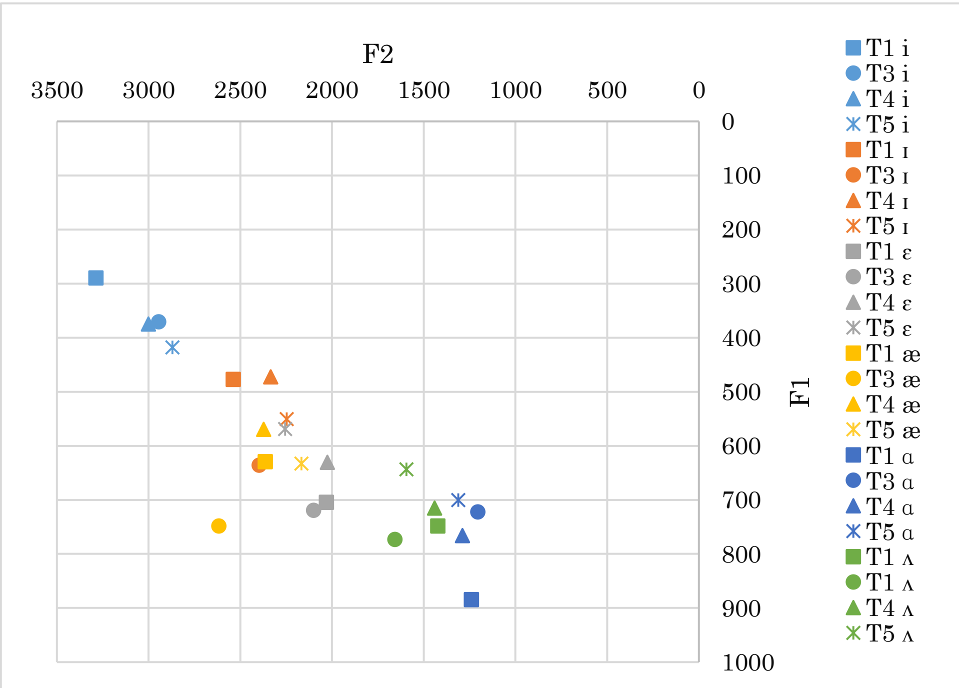

Figure 13. F1, F2 frequencies of six vowels uttered by four talkers in /pVn/ frame

Even though individual differences are observed, in preplosive condition, the relative position of each vowel in the vowel space is more or less similar to the findings of

Bradlow (1993) and Yavas (2003). Talker 5 seems to have a vertically smaller vowel space. Her low vowels are higher than those of the other talkers. For instance, her /æ/ token in /pVt/ context has almost the same formant frequencies as Talker 4's /ɛ/ token. Also in /pVt/ context, Talker 5's /ʌ/ is almost as high as other talkers' /ɑ/ tokens. Talker 1's /ɛ/ tokens are a little backer than those of other talkers. This is especially true in /pVt/ context. Talker 1's /ɛ/ is in a region of other talkers /ʌ/.

In prenasal /pVn/ context, the position of vowels is a little different. The vowel /ɑ/ is higher in this context. Two of /ɪ/ tokens are lower than in two preplosive contexts. The most prominent difference between /pVn/ and preplosive contexts is fronted and raised /æ/. The /æ/ tokens are higher and more fronted than /æ/ in this context, as pointed out by Beddor (1993), Lobov (1994, 2010).

In /pVl/ context, as mentioned earlier, F1 and F2 frequencies were measured at 3 locations. F1 and F2 frequencies measured at 25%, 50 % and 75% of the /Vl/ continua are respectively plotted in Figures 14-16. F2 frequencies are lower than in preplosive conditions already at the first 25%, meaning the coarticulatory effects of the coda /l/ extend to the whole vowel. At 25%, /ɛ/ is especially retracted, but examining the change throughout the continua reveals that /i/ tokens undergo the largest decrease in F2 frequencies because the vowel has intrinsically high F2. The F2 continues descending, and the F1 of low vowels ascends, as at 75%, vowels are hardly distinguishable. Except two high front vowels, vowels uttered by Talker 3 are more retracted than those uttered by the other native speakers. This is especially true with Talker 3's /ɛ/ and /æ/. F2 frequencies are lower than 1500Hz, and these are not typical of front vowels.

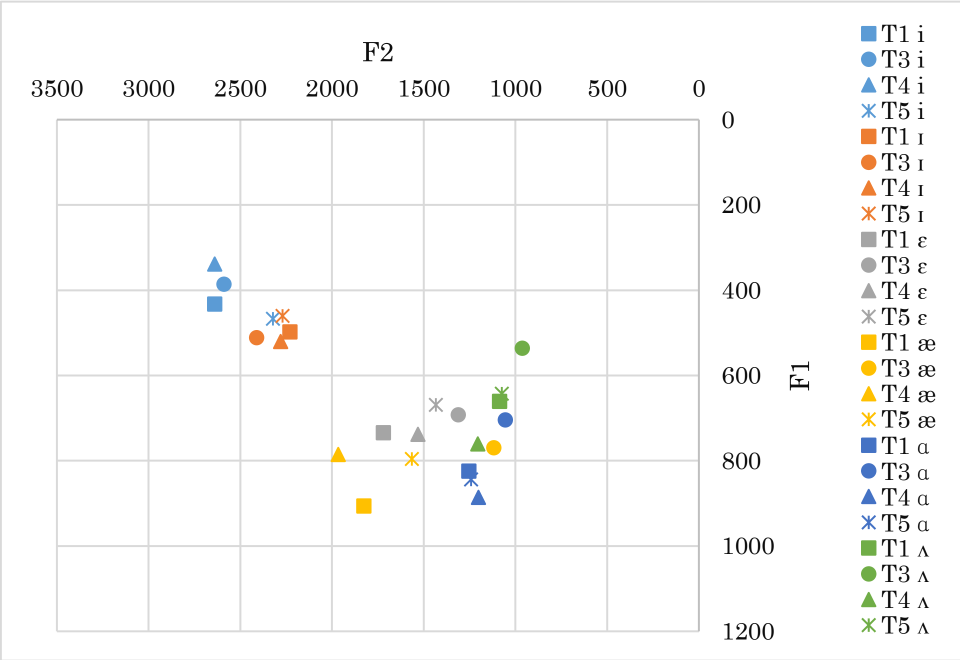

Figure 14. F1, F2 frequencies measured at the first 25% of /pVl/ continua

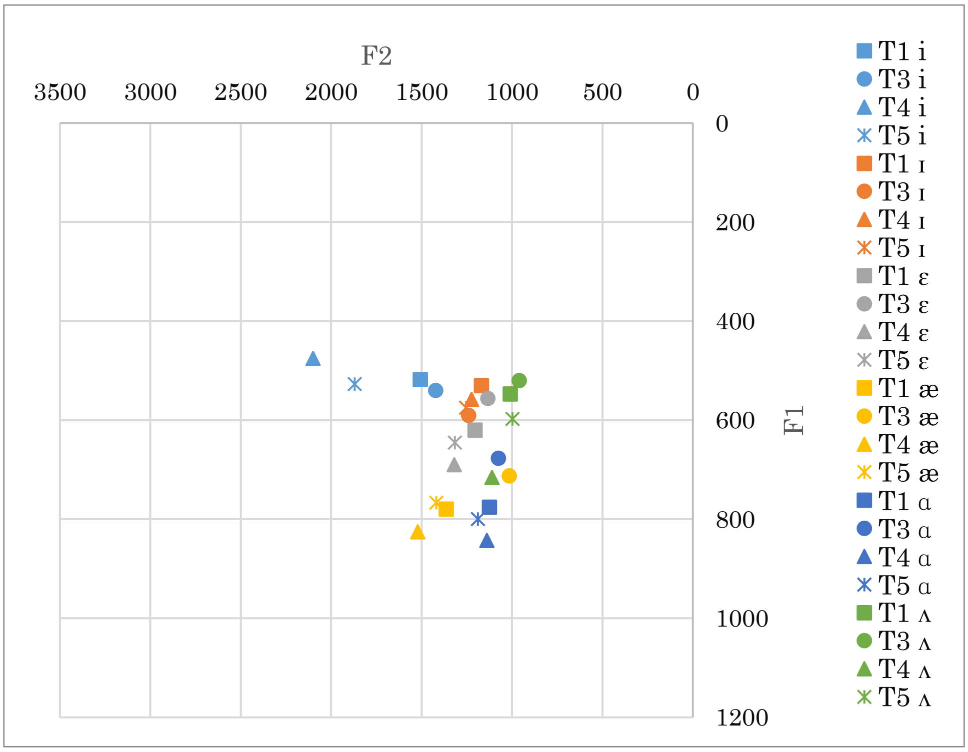

Figure 15. F1, F2 frequencies measured at the first 50% of /pVl/ continua

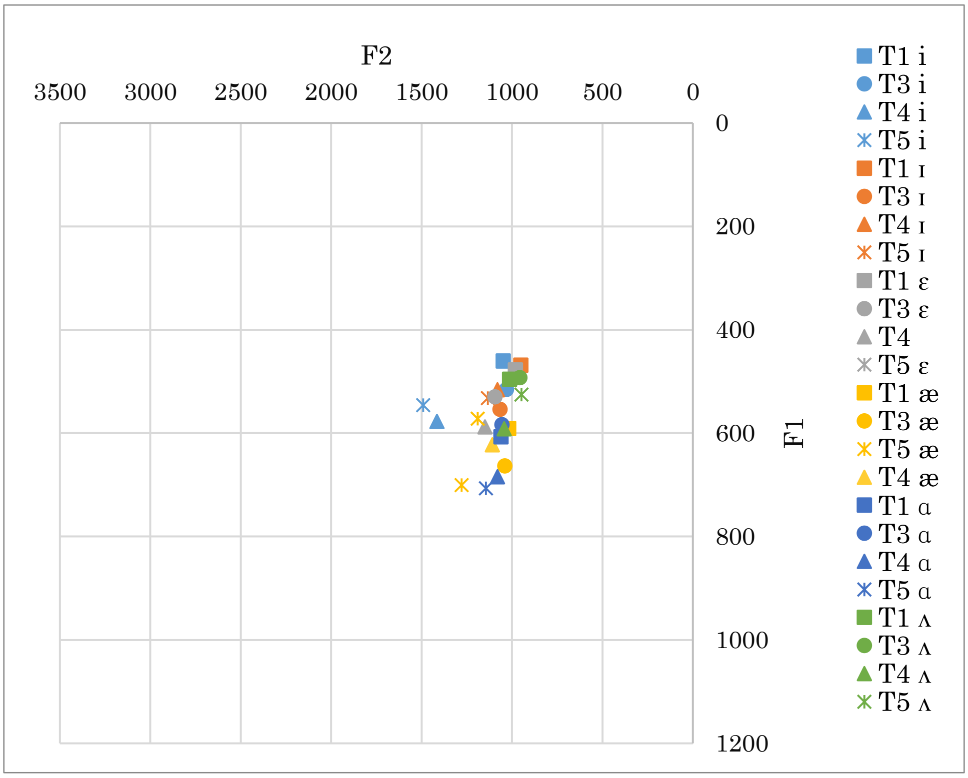

Figure 16. F1, F2 frequencies measured at the first 75% of /pVl/ continua

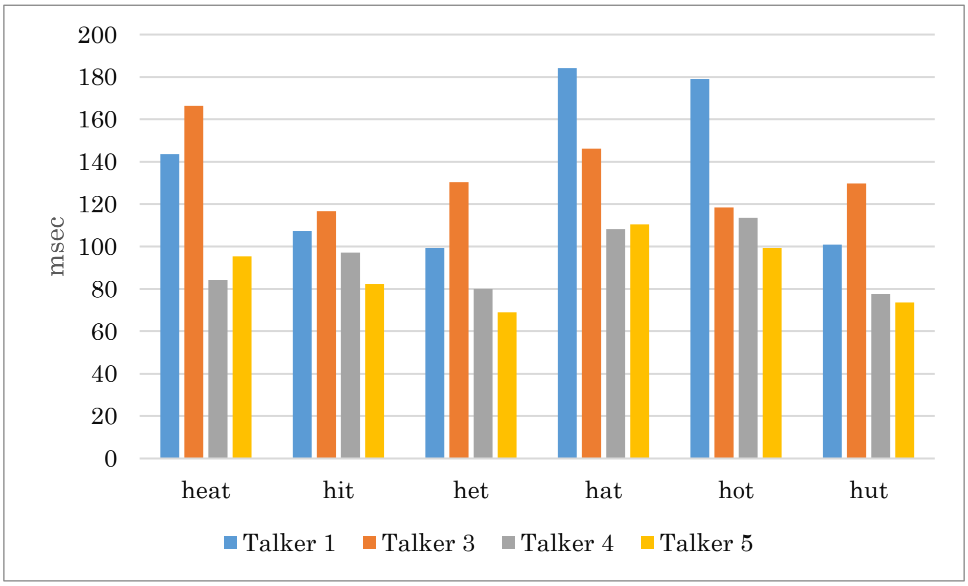

Figure 17. Duration of six vowels uttered by four talkers in /hVt/ context

Figure 17 shows the duration of six vowels in /hVt/ context. A large individual difference is observed. Generally, the utterances of Talkers 1 and 3 are longer than those of the other talkers. The /i/ tokens of Talkers 1 and 3 are longer than their /ɪ/ tokens, but, as for Talkers 4 and 5, there appears to be no significant difference in duration between their /i/ and /ɪ/ tokens. Phonologically short vowels /æ/ and /ɑ/ are both longer than /i/

In the /pVt/ context, durational difference between /i/ and /ɪ/ tokens is smaller than in /hVt/ context, and the /i/ tokens of Talkers 1 and 3 are shorter than in /hVt/ context. As in /hVt/ context, /æ/ and /ɑ/ longer than /i/. Talker 5's tokens are the shortest of all of the six vowels.

In the /pVn/ context, vowels are generally longer than in the two preplosive contexts. Ladefoged and Johnson (2011) state that vowels uttered before a sonorant are longer than those uttered before an obstruent, and vowels uttered before a voiced consonant are longer than those uttered before a voiceless consonant. The measured vowel duration here supports Ladefoged and Johnson's claim. As in /hVt/ and /pVt/ contexts, the tokens of Talkers 1 and 3 are longer, and Talker 5's tokens are the shortest, with the exception of /ɪ/.

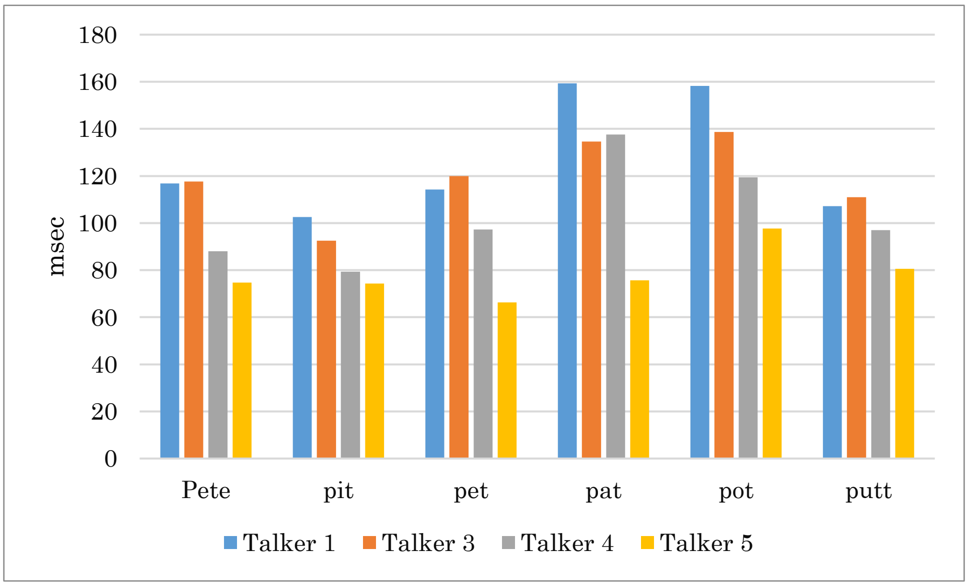

Figure 18. Duration of six vowels uttered by four talkers in /pVt/ context

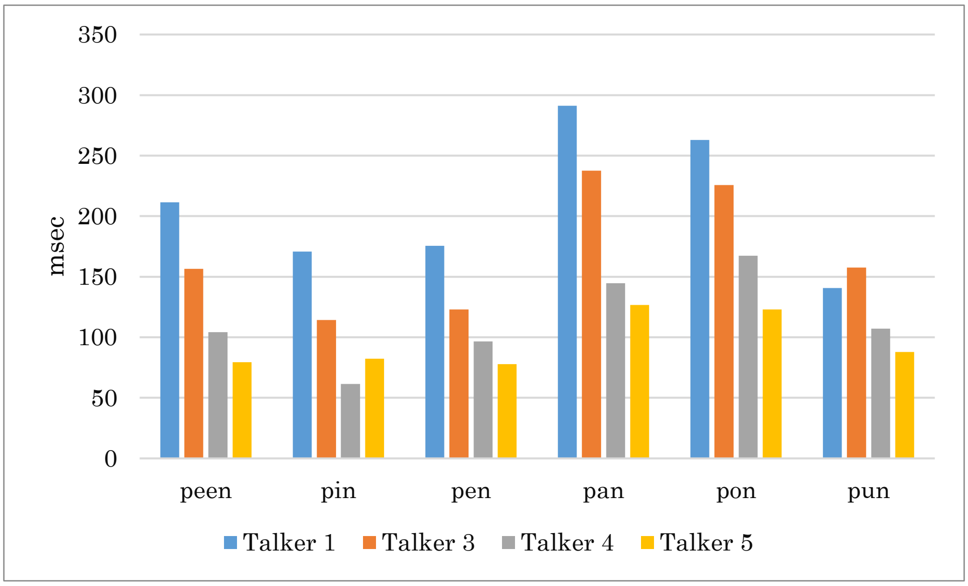

Figure 19. Duration of six vowels uttered by four talkers in /pVn/ context

In /pVl/ context, the duration of /Vl/ sequence is measured. Generally, Talker 1's tokens are the longest, followed by Talker 3, and Talker 4, and Talker 5's tokens are generally shortest. The difference in duration between talkers is larger than the other three contexts. The tokens of Talker 1 are about three times longer than those of Talker 5.

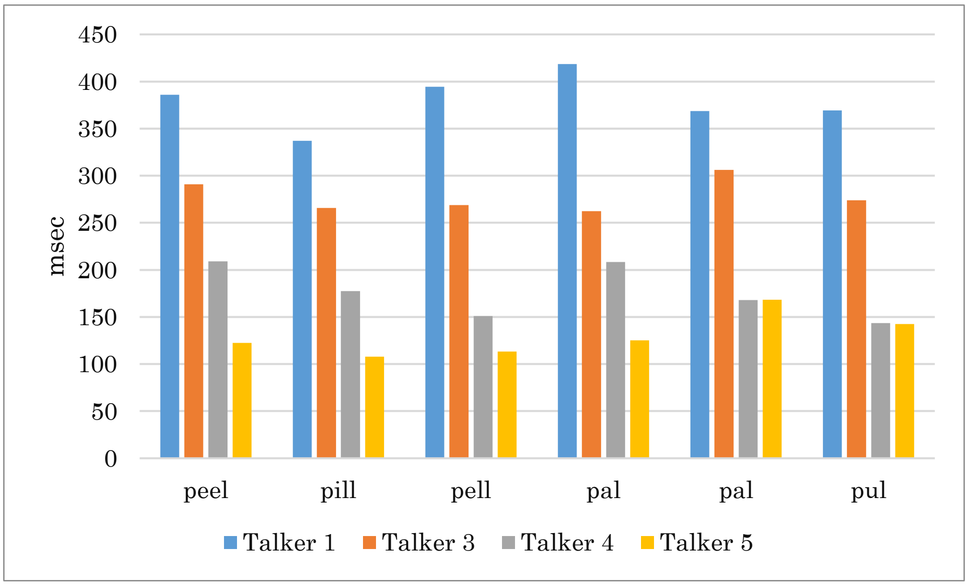

Figure 20. Duration of six /Vl/ sequences uttered by four talkers.

2.3 Participants

Participants answered the language background questionnaire, and their answers are summarized in Table 9. Of the 34 participants, J27 withdrew from the session, and her data were excluded from the analysis. The age noted in Table 9 is the age of each participant when that participant took part in this project (mean=19.4). Each rated his/her own English proficiency on a 10-point scale (1=poor, 10=very good). Table 9 also shows each participant's TOEIC ® L & R Test score, which ranges from 300 to 990. Four of the participants, who were first-year students at a university when they took part in this project, had not yet taken the test. Ten of them had lived abroad. Of these, J4 and J6 were in the United States on university study abroad programs. J14 and J29 each spent 1 year in Australia and Canada, respectively, as part of high school study abroad programs. J32 also spent 3 months in Canada on her high school's study abroad programs. J3, J12, J13, J25 and J34 spent their childhoods abroad; J3 spent 9 years in Thailand and 1 year in New Zealand. J12 spent 11 years in Italy. J13 spent 9 months in Germany and in Italy. J25 was born in Malaysia and spent 2 years there; 8 years in Singapore; and 3 years in Czech Republic. J34 spent 1 year in Brazil. Seven participants reported that they have some proficiency in another language than English. They rated their proficiency in the languages that they can speak on a 10-point scale (1=poor, 10=good). In general, their proficiency rating of non-English proficiency is lower than that of English. The only exception is J7, who rated her English proficiency as 5 and her Chinese proficiency as 7. As for the participants' parents' language background, most of them reported that both of their parents are native speakers of Japanese. J7's parents are from China, and they are both native speakers of Chinese. J11's biological mother is a native speaker of Japanese, but her step-mother is a French speaker. J25's mother is a native speaker of Chinese. The participants' linguistic experience varied widely, and so it was expected that their performance in perception and production experiment would also vary.

Table 9

Participants' language background

#	Gender	DOB (MM/DD/YY)	Age	Birthplace	English Proficiency	TOEIC	TOEFL
J1	female	09/04/96	19	Kyoto		645	477
J2	male	10/27/96	19	Kyoto	6	655	
J3	female	04/27/95	21	Osaka	6	865	
J4	female	12/26/94	21	Kyoto	9	840	iBT78
J5	female	11/06/96	19	Nara	5	630	505
J6	female	06/02/94	22	Shiga	5	625	480
J7	female	05/23/97	19	Gunma	5	650	
J8	female	06/14/94	22	Osaka	6	605	
J9	female	04/08/95	21	Nara	5	600	
J10	female	02/18/98	18	Fukuoka	5	605	
J11	female	03/20/98	18	Aichi	5	500	463
J12	female	03/21/98	18	Kyoto	5	400	
J13	male	09/02/96	19	Osaka	5	645	495
J14	female	12/18/95	20	Kyoto	5	735	
J15	female	06/27/95	21	Kagawa	1	650	
J16	female	02/09/97	19	Hyogo	6	620	
J17	female	08/18/97	19	Kyoto	3	595	420
J18	female	09/12/96	20	Yamaguchi	3	450	
J19	female	12/09/98	18	Kyoto	4	715	493
J20	female	12/31/95	20	Nara	6	630	
J21	female	12/02/97	18	Nara	4	535	
J22	male	11/11/97	19	Hyogo	4	500	
J23	male	09/11/97	19	Shiga	4	500	
J24	male	08/06/97	19	Osaka	2	300	
J25	female	04/15/96	21	Malaysia	10	990	663

J26	female	08/01/96	20	Hyogo	6	765	
J27	~~female~~	Withdrawn			~~6~~	~~630~~	
J28	female	07/25/96	20	Fukui	6	695	470
J29	female	04/07/98	19	Aichi	5	645	
J30	female	04/06/98	19	Mie	2		
J31	female	12/07/98	18	Hyogo	5		
J32	female	03/20/99	18	Shiga	6		480
J33	female	03/26/99	18	Shizuoka	4		
J34	female	12/22/97	19	Nara	5	555	

Table 9

Participants' language background (continued)

#	Other language	Living abroad	Mother's L1	Father's L1
J1	no		Japanese	Japanese
J2	no		Japanese	Japanese
J3	Yes (Thai=2)	9 yrs in Thailand, 1 yr in NZ	Japanese	Japanese
J4	no		Japanese	Japanese
J5	no		Japanese	Japanese
J6	no		Japanese	Japanese
J7	Yes (Chinese=7)		Chinese	Japanese
J8	no		Japanese	Japanese
J9	no		Japanese	Japanese
J10	no		Japanese	Japanese
J11	Yes (French=3)		French	Japanese
J12	Yes (Italian=3)	11yrs in Italy	Japanese	Japanese

J13	no	9 mos in Germany 9 mos in Italy	Japanese	Japanese
J14	no	1 yr in Australia, 6 mos in Hungary	Japanese	Japanese
J15	no		Japanese	Japanese
J16	no		Japanese	Japanese
J17	no		Japanese	Japanese
J18	no		Japanese	Japanese
J19	no		Japanese	Japanese
J20	no		Japanese	Japanese
J21	no		Japanese	Japanese
J22	no		Japanese	Japanese
J23	no		Japanese	Japanese
J24	no		Japanese	Japanese
J25	Yes (Chinese=6)	2 yrs in Malaysia 2yrs,8 yrs in Sigapore、3 yrs in Czech	Chinese	Japanese
J26	no		Japanese	Japanese
J27	~~no~~	Withdrawn		
J28	no		Japanese	Japanese
J29	Yes (Spanish=1)	1 yr in Canada	Japanese	Japanese
J30	no		Japanese	Japanese
J31	no		Japanese	Japanese
J32	no	3 mos in Canada	Japanese	Japanese
J33	no		Japanese	Japanese
J34	Yes (Portugese=1)	1 yr in Brazil	Japanese	Japanese

3. Perceptual assimilation experiment

3.1 Procedure

All the perception experiments were carried out on UAB software.[7] Participants took part in experiments individually. A participant heard one stimulus per trial over a headset, chose a Japanese vowel that best represents the stimulus from the choices on the answer sheets, and rated their category goodness in 7-point scale (1=poor, 7=good). The choices were written in katakana on the answer sheet. The choices were in /hVt/ context, "ha, ha:, hi, hi:, hu, hu:, he, he:, ho, ho:, hua, hea, hei, hia, hie, hoa, hou, hya, hyu, hyo." In the contexts in which the stimuli begin with /p/, the participants' choices begin with "p" instead of "h": "pa, pa:, pi, pi:, pu, pu:, pe, pe:, po, po:, pua, pea, pei, pia, pie, poa, pou, pya, pyu, pyo." The instruction was given prior to the experiment that the participant's job was to circle the Japanese syllable closest to the stimulus they heard, and rate degree of similarity of the stimulus to Japanese on a 7-point scale. In each case, the participant was told to disregard the lexical meaning of each stimulus and to disregard the final consonant /t/, /n/, and /l/. In each frame, 24 stimuli (6 vowels × 4 talkers) were prepared, and each stimulus was played twice. Thus, 48 trials were prepared in each context. The order of the contexts was counterbalanced across participants. Participants were allowed to hear the same stimulus as many times as they wanted. The experimenter held the mouse in each case, and after confirming that the participant completed the answer, the experimenter clicked on the word "NEXT" on the computer screen to move on to the next trial.

3. 2 Results

The most frequent responses (R1) and the second most frequent responses (R2) are shown in Table 10. High front vowels /i/ and /ɪ/ are generally equated with Japanese high front vowel /i/ (*hi* and *pi*) and /i:/ (*hii* and *pii*). The vowel /i/ is classified as two-mora

---

[7] Software program developed at the Department of Biocommunication, University of Alabama at Birmingham. The program is not commercially available.

/i:/ in approximately 50% of instances in /hVt/ and /pVt/ contexts, but before /n/, it is more likely to be equated with one-mora /i/, and only in 34.9% of instances was it equated with /i:/, despite the fact that vowel duration is generally longer before /n/ than before /t/ (Compare Figures 17-19). And before /l/, /i:/ is not included in the two Japanese vowels with which /i/ is most frequently equated. This may be attributed to the retracting effect of the following /l/. F2 keeps descending as shown in Figure 10, and because of this, /i/ is less likely to be heard as two-mora /i:/ in this context. In the context of /pVl/, the mean category goodness rating of /i/ as Japanese /i/ is lower (3.9) than in the other contexts. The vowel /ɪ/ is equated with one-mora /i/ most frequently in all of the four contexts, but in less than 50% of instances before /n/ and /l/. In /hVt/ and /pVt/ contexts, /ɪ/ is classified as /i/ in about 70% of instances. Even though both /i/ and /ɪ/ are equated with Japanese /i/, category goodness rating of /ɪ/ is lower in all the contexts, so participants may have discerned spectral or qualitative differences between the two English vowels. In the context of /pVn/, it is equated with /e/ more frequently than in other contexts. Thus, /ɪ/ may be perceived lower in this context.

The vowel /ɛ/ is most frequently identified as instances of Japanese /e/ in all the contexts. This vowel may be least affected by the effect of the following consonant, but in /pVl/ context the mean category goodness rating of /ɛ/ as Japanese /e/ is lower (3.1) than in the other context. This is even lower than the mean category goodness rating of /æ/ as the Japanese /e/ in the same context (3.7). The perceptual assimilation pattern of /æ/ is clearly different, depending on whether or not the following consonant is a plosive /t/. In /hVt/ and /pVt/ contexts, the vowel is equated with the Japanese low vowel /a/ or /a:/, but it is more frequently equated with /e/ in /pVn/ and /pVl/ contexts.

Vowels /ɑ/ and /ʌ/ show similar perceptual assimilation patterns. In /pVt/ context, /ɑ/ and /ʌ/ are identified as /a/ and /u/, respectively, but in the other contexts, these two vowels are equated with the same or similar Japanese vowel categories. In /hVt/ contexts, these vowels are both identified as /a/, and in the /pVn/ context, /ʌ/ is most frequently identified as /a/, but it is also identified as /o/ in 26.5% of instances, and /ɑ/ is also identified as /o/ in 24.2% of instances in the same context. In /pVl/ context, these two vowels are both equated with /o/, /ou/, or /o:/.

Table 10.

Results of perceptual assimilation experiment in percent

	i		ɪ		ɛ	
	R1	R2	R1	R2	R1	R2
hVt	hii 49.6%	hi 46.6%	hi 75.0%	hii 13.6%	he 60.6%	hi 14.2%
	4.8	4.5	3.8	3.3	3.8	3.3
pVt	pi 53.0%	pii 42.0%	pi 77.2%	pe 6.8%	pe 50.4%	pu 9.8%
	5.2	4.7	3.9	4.2	4	3.2
pVn	pi 62.5%	pii 34.9%	pi 40.5%	pe 33.7%	pe 62.5%	pi 10.2%
	4.5	4.5	3.5	3.9	4.2	2.7
pVl	pi 37.1%	pyu19.7%	pi 39.4%	pii 8.3%	pe 42.8%	pea 9.1%
	3.9	4.2	3.5	3.7	3.1	1.7

	æ		ɑ		ʌ	
	R1	R2	R1	R2	R1	R2
hVt	ha 45.5%	haa 27.7%	ha 42.4%	haa 26.1%	ha 39.0%	ho 23.1%
	3.3	3.7	3.9	4.1	3.6	3.7
pVt	pa 31.0%	paa15.5%	pa 41.7%	paa 14.8%	pu 40.8%	pa 20.8%
	3.8	3.5	4.7	4.6	3.8	3.5
pVn	pe 33.7%	pea 16.'%	po 24.2%	pa 23.5%	pa 47.4%	po 26.5%
	3.6	3.2	4.5	4.1	4.5	4.1
pVl	pe 26.1%	pa 16.3%	pou 29.5%	po 17.4%	po 34.5%	poo 27.7%
	3.7	3.3	3.6	3.8	4.1	4.4

Note. The most frequent response (R1) and the second most frequent response （R2） are shown. The numbers in lower stand show mean category goodness rating.

Based on the results of this perceptual assimilation experiment, the six American English vowels can be classified into three groups.

American English vowels	matching Japanese vowels
/i/, /ɪ/	/i/, /ii/
/ɛ/	/e/
/æ/	/a/, /e/
/ɑ/, /ʌ/	/a/, /o/

PAM (Best 1995) classifies non-native phone pairs into six different types, based on whether or not a non-native phone is categorized in a learner's native categories and whether or not two non-native phones of a pair fall into one single native category. But here, it follows that American English vowels are rarely categorized because, in most cases, American English vowels are not labeled as exemplars of one Japanese vowel category in more than 50% of instances, at least when the labeling is averaged across four talkers. According to PAM, discriminability is difficult to predict when a non-native phone is not categorized. So here I would like to withhold labeling each vowel pair in each context as to what type they belong to.

However, it can be predicted that /i/-/ɪ/ would be discriminated more accurately when /i/ is labeled as two-mora /iː/ than when it is classified as one-mora /i/. It can also be predicted that /æ/ would be discriminated better against /ɑ/ and /ʌ/, in /pVn/ and /pVl/ contexts, in which /æ/ is more frequently labeled as /e/ than as /a/. Moreover, /ɑ/-/ʌ/ would be difficult to discriminate in all the contexts, but discrimination accuracy may be a little higher in the /pVt/ context in which /ʌ/ is labeled as /u/.

What prediction can be made as to identification? If Japanese listeners expect English vowels to sound like Japanese transcription of English vowels, then /i/ may be identified better in contexts in which it is labeled as two-mora /iː/. Similarly, /ɪ/ may be identified better in /hVt/ and /pVt/ contexts, in which /ɪ/ is equated with Japanese /i/ in about 70% of instances, than in /pVn/ and /pVl/ contexts, in which /ɪ/ is equated with /i/ in only about 40% of instances. The /ɛ/ may identified better in /hVt/ and /pVn/ contexts in which the vowel is identified as an exemplar of /e/ in about 60% of instances, and it may be least accurately identified in the context of /pVl/ in which /ɛ/ is equated with /e/ in about 42.8% of instances. The identification accuracy of /æ/, /ɑ/, and /ʌ/ is a bit difficult to predict from the results of the perceptual assimilation experiment because, for example, all of these three vowels are labeled as /a/ or /aa/, and so these three vowels may be mutually confusable for Japanese listeners, but if Japanese listeners assume /ɑ/ to sound like Japanese /o/ as the common transcription of the vowel, then Japanese listeners may not choose /ɑ/ if the vowel sounds like Japanese /a/. The /æ/ and /ʌ/ are both transcribed as /a/ in Japanese, so, if the stimulus sounds like /a/, Japanese listeners may not be able to tell which is correct. If a stimulus sounds like Japanese /o/, Japanese listeners would choose /ɑ/. In /pVn/ and /pVl/ contexts, /æ/ is more frequently equated with /e/ than with /a/, which probably makes the confusion of the vowel with /ɑ/ and /ʌ/ less likely, but it may instead lead to the misidentification of /æ/ with /ɛ/.

4. Identification experiment

4.1 Procedure

A participant heard one stimulus per trial. Six choices were given, and the choices were spelled out as shown in Table 11, rather than shown in phonetic symbols, because participants were not familiar with phonetic symbols. In all the contexts or frames, the choices were aligned in the order /i, ɪ, ɛ, æ, ɑ, ʌ/. Each participant was told what vowel (or syllable) each choice represented. Participants were told, for example, that the leftmost choice represented /i/ even though they are spelled differently. Thus, they were told “Pete” and “peen” rhyme with “beat” “bean” respectively. Among the stimuli were unfamiliar words like “putt.” Thus, each participant was told that “putt” is different from “put,” and it rhymes with “but” and “cut.” Also among the stimuli were non-words like “het” and “pul.” A participant was told that “het” was a non-word that sounds like “head,” but that it ends with “t” instead of “d.” A participant was also told that “pul” was a non-word that sounds like “pulse” without the final “s” and that it was different from “pull.”

Table 11.

Six choices in each consonantal context in Identification Experiment

	i	ɪ	ɛ	æ	ɑ	ʌ
hvt	heat	hit	het	hat	hot	hut
pVt	Pete	pit	pet	pat	pot	putt
pVn	peen	pin	pen	pan	pon	pun
pVl	peel	pill	pell	pal	pol	pul

Each participant heard one stimulus per trial. Twenty-four stimuli (6 vowels × 4 talkers) were prepared for each consonantal context, and each participant heard each stimulus twice in different order. Thus, 48 trials were prepared in each consonantal context. Participants were allowed to take a break after the first 24 trials. Stimuli were played on a computer, and participants heard stimuli over headsets. Participants responded by moving a cursor to a box on the computer screen and clicking on the box. If a participant waited more than 10 seconds before providing a response, he/she was asked if he/she would like to hear the stimulus again. The inter-trial-interval (ITI) was 1,00ms, so that, for example, 1,000ms after a participant gave a response, the next stimulus was played. Thus, 48 trials (6 vowels × 4 trials × 2 times) were created in each consonantal context.

The order of the consonantal context was counterbalanced across participants. Each participant worked on 10 practice trials to get familiar with the task.

4. 2 Results

The percentages of correct responses and confusion matrices are shown in Tables 11-14.

Table 12.

The percentages of correct responses and confusion matrix in /hVt/ context (Correct responses are in bold.)

	Responses					
heard	i	ɪ	ɛ	æ	ɑ	ʌ
i	**56.8**	40.5	2.3	0	0	0.4
ɪ	7.2	**87.9**	4.5	0	0	0.4
ɛ	0.4	11.0	**75.4**	4.2	2.7	6.4
æ	2.3	0.4	6.8	**71.6**	7.6	11.4
ɑ	1.9	0.4	0.8	35.2	**43.2**	18.6
ʌ	1.5	0	0.8	25.8	38.3	**33.7**

Table 13.

The percentages of correct responses and confusion matrix in /pVt/ context (Correct responses are in bold.)

	Responses					
heard	i	ɪ	ɛ	æ	ɑ	ʌ
i	**47.3**	50.4	2.3	0	0	0
ɪ	6.1	**81.1**	10.6	0	0.4	1.9
ɛ	1.1	4.5	**70.1**	9.1	3.4	11.7
æ	1.5	1.1	20.1	**62.9**	3.8	10.6
ɑ	0	0.38	0	31.8	**56.8**	11.0
ʌ	0	0.4	0.8	14.8	37.1	**47.0**

Table 14.

The percentages of correct responses and confusion matrix in /pVn/ context (Correct responses are in bold.)

	Responses					
heard	i	ɪ	ɛ	æ	ɑ	ʌ
i	**21.6**	75.4	1.5	0.4	0.4	0.76
ɪ	18.6	**46.6**	30.3	3.0	0.8	0.8
ɛ	3.0	15.2	**64.8**	11.4	1.1	4.5
æ	5.3	7.6	43.6	**36.7**	1.1	5.7
ɑ	0	0	0.8	23.5	**59.0**	16.7
ʌ	0	0.4	0.8	28.0	42.4	**28.4**

Table 15.

The percentages of correct responses and confusion matrix in /pVl/ context (Correct responses are in bold.)

	responses					
heard	i	ɪ	ɛ	æ	ɑ	ʌ
i	**36.0**	59.1	3.0	0	0	1.9
ɪ	9.5	**69.7**	18.9	0.8	0	1.1
ɛ	3.0	9.1	**76.5**	4.9	1.1	5.3
æ	1.5	1.9	43.9	**36.0**	4.9	11.7
ɑ	0	0.8	9.8	23.1	**42.8**	23.5
ʌ	0	0	1.5	12.5	62.5	**23.5**

The results were submitted to a repeated-measures ANOVA with 4 Contexts and 6 Vowels as within-subject variables. The results revealed that over vowels are better identified in /hVt/ and /pVt/ contexts than in /pVn/ and /pVl/ contexts (p<.001). The results also revealed that the main effect of Context [F(3, 96)=47.46, p<.001] and Vowels [F(5, 160)=40.80, p<.001] are both significant, and the two-way interaction between Context × Vowels is also significant [F(15, 480=11.38, p<.001]. Bonferroni-adjusted post-hoc pair-wise comparisons revealed that the identification of all the vowels but /ɛ/ was affected by the context. The vowels /i/, /ɪ/, and /æ/ were all significantly less accurately identified in /pVn/ and /pVl/ contexts than in /hVt/ and /pVt/ contexts (p<.001). Furthermore, the identification of /ɪ/ was significantly less accurate in the context of /pVn/ than in the context of /pVl/ (p=.024). On the contrary, /ɑ/ was significantly less accurately identified in /hVt/ context than in /pVt/ (p=.032) and /pVn/ (p=.039) contexts. /ʌ/ was more accurately identified in /pVt/ context than in the other three contexts (p<.05).

The results of the identification experiment seem to agree with the results of the perceptual assimilation experiment. The vowel /i/ is most frequently equated with two-mora /iː/ in /hVt/ context, and /i/ is most accurately identified in the same context. /ɪ/ is more frequently labeled as one-mora /i/ in /hVt/ and /pVt/ contexts than in /pVn/ and /pVl/ contexts, and /ɪ/ is identified more correctly in /hVt/ and /pVt/ contexts. The /ɪ/ is also labeled as /e/ in 33.7% of instances in /pVn/ context, and /ɪ/ is mistakenly identified as /ɛ/ in 30.3% of instances in /pVn/. /æ/ is more frequently equated with /a/ or /aː/ in the contexts of /hVt/ and /pVt/ than in the /pVn/ and /pVl/ contexts, and the identification accuracy of /æ/ is higher in the /hVt/ and /pVt/ contexts. In the contexts of /pVn/ and /pVl/, /æ/ is more frequently equated with /e/ than with /a/, and, in these contexts, /æ/ is mistakenly identified as /ɛ/ in 43.6% and 40% of instances, respectively. Therefore, it appears that for /æ/ to be identified correctly by native Japanese listeners, it is important that the vowel is perceived as an exemplar of /a/. /ɑ/ is mistakenly identified as /æ/ in 35.2% of instances. /ʌ/ is least accurately identified in all of the contexts. It seems that Japanese listeners may not have a clear image of the vowel. In /pVn/ and /pVl/ context, /ʌ/ is more frequently misidentified as /ɑ/ than is correctly identified. In these contexts, /ʌ/ tends to be equated with /o/, /ou/, or /oː/ than with /a/ or /aː/.

It has to be stressed that there is a very significant individual difference. Figures 21-24 show the boxplots of identification accuracy (in percent) in four different consonantal contexts. While at least one participant gave correct responses to all the /ɛ/and /ɑ/ trials in /pVt/ context, /i/, /ɛ/, /ɑ/, and /ʌ/ trials in /pVn/ context, and /æ/ and /ɑ/ trials in /pVl/ contexts, at least one participant gave no correct responses to these trials. Nine participants responded incorrectly to all of the /ʌ/ trials in /pVl/ context. Individual differences in identification accuracy and perceptual assimilation will be discussed in Chapter 8.

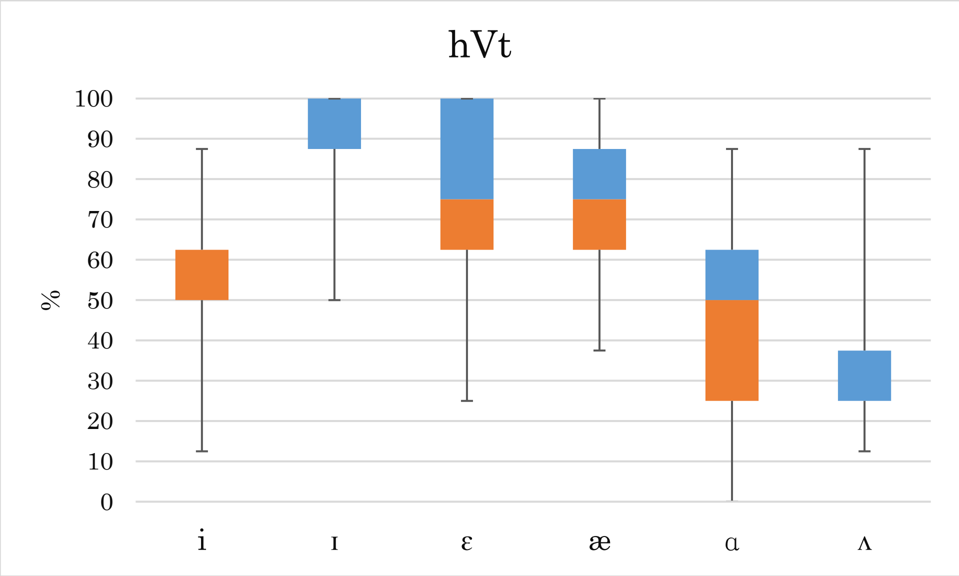

Figure 21. Boxplot of identification accuracy of /i, ɪ, ɛ, æ, ɑ, ʌ/ in /hVt/ context

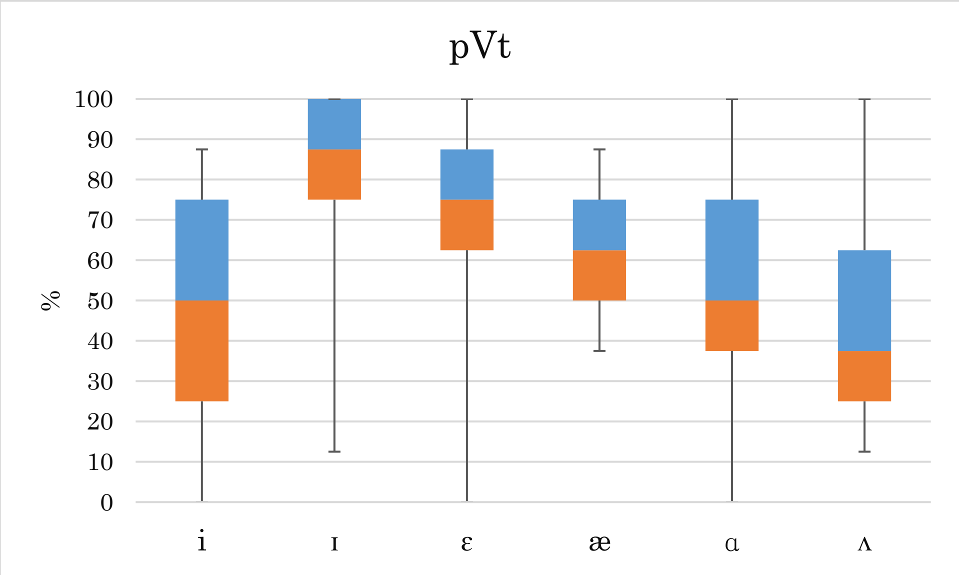

Figure 22. Boxplot of identification accuracy of /i, ɪ, ɛ, æ, ɑ, ʌ/ in /pVt/ context

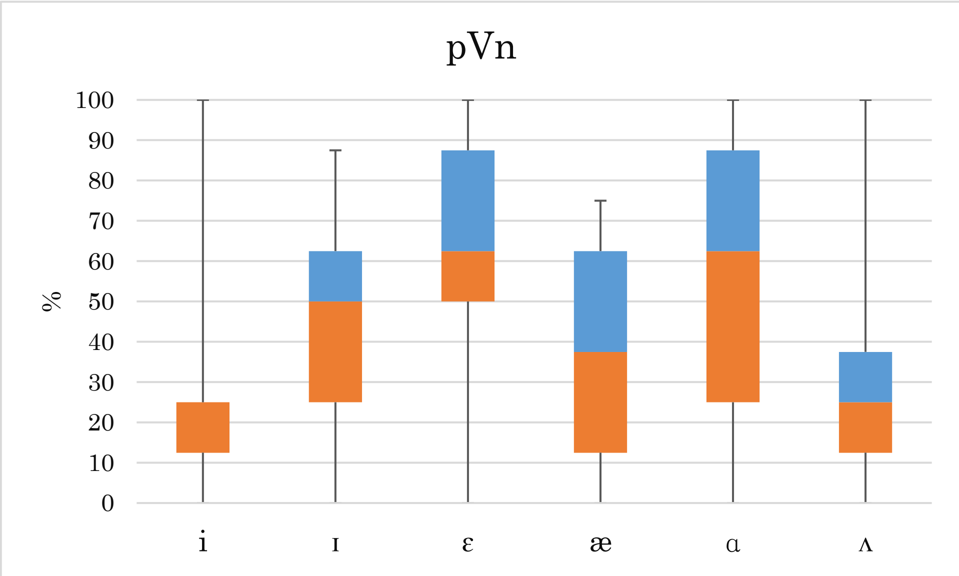

Figure 23. Boxplot of identification accuracy of /i, ɪ, ɛ, æ, ɑ, ʌ/ in /pVn/ context

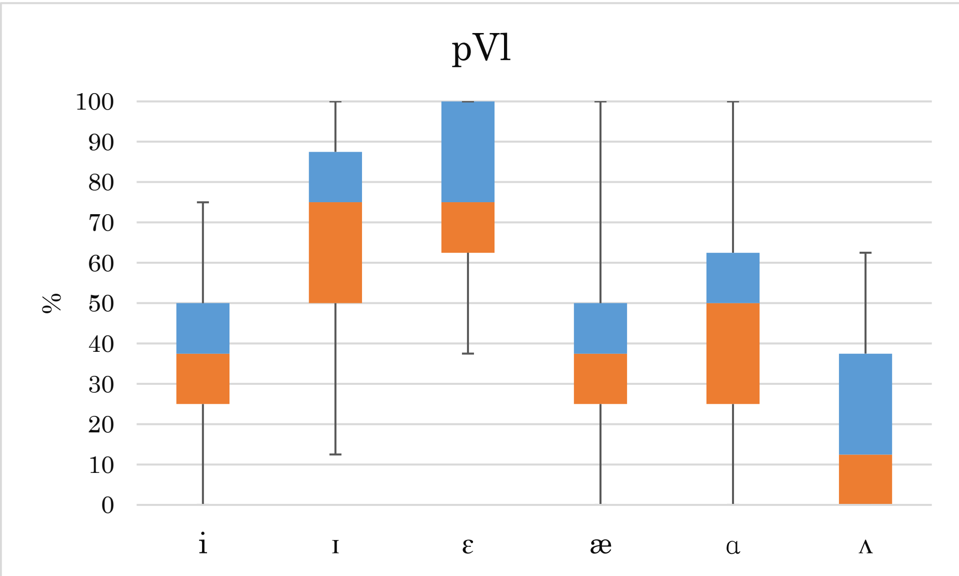

Figure 24. Boxplot of identification accuracy of /i, ɪ, ɛ, æ, ɑ, ʌ/ in /pVl/ context

There is no perfect match between the results of the perceptual assimilation experiment and the identification experiment, but it appears to be the case that the ways in which American English vowels are labeled in terms of native Japanese vowel categories play an important role in the identification of American English vowels by native Japanese listeners, and the ways American English vowels are mapped into Japanese vowel categories are strongly affected by the manner of articulation of the following consonant.

5. Discrimination Experiment

5.1 Procedure

Six vowel pairs made up of spectrally close American English vowels were created: /i/-/ɪ/, /ɛ/-/ɪ/, /æ/-/ɛ/, /æ/-/ɑ/, /æ/-/ʌ/, /ɑ/-/ʌ/. Japanese listeners' sensitivity levels to categorically discriminate six vowel pairs were assessed by AXB format. A participant heard three stimuli per trial and decided whether the second stimulus was categorically the same as the first or the third stimulus. The three stimuli in each trial were from utterances by different talkers. A participant's sensitivity to discriminate each vowel pair was assessed by 12 trials. For example, 12 trials to assess the /i/-/ɪ pair were composed of 3 /i/-/i/-/ɪ/ trials, 3 /i/-/ɪ/-/ɪ/ trials, 3 /ɪ/-/i/-/i/ trials, and 3 /ɪ/-/ɪ/-/i/ trials. Thus, 72 trials (12 trials × 6 vowel pairs) were created. The inter-stimulus interval (ISI) and the inter-trial stimulus (ITI) were both 1,000ms. Each participant responded by moving the cursor to the "First" or "Last" box on the computer screen and click on it.

After the instruction, a participant worked on 10 practice trials. The order of consonantal context was counterbalanced across participants.

5.2 Results

The discrimination accuracy of each vowel pair in each consonantal context is shown in percent in Figure 25.

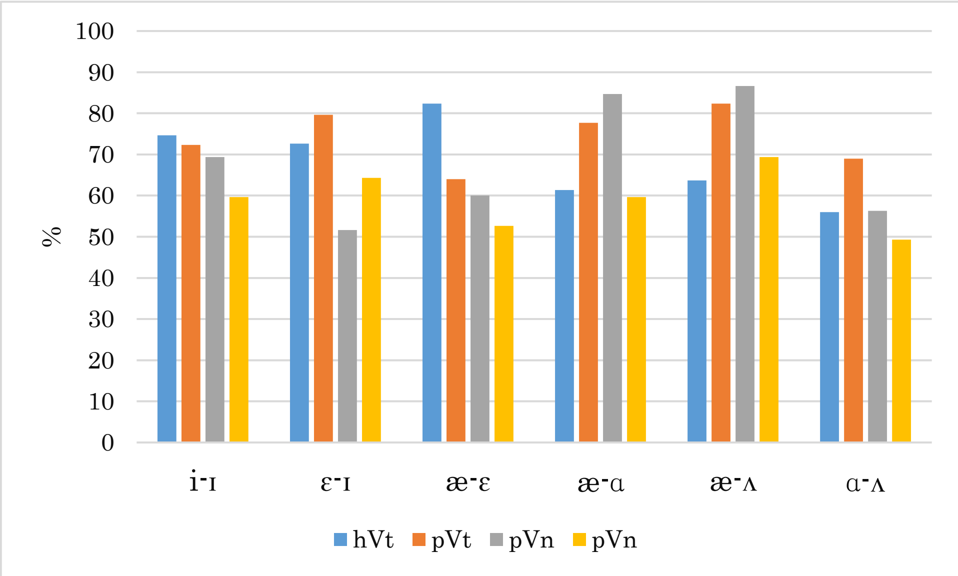

Figure 25. The percentages of correct discrimination of six vowel pairs in four consonantal contexts

The results were submitted to a repeated-measures ANOVA with 4 Consonantal contexts and 6 Vowel pairs as within-subject variables. The obtained results revealed that the main effect of Consonantal context [$F(3, 96)=13.46$, $p<.001$], Vowel pairs [$F(5, 160)=21.63$, $p<.001$], and the interaction between Consonantal Context × Vowel pairs [$F(15, 480)=12.96$, $p<.001$] are all significant. Overall, vowel pairs were discriminated most accurately in the context of /pVt/ and least accurately in the context of /pVl/. The vowel pair /ɑ/-/ʌ/ was least accurately discriminated, and the discrimination accuracy is significantly lower than that of other pairs, except /æ/-/ɛ/, at least at the $p<.005$ level. /æ/-/ɛ/ was significantly less accurately discriminated than /æ/-/ɑ/ and /æ/-/ʌ/ at $p<.001$ level. Bonferroni-adjusted post-hoc pair-wise comparisons revealed that vowel pairs were discriminated significantly more accurately in /pVt/ context than in /hVt/ context ($p<.001$) or in /pVl/ context ($p<.001$). Bonferroni-adjusted pair-wise comparisons also revealed that there is a significant difference in discrimination accuracy of vowel pairs between consonantal contexts at least at $p<.05$ level as shown in Table 16.

Table 16.

Results of pair-wise comparisons (“＜” denotes less accurately than, “＞” denotes more accurately than)

/i/-/ɪ/	/ɛ/-/ɪ/	/æ/-/ɛ/	/æ/-/ɑ/	/æ/-/ʌ/	/ɑ/-/ʌ/
pVl<hVt, pVt	pVt>pVn, pVl hVt>pVn	hVt> all the others pVt >pVl	hVt<pVt, pVn pVl<pVt, pVn	hVt<pVt, pVn pVl<pVn	pVt > all the others

Overall vowels seem to be less distinctive when /n/ or /l/ follows. Low discrimination accuracy of /i/-/ɪ/, /ɛ/-/ɪ/ and /æ/-/ɛ/ may be attributed to the fact that F2 descends in this context. Descending F2s of front vowels draw similar trajectory curves. Low discrimination accuracy of /ɛ/-/ɪ/ in /pVn/ context is due to the fact that /ɪ/ is perceived lower in this context. /ɪ/ was labeled as Japanese /i/ more frequently in this context. Vowel pairs /æ/-/ɑ/ and /æ/-/ʌ/ were discriminated better in /pVn/ context. This is primarily because /æ/ is shifted forward and upward in this context and is more distant from /ɑ/ and /ʌ/.

As in the Identification Experiment, large individual differences are observed, but because the chance level is higher than in the identification task, no participant’s discrimination accuracy is 0%. In many vowel pairs, at least one participant’s discrimination accuracy reaches 100%.

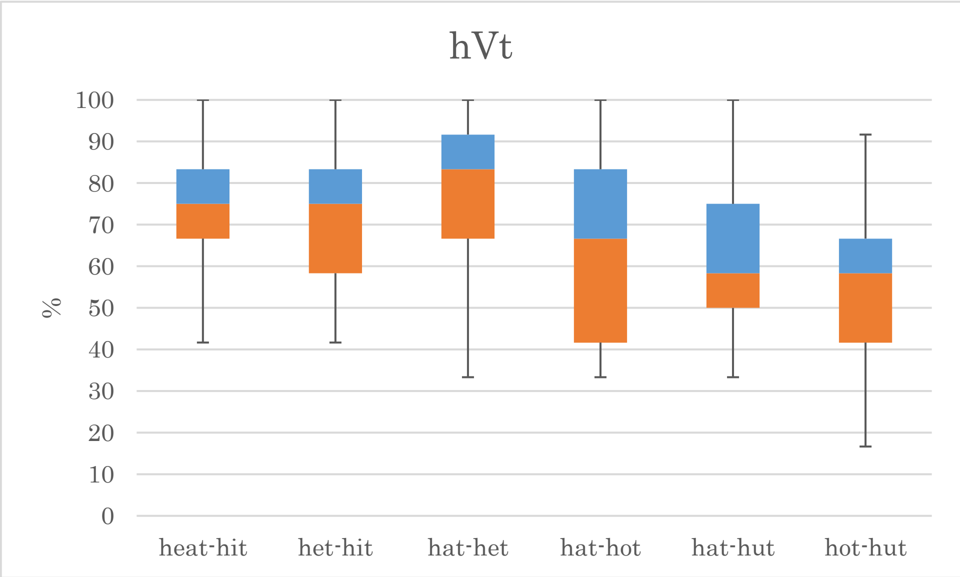

Figure 26. Boxplot of discrimination accuracy of six vowel pairs in /hVt/ context

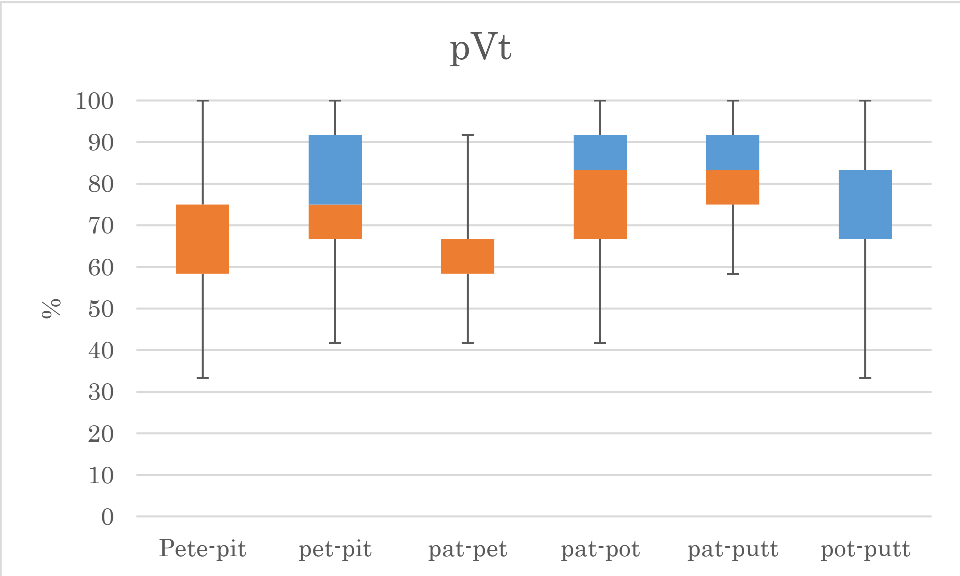

Figure 27. Boxplot of discrimination accuracy of six vowel pairs in /pVt/ context

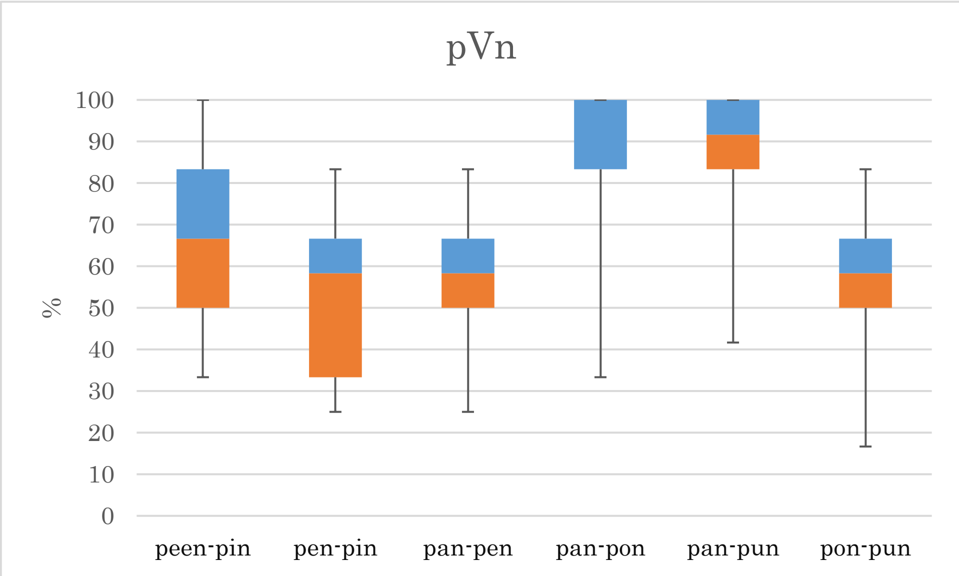

Figure 28. Boxplot of discrimination accuracy of six vowel pairs in /pVn/ context

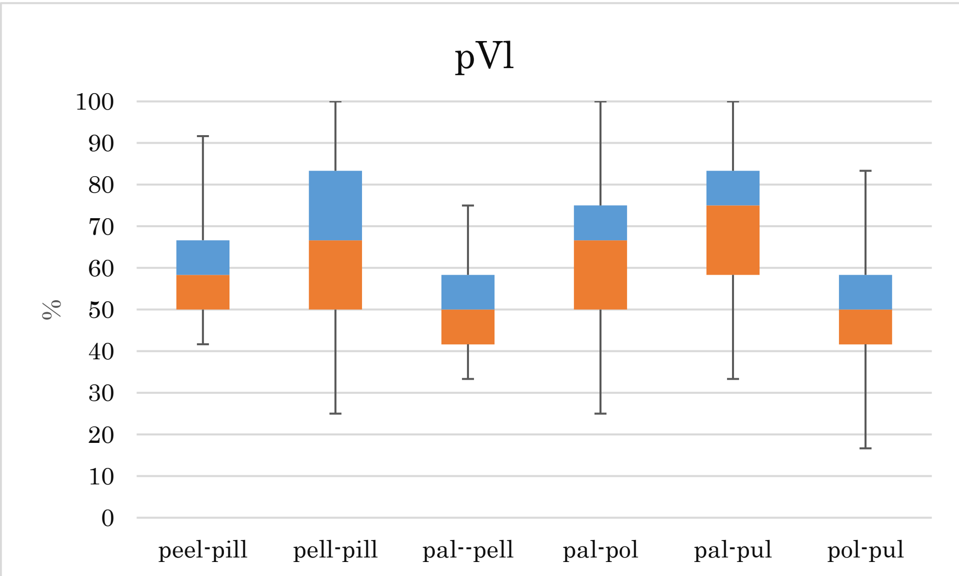

Figure 29. Boxplot of discrimination accuracy of six vowel pairs in /pVl/ context

6. Production Experiment

6.1 Procedure

Six American English vowels produced in four different consonantal contexts by 33 participants were recorded and collected in two different formats. First, a participant read aloud words on the word list (Read Aloud Condition). A word list, as shown in Table 11, was handed to a participant, and he/she was asked to read each word three times in isolation and in a carrier sentence: "Now I say 'the word' to you." Once again, a participant was told that words in the same column contain the vowel even though they are spelled differently. For instance, they were told that "heat," "Pete," "peen," and "peel" contain the same vowel /i/ as in "beat" and "bean." They were also told that the word list contains non-words like "het," "pol," and "pul."

Participants were told that they should produce these words in a way they think is correct. American English vowels produced in this way should represent what Japanese speakers believe English vowels sound like, or they should be auditory images of English vowels that Japanese speakers have. And a participant was expected to use the same phonetic cues or features to differentiate American English vowels as in the perception experiment. So, for example, if a participant relies on durational difference to differentiate /i/ and /ɪ/ in perception, he/she should differentiate these two vowels by duration in production as well. And if Japanese speakers' auditory images of English vowels is distant from how English vowels actually sound to them, that should lead to inaccurate perception.

Next, a participant heard each stimulus in a randomized order over a headset, and repeated (or reproduced) each word (Reproduction Condition). The same set of stimuli used in perception experiments was used. The stimuli were blocked between consonantal contexts, and each stimulus was played once. Participants were told in which consonantal context they would hear vowels next, but they were not told which vowel they would hear in each trial. Vowels produced in this way should represent how each vowel sounds to Japanese speakers, and the gap between features of vowels recorded in the two different conditions should help figure out what causes inaccurate identification and discrimination.

Participants' utterances were digitally recorded on Maranz Portable SD Card Recorder PMD620MKII, using a head-worn Shure 10 microphone.

6.2 Results

6.2.1 Read Aloud Condition

Formant frequencies were measured in the same manner as the stimuli. Because both male and female participants were involved, mean formant frequencies were calculated separately for male and female participants. Figure 30 shows mean F1 and F2 frequencies of six American English vowels uttered in the context of /hVt/ in the Read Aloud Condition by 28 female participants, and Figure 31 shows mean F1 and F2 frequencies of 6 vowels uttered by 5 male participants. Tokens uttered in the carrier sentence are excluded from the calculation.

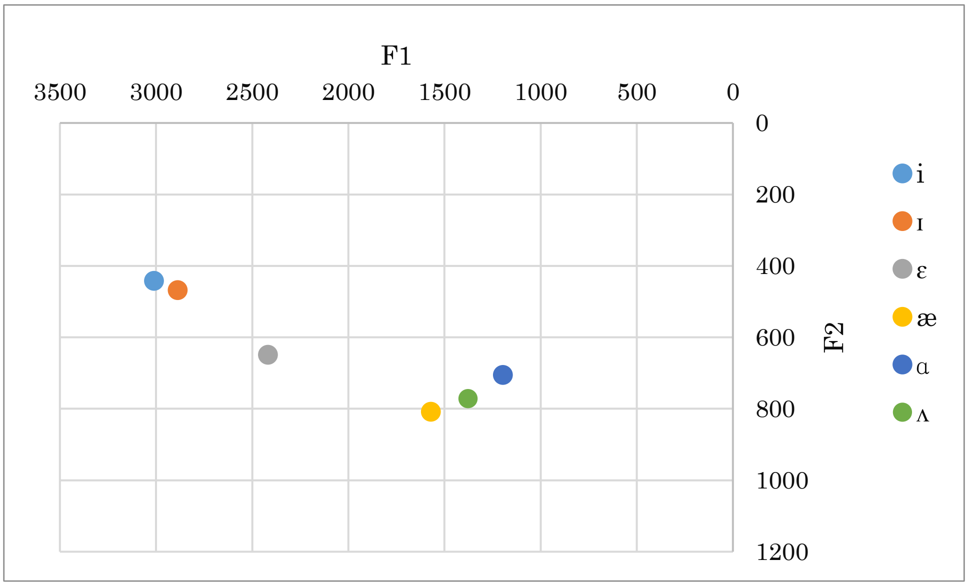

Figure 30. Mean F1 and F2 frequencies of /i, ɪ, ɛ, æ, ɑ, ʌ/ uttered in /hVt/ context in "Read Aloud" condition by 28 female Japanese participants

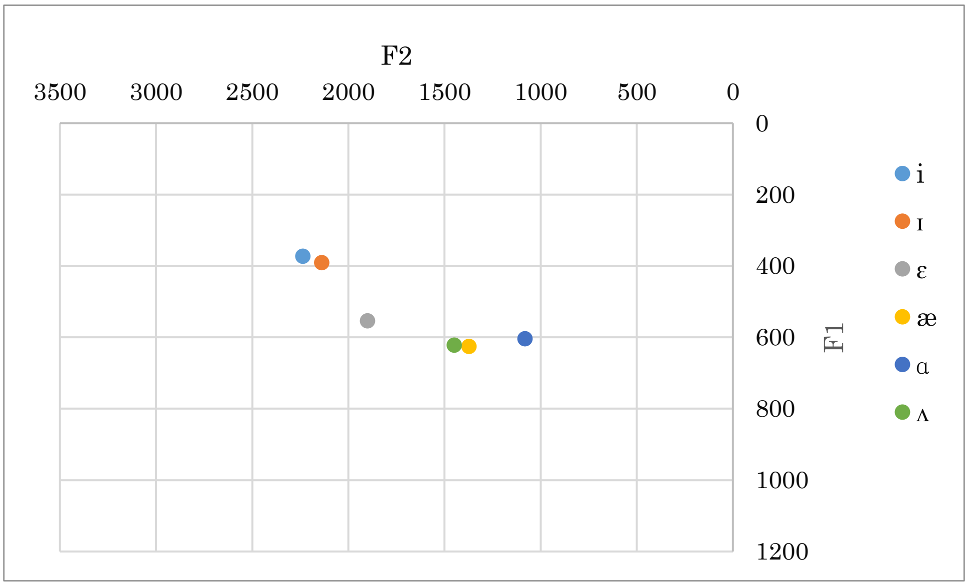

Figure 31. Mean F1 and F2 frequencies of /i, ɪ, ɛ, æ, ɑ, ʌ/ uttered in /hVt/ context in "Read Aloud" condition by 5 male Japanese participants

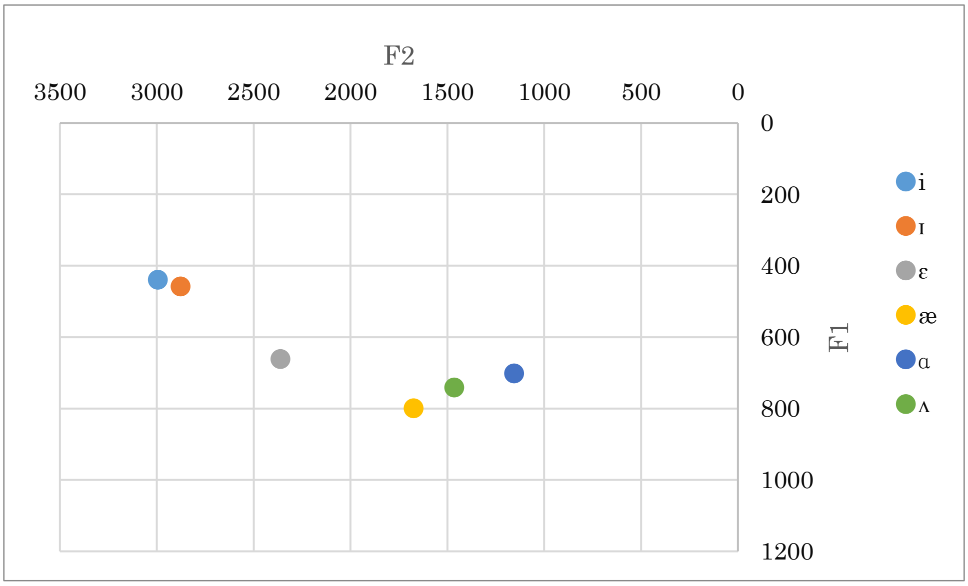

Figure 32. Mean F1 and F2 frequencies of /i, ɪ, ɛ, æ, ɑ, ʌ/ uttered in /pVt/ context in "Read Aloud" condition by 28 female Japanese participants

In female utterances, /æ/ is the lowest, and /ʌ/ is located behind /æ/, but in male utterances, /æ/ is behind /ʌ/. The vowels /i/ and /ɪ/ are close in both male and female utterances, and /ɛ/ is located between /i/, /ɪ/ and low or central vowels. /ɑ/ is backer than the other vowels. Compared with Japanese vowels data, /ɛ/ appears to be lower than

Japanese /e/, and male participants' /æ/ and /ʌ/ seem to be more fronted than Japanese /a/. The /ɛ/ that Japanese participants produced is a mid front vowel, close to the Japanese /e/.

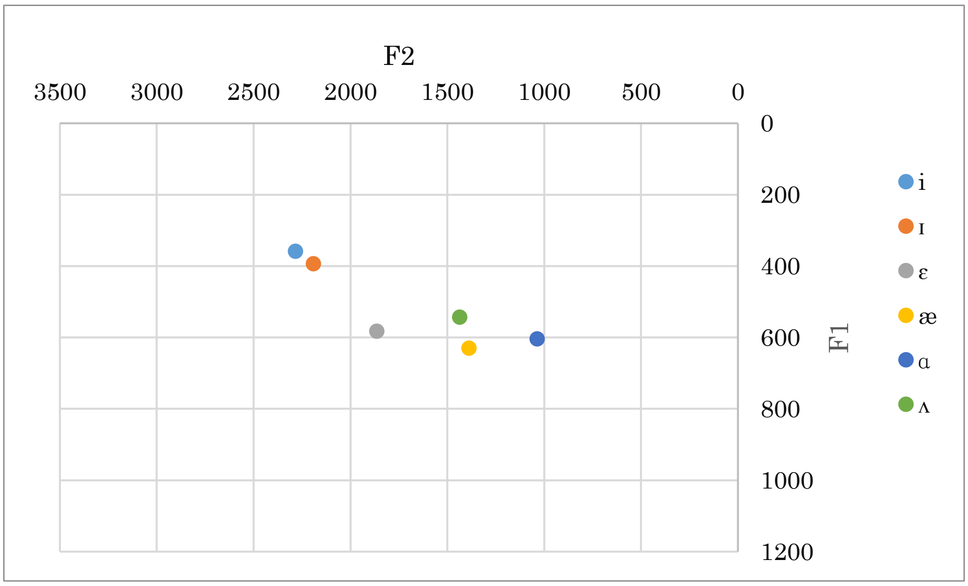

Figure 33. Mean F1 and F2 frequencies of /i, ɪ, ɛ, æ, ɑ, ʌ/ uttered in /pVt/ context in "Read Aloud" condition by 5 male Japanese participants

Mean F1 and F2 frequencies of female utterances are basically the same as those in the /hVt/ context. Male utterances show a slightly different pattern, and /ʌ/ is higher in the /pVt/ context. Male participants' data are an average of just five speakers' utterances, and so individual speakers' characteristics exercise a stronger influence on the averaged data.

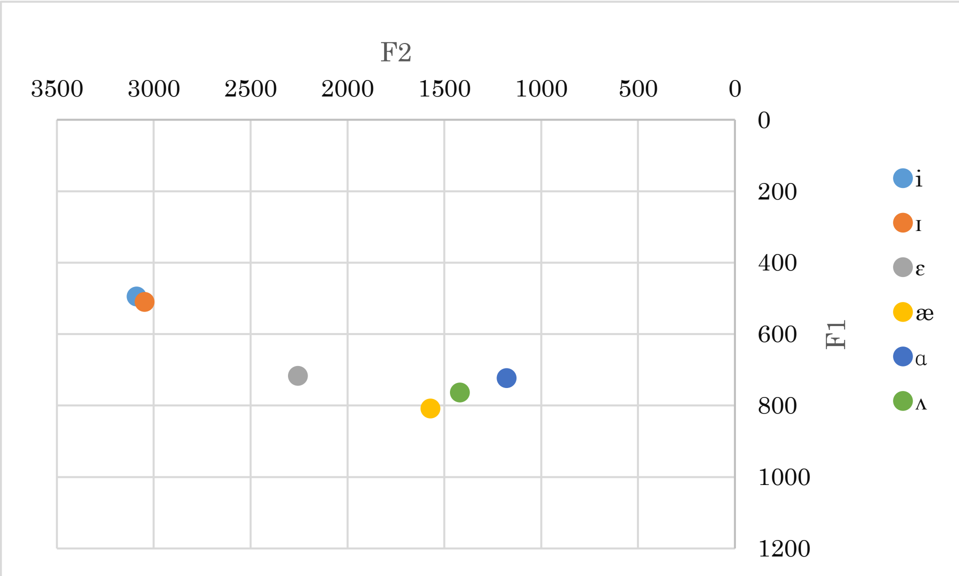

Figure 34. Mean F1 and F2 frequencies of /i, ɪ, ɛ, æ, ɑ, ʌ/ uttered in /pVn/ context in "Read Aloud" condition by 28 female Japanese participants

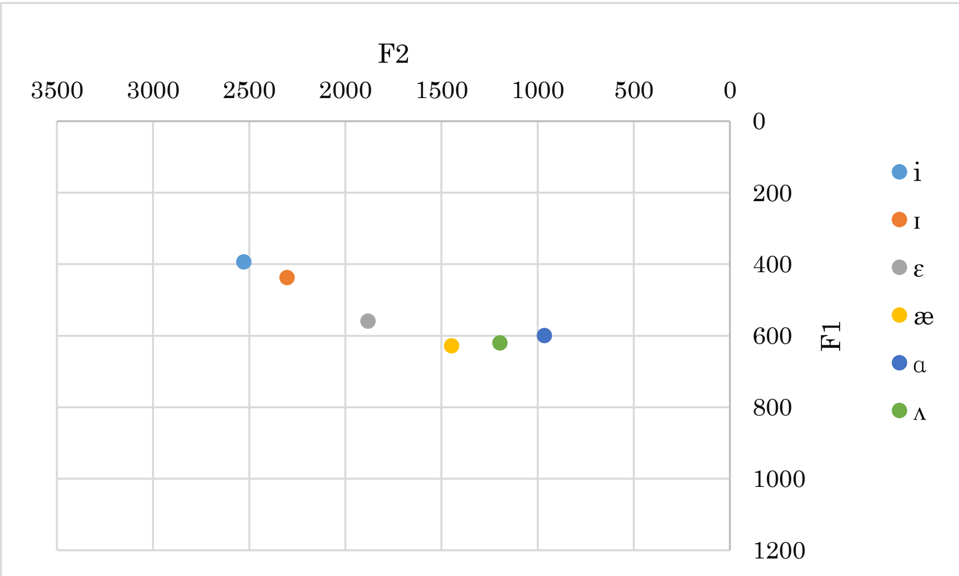

Figure 35. Mean F1 and F2 frequencies of /i, ɪ, ɛ, æ, ɑ, ʌ/ uttered in /pVn/ context in "Read Aloud" condition by 5 male Japanese participants

Formant frequencies are not much different from those in /hVt/ and /pVt/ contexts. Male participants' /ɪ/ is further separated from /i/ than in /hVt/ and /pVt/ contexts. Other than this, no difference here is noteworthy.

Formant frequencies of six vowels in the /pVl/ context were measured at three locations in the same way as were the four native speakers' utterances. Figure UU shows F1 and F2 frequencies, respectively, of six American English vowels uttered by 28 female participants in "read aloud" condition. Generally, the relative positions of vowels are the same as in the other three consonantal contexts. As we compare F1 and F2 frequencies at 25%, 50%, and 75% of /Vl/ sequences, it can be seen that F2 frequencies gradually decrease, but compared with native speakers' utterances, vowels are still defused, meaning that vowels are not retracted as much as in native speakers' utterances. This can be attributed to Japanese speakers' inability to produce postvocalic /l/ authentically. The same essential tendency is seen in male participants' utterances. Their front vowels are slightly less retracted than female participants' utterances.

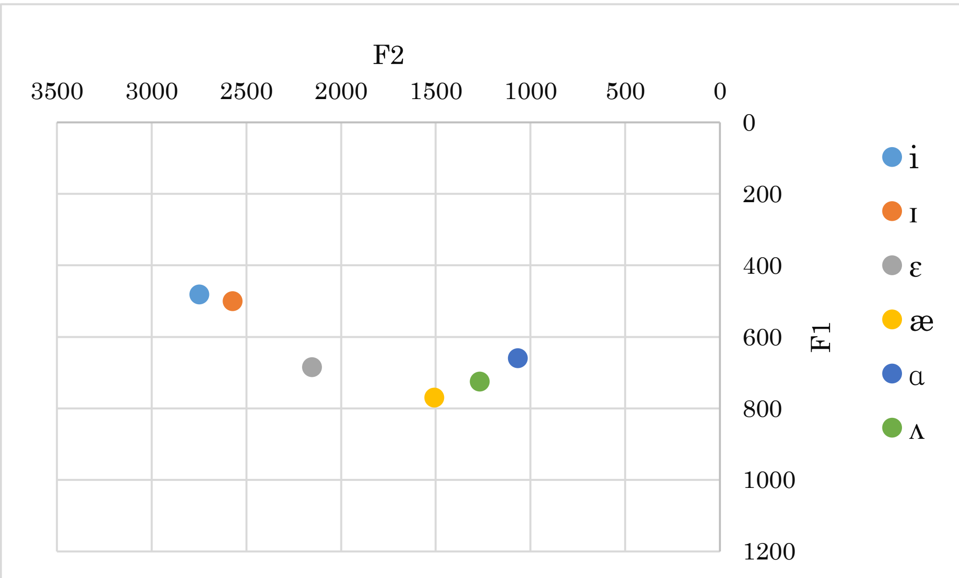

Figure 36. Mean F1 and F2 frequencies at the 25 % of six /Vl/ sequences uttered in "Read Aloud" condition by 28 female Japanese participants

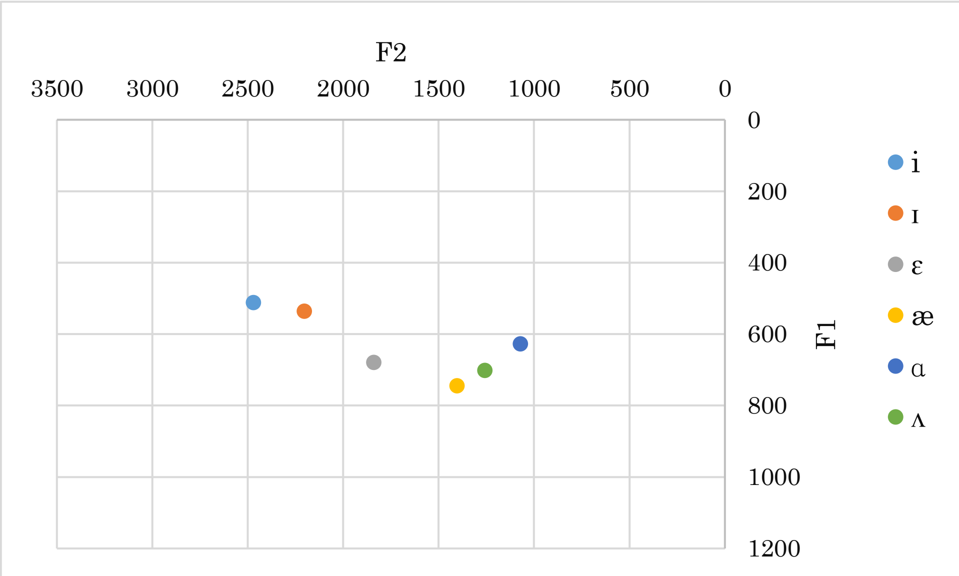

Figure 37. Mean F1 and F2 frequencies at the 50 % of six /Vl/ sequences uttered in "Read Aloud" condition by 28 female Japanese participants

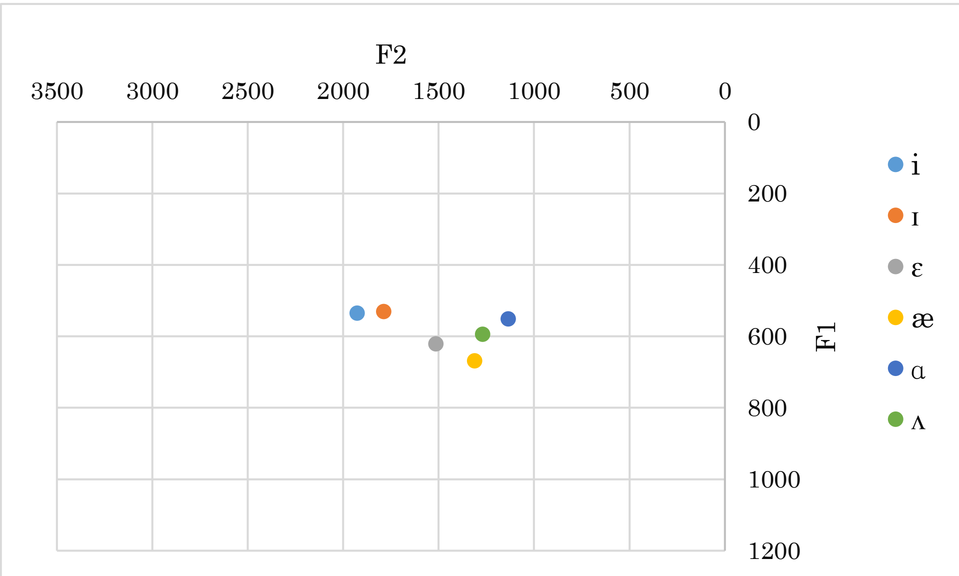

Figure 38. Mean F1 and F2 frequencies at the 75 % of six /Vl/ sequences uttered in "Read Aloud" condition by 28 female Japanese participants

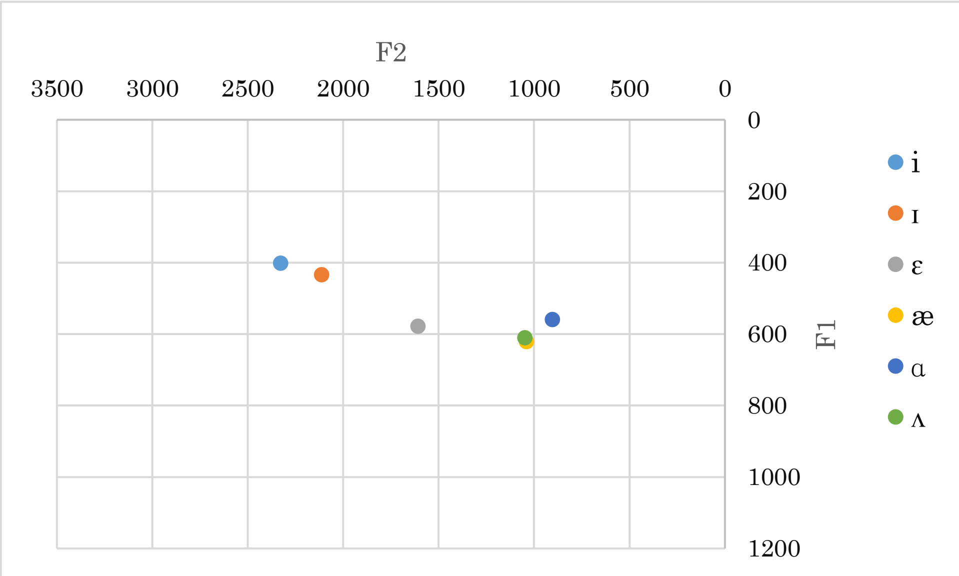

Figure 39. Mean F1 and F2 frequencies at the 25 % of six /Vl/ sequences uttered in "Read Aloud" condition by 5 male Japanese participants

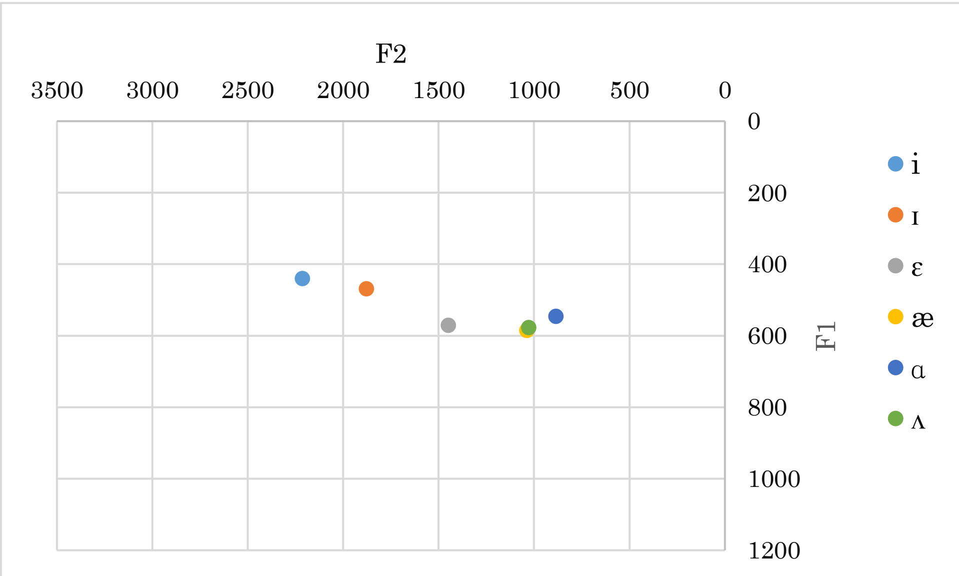

Figure 40. Mean F1 and F2 frequencies at the 50 % of six /Vl/ sequences uttered in "Read Aloud" condition by 5 male Japanese participants

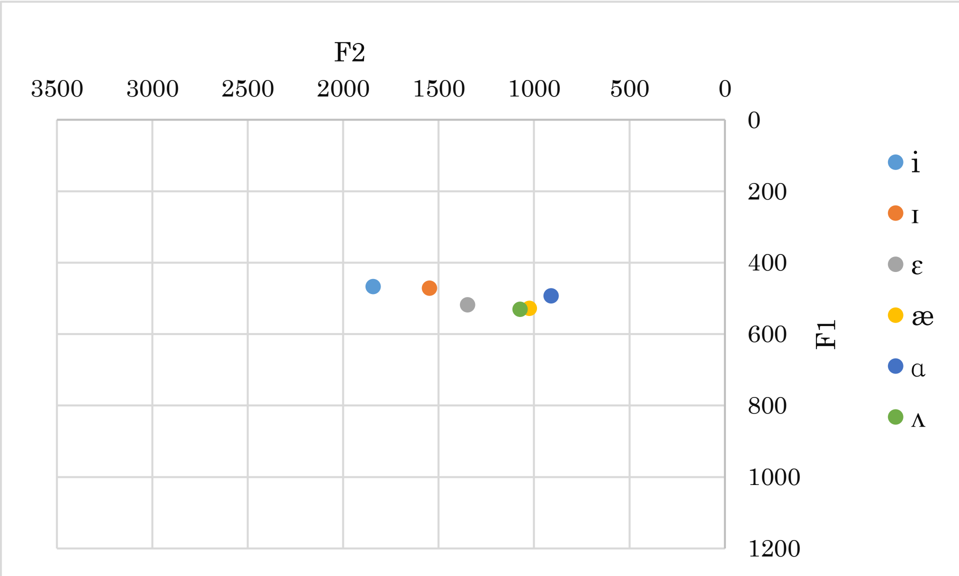

Figure 41. Mean F1 and F2 frequencies at the 75 % of six /Vl/ sequences uttered in "Read Aloud" condition by 5 male Japanese participants

The mean duration of each vowel is shown in Figures 42-45. In /pVl/ context, the mean duration of /Vl/ continua is shown. As seen in these figures, /i/ is produced with a longer duration than any other vowel in this study, in all of the contexts. In the context of /pVl/, the durational difference appears smaller because the following /l/ is included.

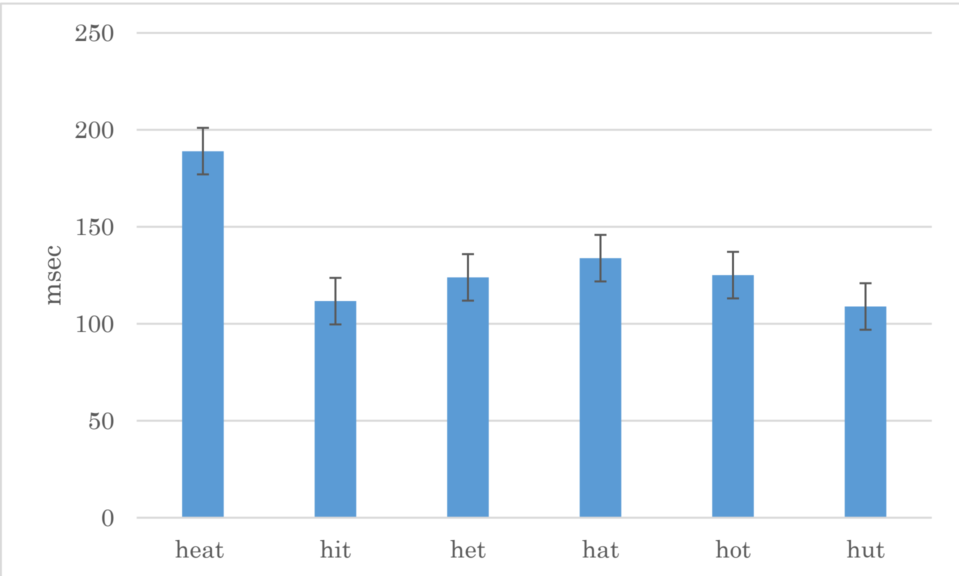

Figure 42. Mean vowel duration in milliseconds uttered by Japanese participants in /hVt/ context in "Read Aloud" condition

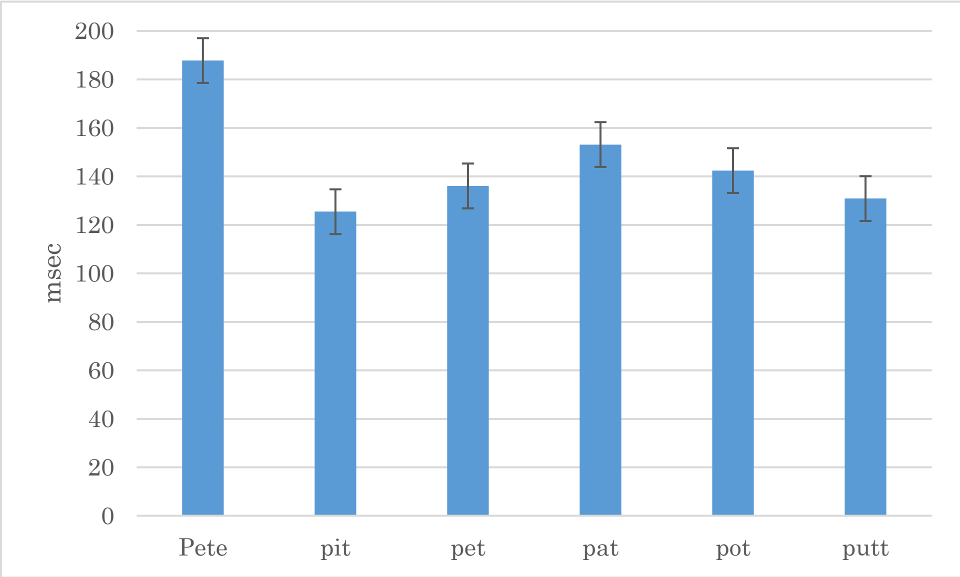

Figure 43. Mean vowel duration in milliseconds uttered by Japanese participants in /pVt/ context in “Read Aloud” condition

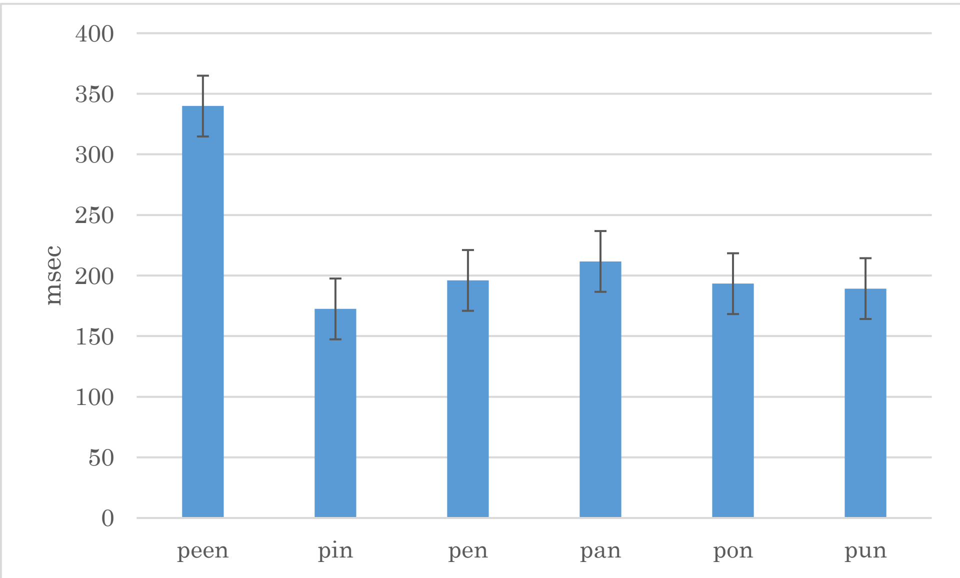

Figure 44. Mean vowel duration in milliseconds uttered by Japanese participants in /pVn/ context in “Read Aloud” condition

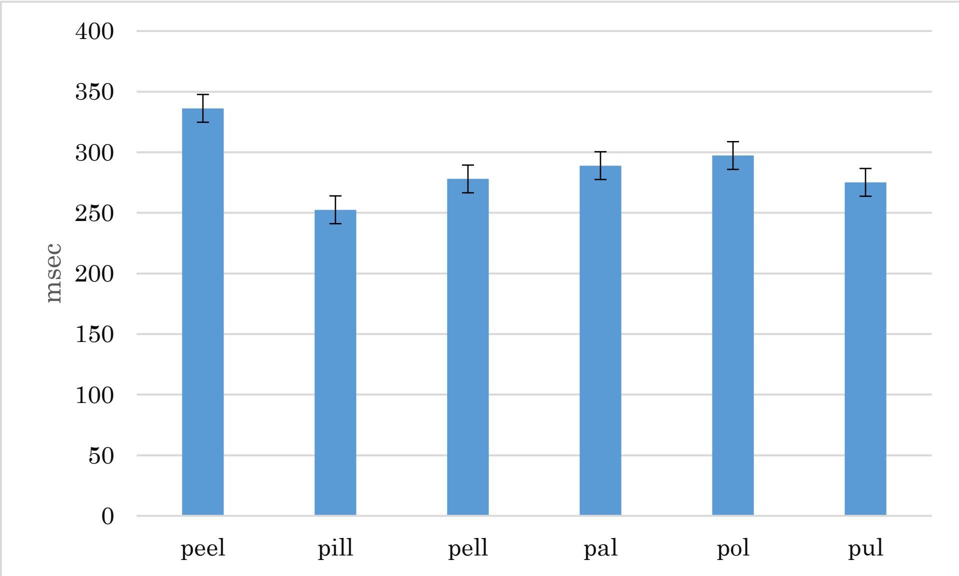

Figure 45. Mean duration of /Vl/ sequences in milliseconds uttered by Japanese participants in "Read Aloud" condition

Generally, the results of the production experiment in the "Read Aloud" condition agree with the commonly accepted Japanese adaptation of English vowels. The vowel /i/ is treated as a long vowel, and the other five vowels are treated as short vowels. The /i/ and /ɪ/ are mainly differentiated by duration rather than by quality. In /pVl/ context, the spectral difference between /i/ and /ɪ/ are larger probably because F2 of /ɪ/ descends earlier in a /Vl/ continuum because the vowel is shorter. Japanese participants' production of /æ/ is a low central vowel, close to Japanese /a/. Their production of /ɑ/ is backer than the other vowels, and higher than /æ/, and often higher than /ʌ/ as well. It seems that Japanese participants' image of /ɑ/ is a vowel close to Japanese /o/. The vowel /ʌ/ is also produced as a vowel close to Japanese /a/, but its position in the vowel space with reference to /æ/ is slightly unstable. Both /æ/ and /ʌ/ are adapted as /a/, it seems that Japanese participants are not sure of how to differentiate these two vowels. To sum up, Japanese speakers' auditory image of English vowels is strongly affected by Japanese adaptation of English vowels.

6.2.2 Reproduction Condition

As in "Read Aloud" condition, F1 and F2 frequencies were measured at the midpoint of each vowel token. Participants repeated after four native speakers' utterances, and because they may not have perceived each native speaker's utterance equally, mean F1 and F2 frequencies were calculated separately, depending on which native speaker made the utterance.

6.2.2.1 /hVt/ context

In the following figures, mean F1 and F2 frequencies of American English vowels uttered by Japanese participants are shown in circles ● and F1 and F2 frequencies of native speakers are shown in squares ■. Mean F1 and F2 frequencies of female participants are compared with those of a native speaker in the same figure, and mean F1 and F2 frequencies of male participants are shown separately in a different figure.

In general, native speakers of English use a vertically larger vowel space than do Japanese speakers. Native speakers' /i/ tokens are higher than those of Japanese speakers, and native speakers' /æ/ and /ɑ/ tokens are lower than those of Japanese speakers. However, Japanese speakers' /æ/ and /ɑ/ tokens are lower than they are in "Read Aloud" condition.

Japanese speakers' /i/ and /ɪ/ tokens are spectrally more distant than in "Read Aloud" condition, indicating that Japanese speakers discerned at least some spectral difference between the two vowels. However, spectral difference between native speakers' /i/ and /ɪ/ tokens is still larger than that of Japanese speakers. Japanese speakers' /i/ and /ɪ/ are closest when they repeated after Talker 4 (Figures 50 and 51). These findings agree with the results of the Perceptual Assimilation Experiment. Talker 4's /i/ and /ɪ/ are both most frequently classified as exemplars of the Japanese short high front vowel /i/, and they respectively received mean category goodness ratings of 4.9 and 4.3. The mean category goodness rating of /ɪ/ in the context of /hVt/ averaged across four talkers is 3.8, as shown in Table 10, and so Talker 4's /ɪ/ is closer to the category ideal of Japanese /i/ than are /ɪ/ tokens of the other three talkers. The spectral difference between Talker 4's /i/ and /ɪ/ may be less discernible. Among the Japanese speakers are those who seemed to discern the spectral difference between the two vowels and those who seemed to be unaware of the difference. Individual differences will be dealt with in Chapter 8.

Native speakers' /ɛ/ tokens are generally lower and more centralized than those of Japanese speakers. Japanese speakers' /ɛ/ tokens, which they repeated after Talkers 1 and 3, are a little lower than /ɛ/ in "Read Aloud" condition, and Japanese speakers' /ɛ/ tokens which they repeated after Talkers 1, 3, and 4 are a little more centralized than in "Read Aloud" condition.

Japanese speakers' /æ/ is a little bit forwarded in "Repetition" condition. This is especially true for the tokens they produced after Talker 5. Japanese speakers' /ɑ/ tokens are generally lower than in "Read Aloud" condition, but still higher compared with native speakers' tokens.

As far as female participants are concerned, their low vowels /æ/ and /ɑ/ are higher when they repeated after Talker 5 (Figure 52), indicating that the participants perceived the relatively higher /æ/ and /ɑ/ that Talker 5 produced. As for male participants, their /æ/ and /ɑ/ are not higher in terms of mean formant frequencies, when they repeated after Talker 5.

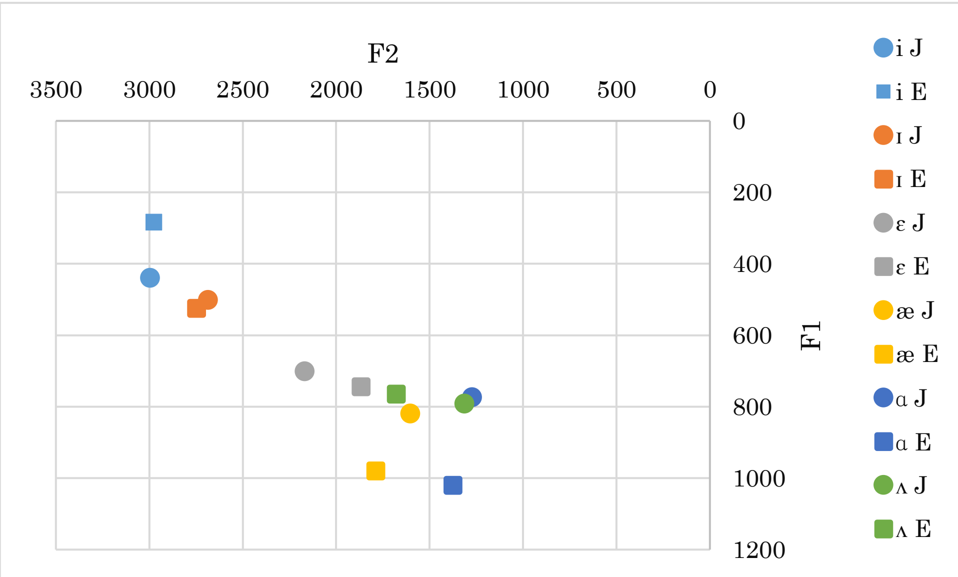

*Figure 46.*Mean F1 and F2 frequencies of six American English vowels uttered by 28 female participants (Repeated after Talker 1)

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 1.

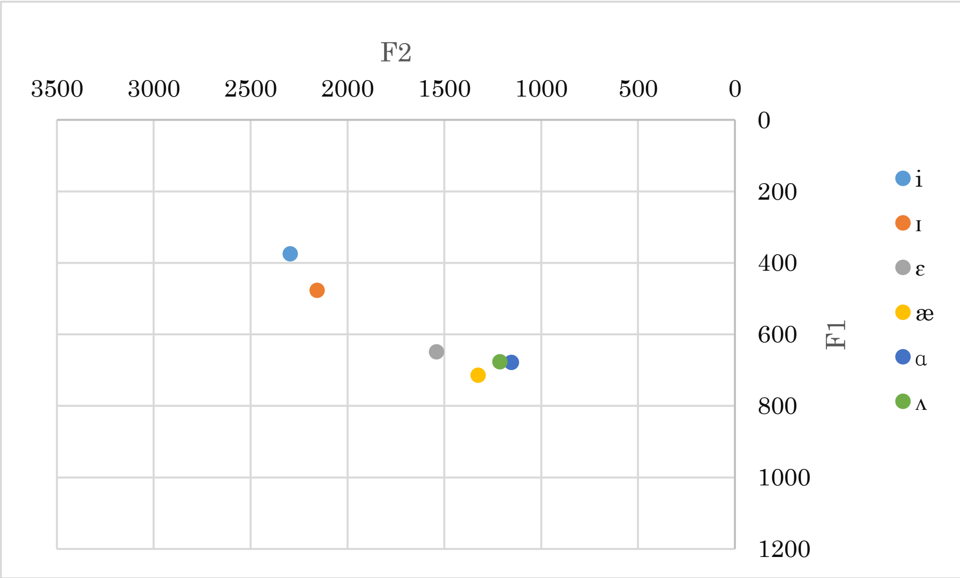

Figure 47. Mean F1 and F2 frequencies of six American English vowels uttered by five male participants (Repeated after Talker 1)

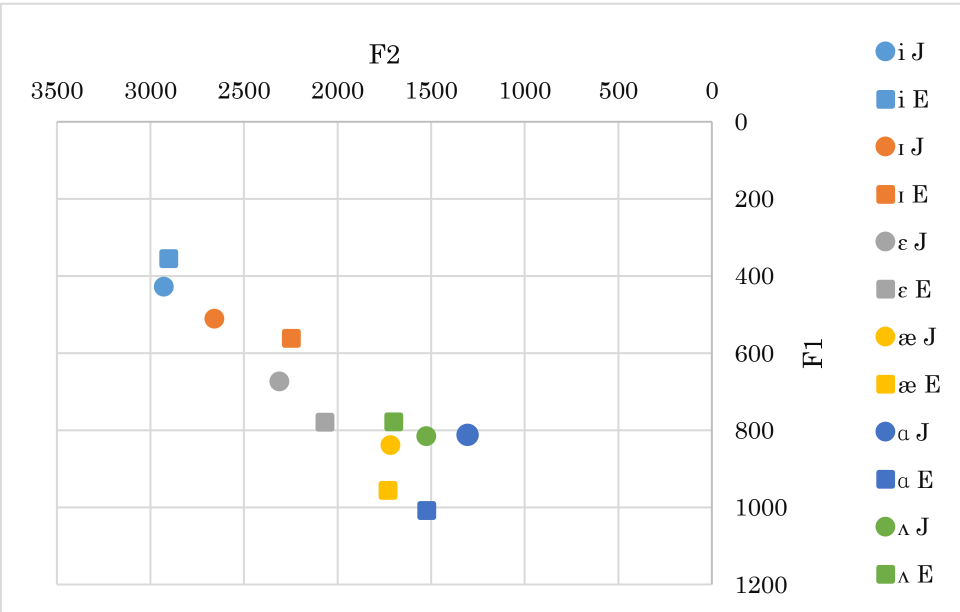

Figure 48. Mean F1 and F2 frequencies of six American English vowels uttered by female participants (Repeated after Talker 3).

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 3.

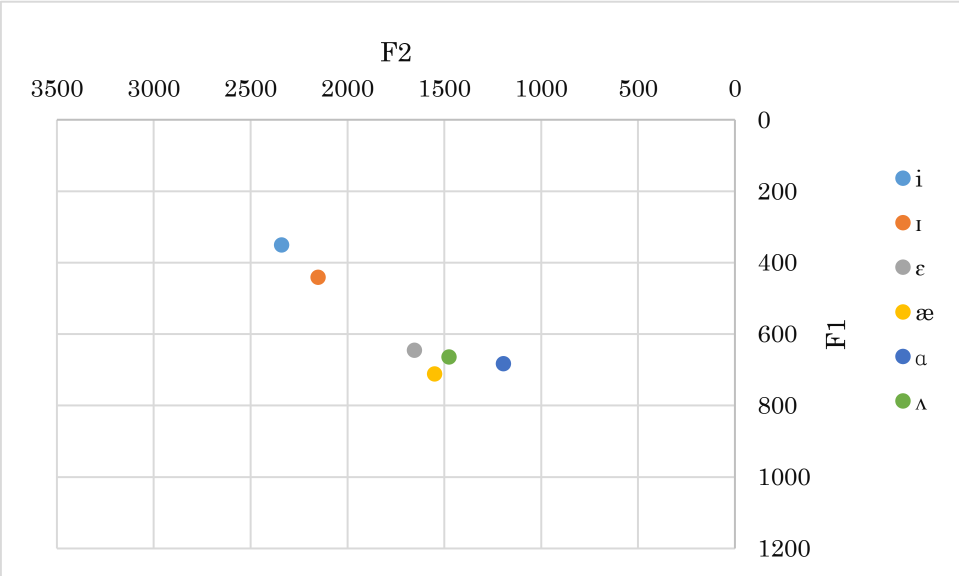

Figure 49. Mean F1 and F2 frequencies of six American English vowels uttered by male participants (Repeated after Talker 3)

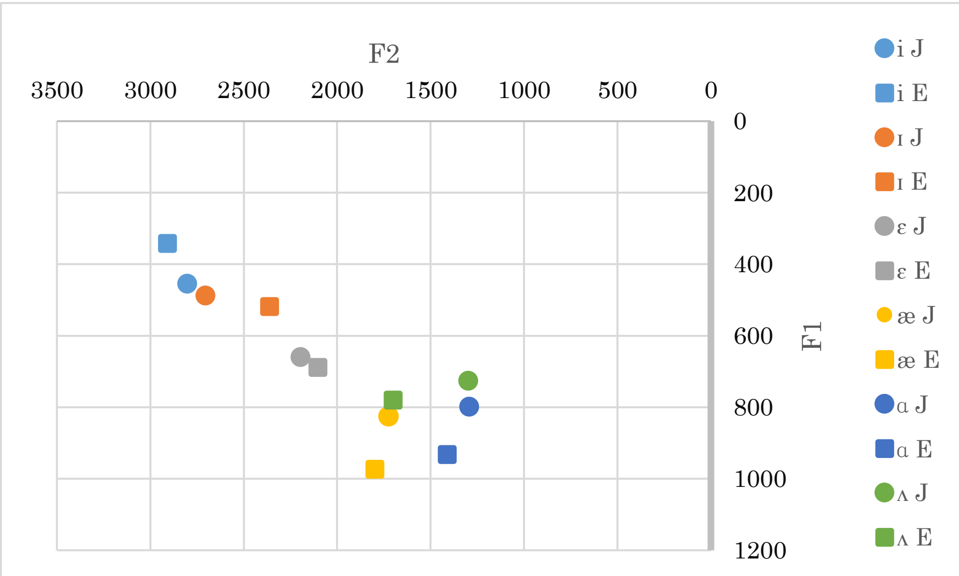

Figure 50. Mean F1 and F2 frequencies of six American English vowels uttered by female participants (Repeated after Talker 4).

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 4.

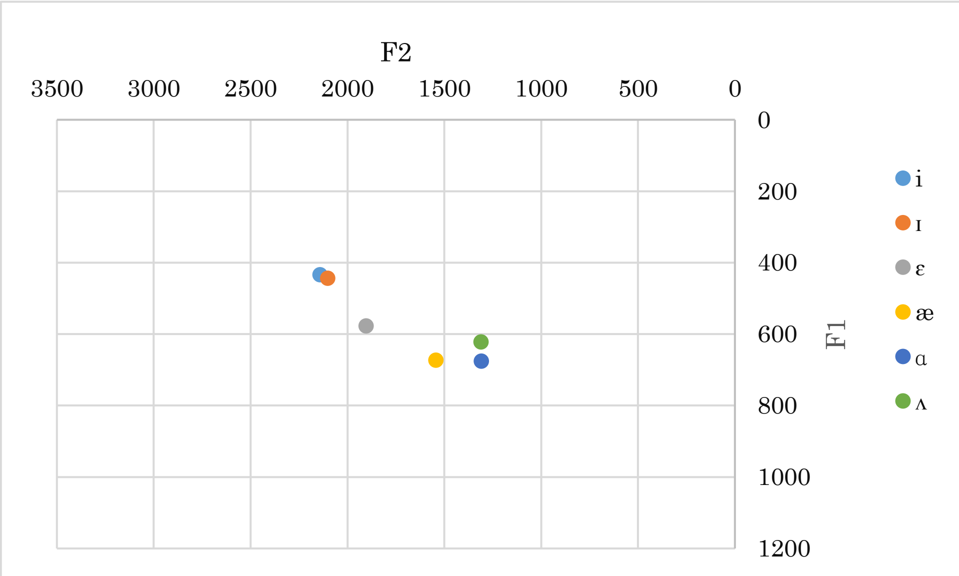

Figure 51: Mean F1 and F2 frequencies of six American English vowels uttered by male participants (Repeated after Talker 4)

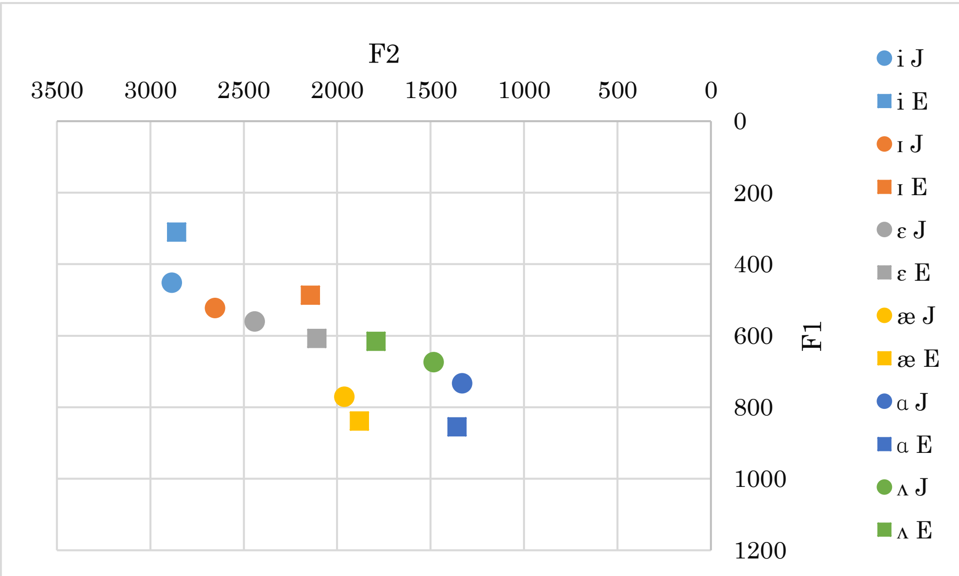

Figure 52: Mean F1 and F2 frequencies of six American English vowels uttered by female participants (Repeated after Talker 5): ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 5.

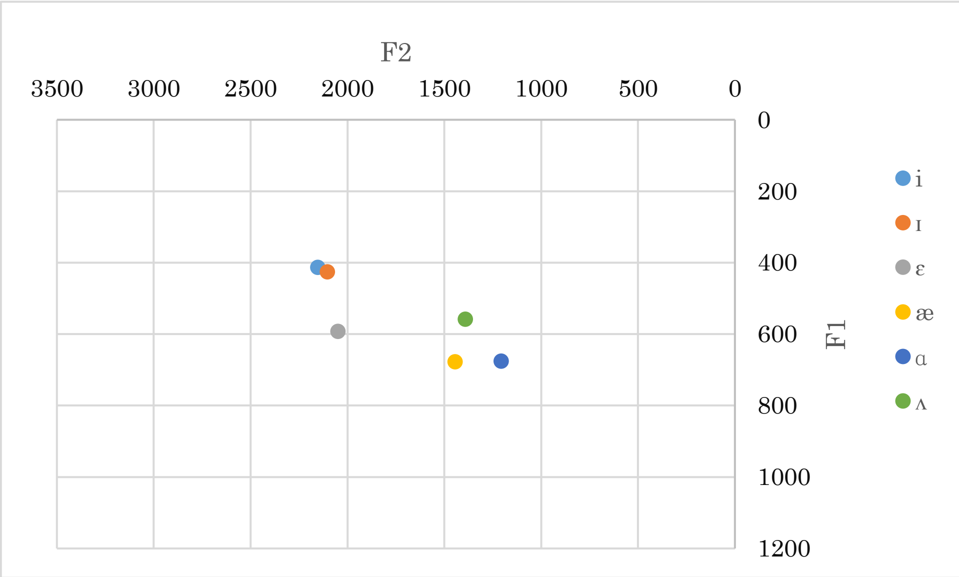

Figure 53: Mean F1 and F2 frequencies of six American English vowels uttered by 5 male participants (Repeated after Talker 5)

Figure 54 shows mean duration of six vowels uttered by 33 Japanese participants in /hVt/ context. The mean duration in general reflects the native speakers' tokens that the participants repeated (Compare Figures 17 and 54). Mean duration of each vowel that participants produced is longest when the vowel was repeated after a native speaker whose token is longer than that of the other native speakers.

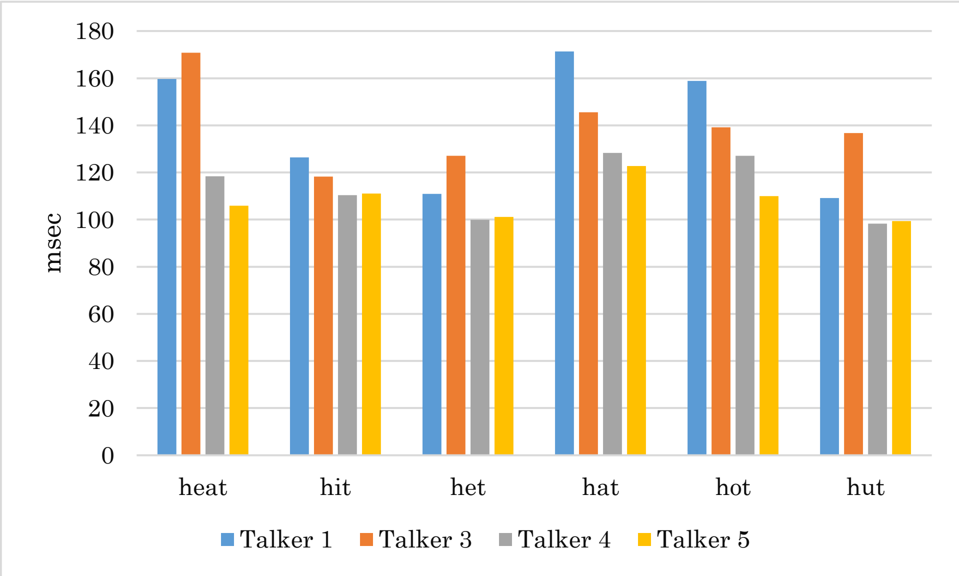

Figure 54. Mean vowel duration in /hVt/ context uttered by Japanese participants in "Repetition" condition

6.2.2.2 /pVt/ context

As in the context of /hVt/, native speakers generally use a vertically larger vowel space than do Japanese speakers. Native speakers' /i/ tokens are higher, and their /æ/ /ɑ/ tokens are lower than those of Japanese speakers. Japanese speakers' /i/ and /ɪ/ tokens are spectrally more separate than in "Read Aloud" Condition, but as in the context of /hVt/, Japanese speakers' /i/ and /ɪ/ tokens are not as separate as those of native speakers.

Native speakers' /ɛ/ tokens are more centralized than those of Japanese speakers, with the exception of the token of the /ɛ/ token of Talker 1. One particular native speaker had a more centralized /ɛ/ token than that of the other native speakers, and Japanese speakers seem to have produced more centralized /ɛ/ when they repeated after this native speaker. This particular talker's /ɛ/ is classified as Japanese /e/ only in 22.7% of instances. Perceptual assimilation of /ɛ/ to Japanese /u/ in the context of /pVt/ is mostly attributed to this talker's /ɛ/, as shown in Table 10.

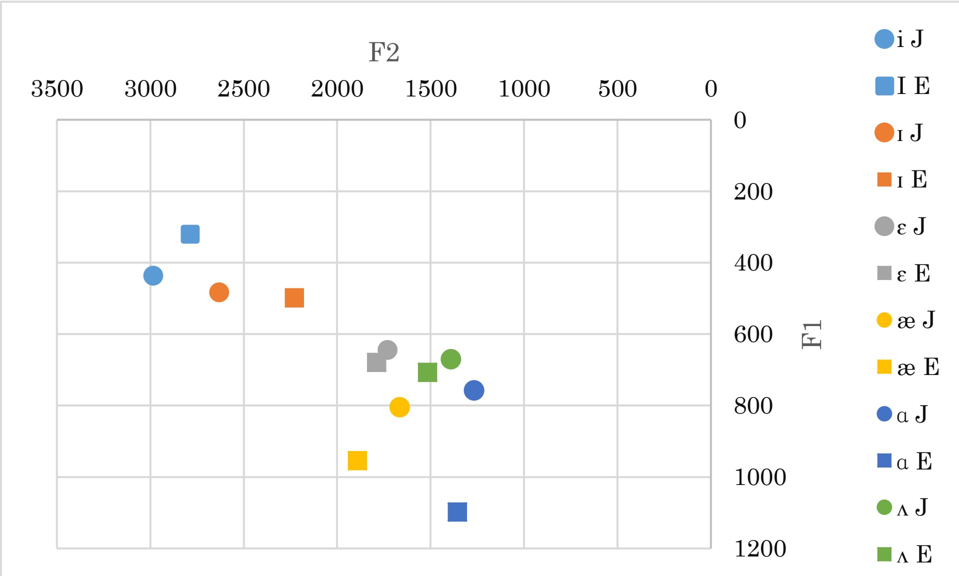

Figure 55. Mean F1 and F2 frequencies of six American English vowels uttered in /pVt/ context by female participants (Repeated after Talker 1

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 1.

.

Japanese speakers' /æ/ tokens are more centralized than those of native speakers, with the exception of Talker 5's token, but, still, Japanese speakers' /æ/ is somewhat fronted compared with "Read Aloud" condition. Japanese speakers' /æ/ uttered after hearing Talker 5 is even more fronted and close to their /ɛ/ tokens. This reflects the results of the Perceptual Assimilation Experiment in which Talker 5's /æ/ is equated with Japanese /e/ in 50% of instances.

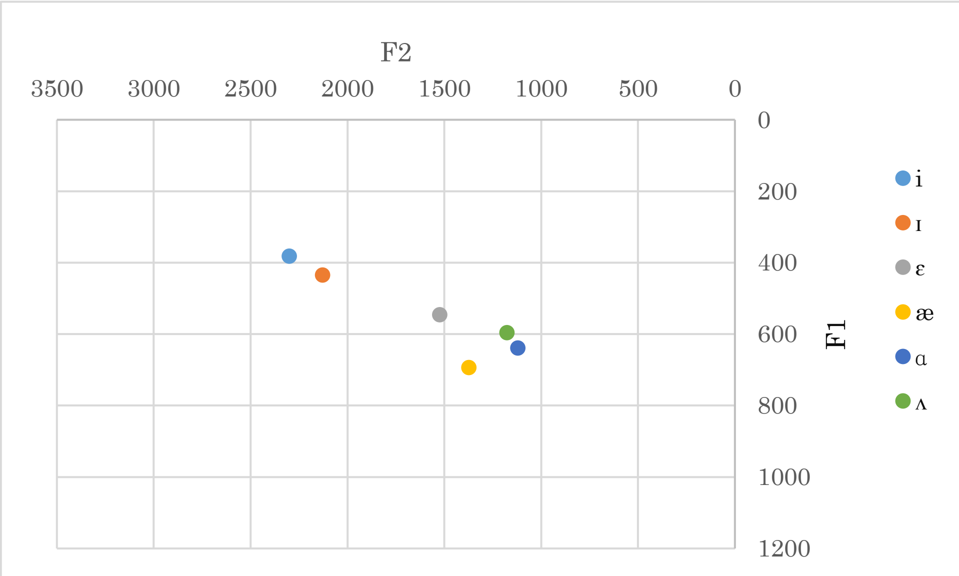

Figure 56. Mean F1 and F2 frequencies of six American English vowels uttered in /pVt/ context by five male participants (Repeated after Talker 1)

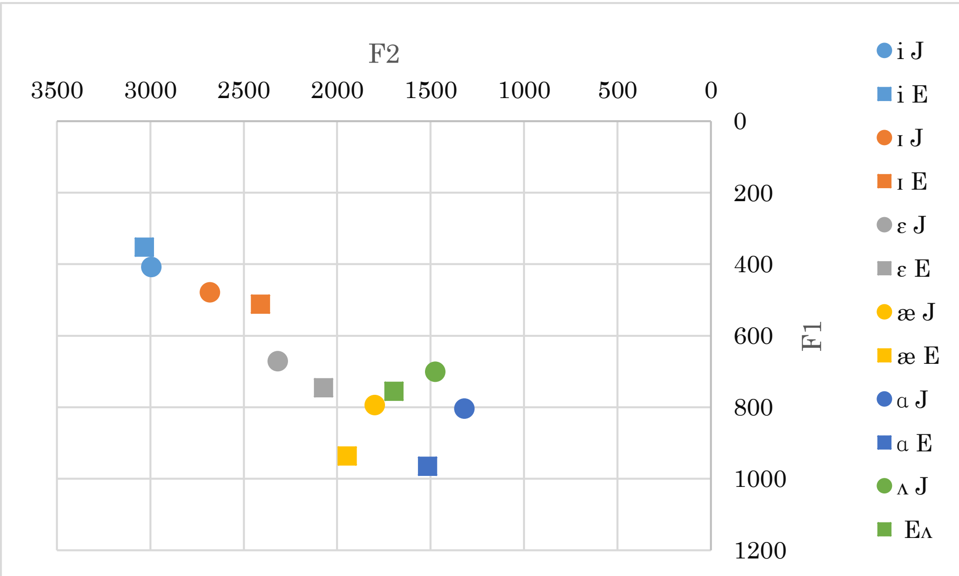

Figure 57. Mean F1 and F2 frequencies of six American English vowels uttered by female participants (Repeated after Talker 3).

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 3.

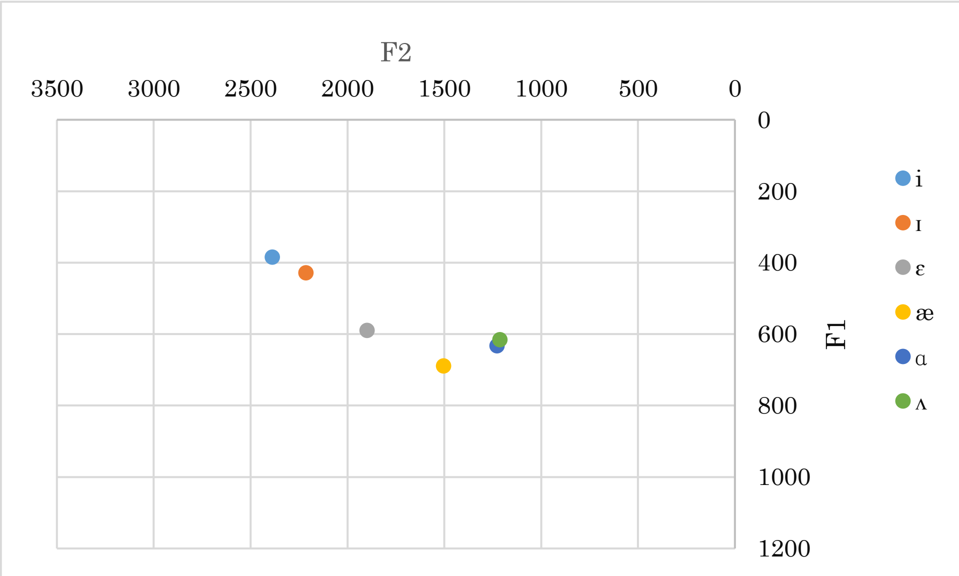

Figure 58. Mean F1 and F2 frequencies of six American English vowels uttered in /pVt/ context by male participants (Repeated after Talker 3)

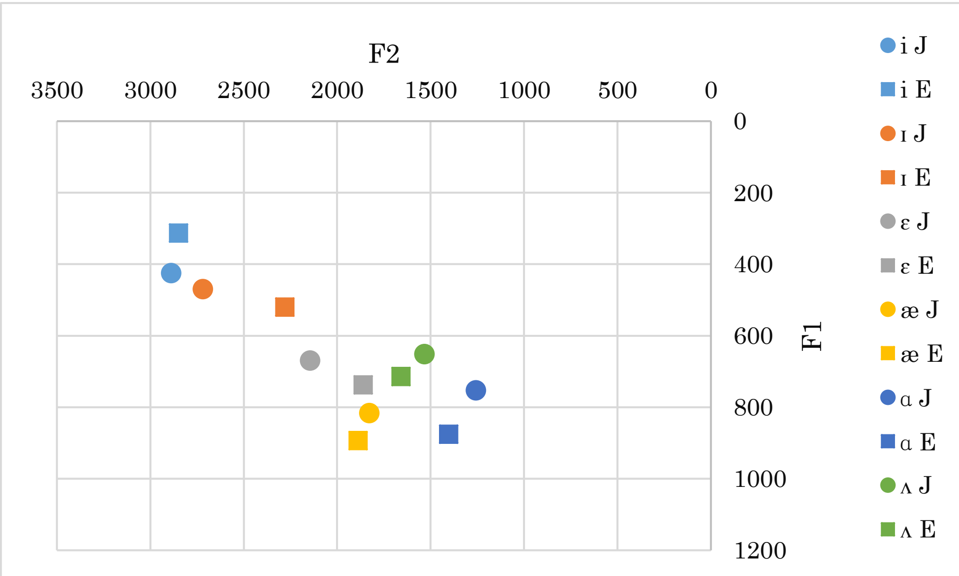

Figure 59: Mean F1 and F2 frequencies of six American English vowels uttered in /pVt/ context by female participants (Repeated after Talker 4): ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 4.

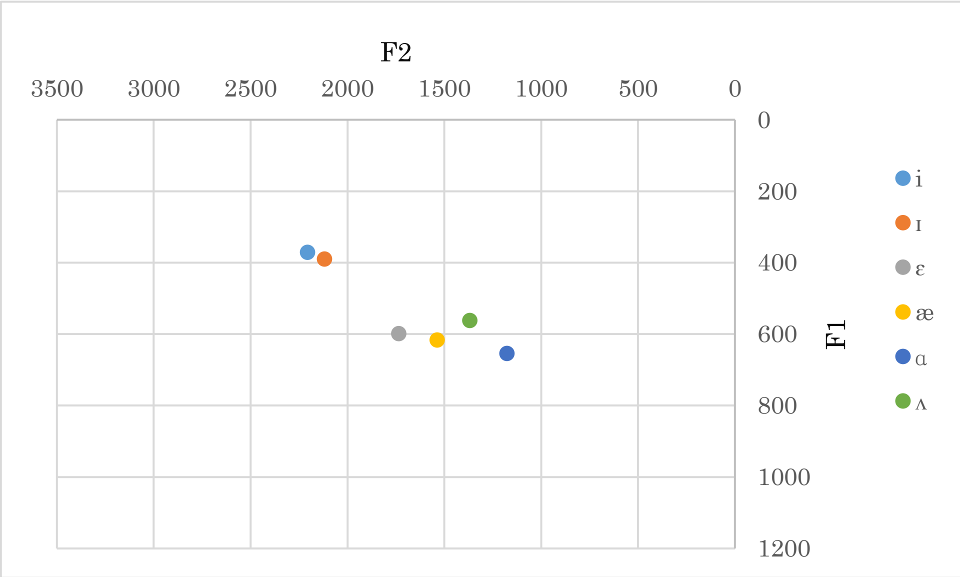

Figure 60. Mean F1 and F2 frequencies of six American English vowels uttered in /pVt/ context by male participants (Repeated after Talker 4)

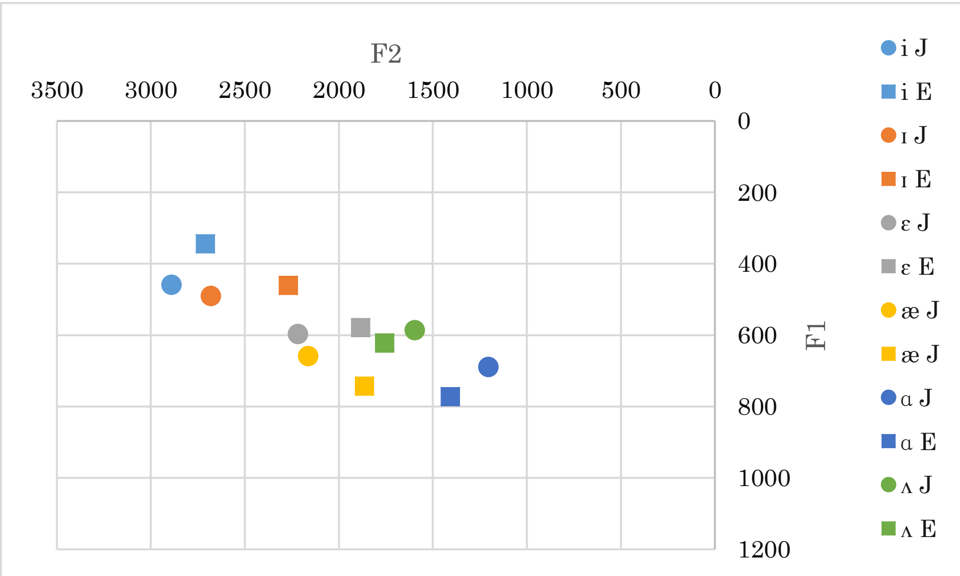

Figure 61. Mean F1 and F2 frequencies of six American English vowels uttered n /pVt/ context by female participants (Repeated after Talker 5).

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 5.

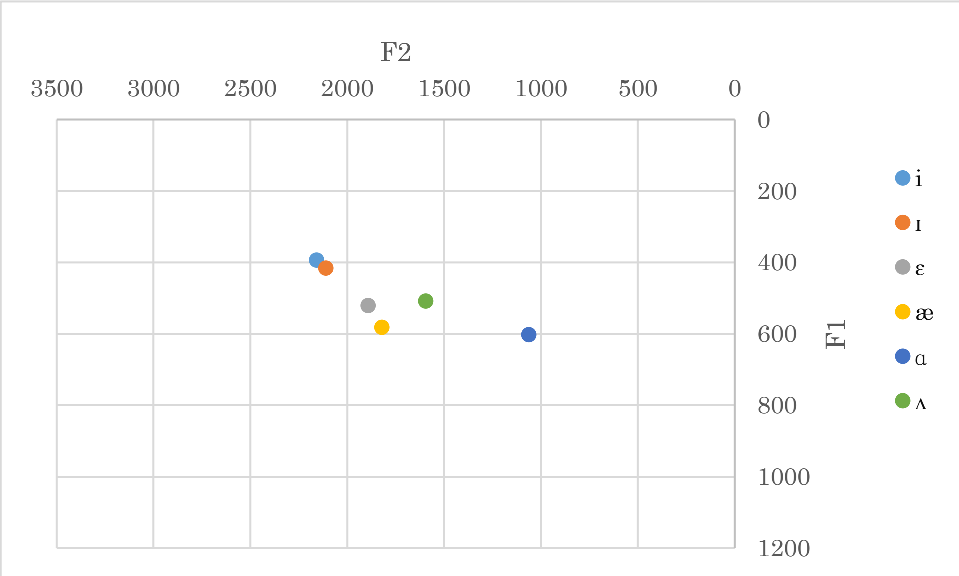

Figure 62. Mean F1 and F2 frequencies of six American English vowels uttered n /pVt/ context by male participants (Repeated after Talker 5)

Figure 63 shows mean duration of six vowels uttered by 33 Japanese participants in /pVt/ context. As in the context of /hVt/, mean duration of each vowel roughly reflects the duration of tokens that were repeated by each of the participants, but the durational difference is smaller than in the /hVt/ context, and, as for /æ/ (pat) and /ɑ/ (pot), Talker 1's tokens are the longest (See Figure 17), but Japanese participants' tokens of /æ/ and /ɑ/ are not the longest when they repeated the exact utterances of Talker 1. Durational difference between mean duration of Japanese participants' /i/ tokens (Pete) and their /ɪ/ (pit) is smaller than it is in the /hVt/ context.

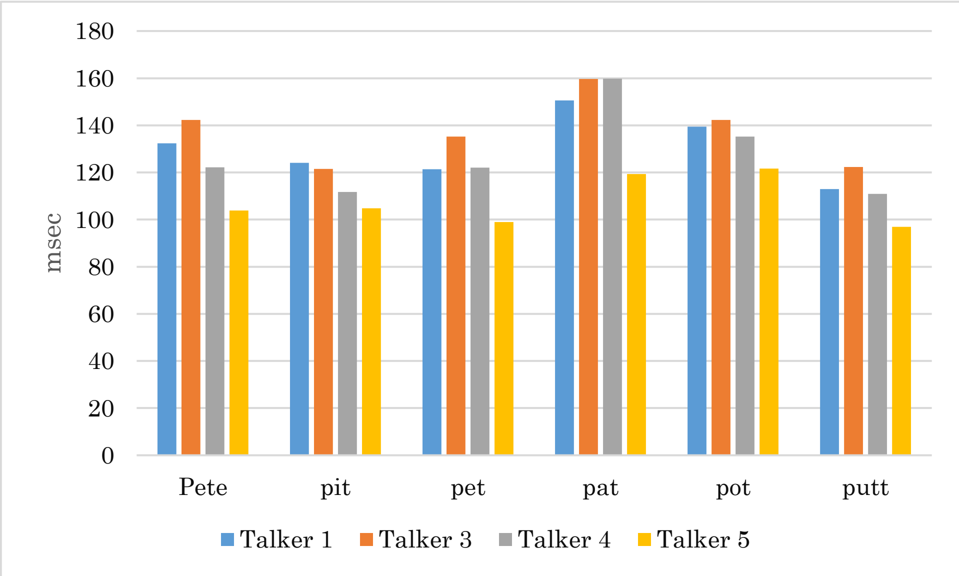

Figure 63. Mean vowel duration in /pVt/ context uttered by Japanese participants in "Repetition" condition

6.2.2.3 /pVn/ context

In /pVn/ context, larger differences in mean F1 and F2 frequencies are found between "Read Aloud" condition and "Repetition" condition than in two preplosive contexts. /i/ and /ɪ/ are more apart, /æ/ is fronted, and /ɑ/ and /ʌ/ are close. These differences are largely due to the four native speakers' utterances that were repeated by the Japanese participants. The four native speakers' /ɪ/ tokens are more apart from their respective /i/ tokens in /pVn/ context, and the Japanese participants seem to have discerned the phonetic difference, but the Japanese participants' /i/ and /ɪ/ tokens are still closer to each other than those of the four native speakers. The four native speakers' /i/ is higher than that of the Japanese participants. The native speakers use a larger vowel space.

The /æ/ of Japanese participants is fronted, and it is distant from their /æ/ in "Read Aloud" condition. Still, Japanese speakers' /æ/ is not as fronted as that of three of the native speakers. It is possible that the Japanese participants may have overreacted to Talker 5's fronted /æ/. Japanese participants' /æ/ and /ɛ/ are close, and this is especially true when they repeated after Talkers 3 and 4. The four native speakers' /ɑ/ and /ʌ/ are more apart than those of the Japanese participants. Apparently Japanese participants could not tell /ɑ/ and /ʌ/ apart when they heard these two vowels.

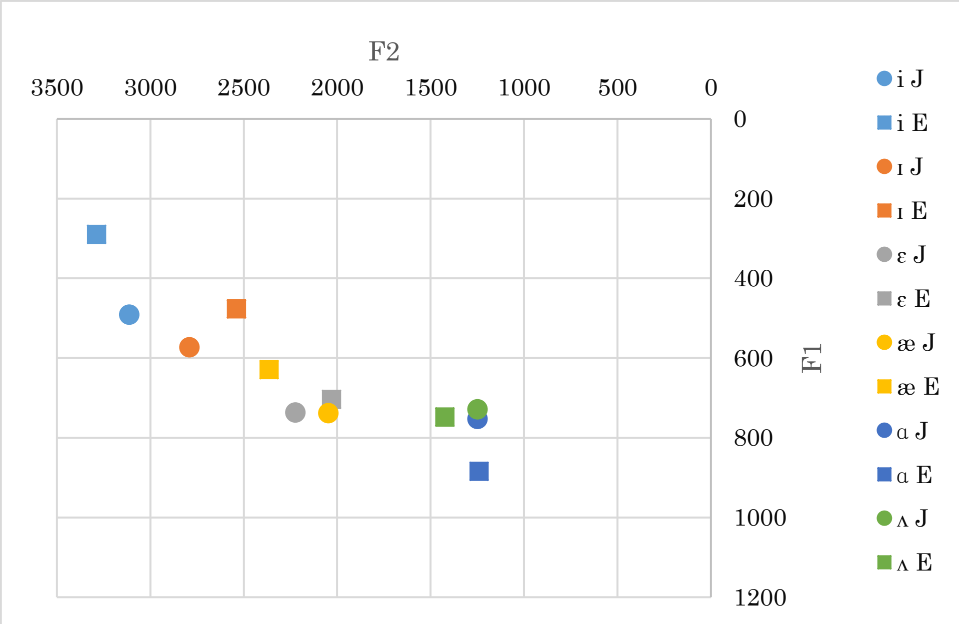

Figure 64. Mean F1 and F2 frequencies of six American English vowels in /pVn/ context uttered by female participants (Repeated after Talker 1).

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 1.

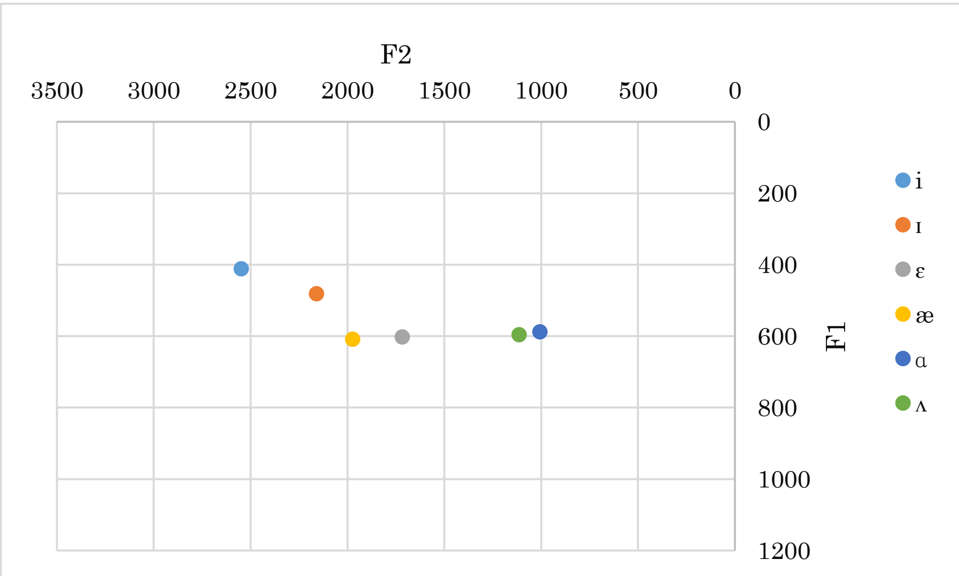

Figure 65. Mean F1 and F2 frequencies of six American English vowels in /pVn/ context uttered by male participants (Repeated after Talker 1)

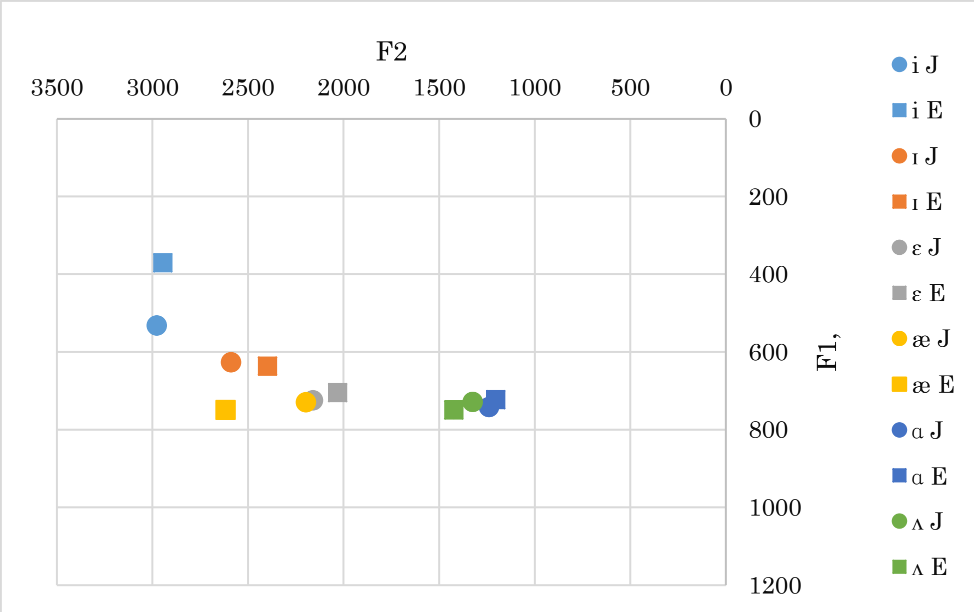

Figure 66. Mean F1 and F2 frequencies of six American English vowels in /pVn/ context uttered by female participants (Repeated after Talker 3).

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 3.

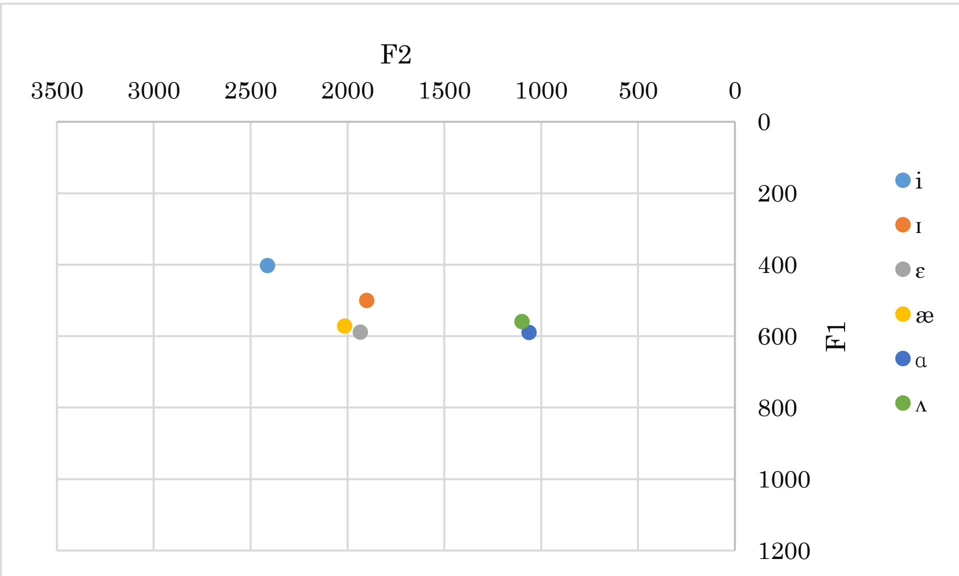

Figure 67. Mean F1 and F2 frequencies of six American English vowels in /pVn/ context uttered by male participants (Repeated after Talker 3)

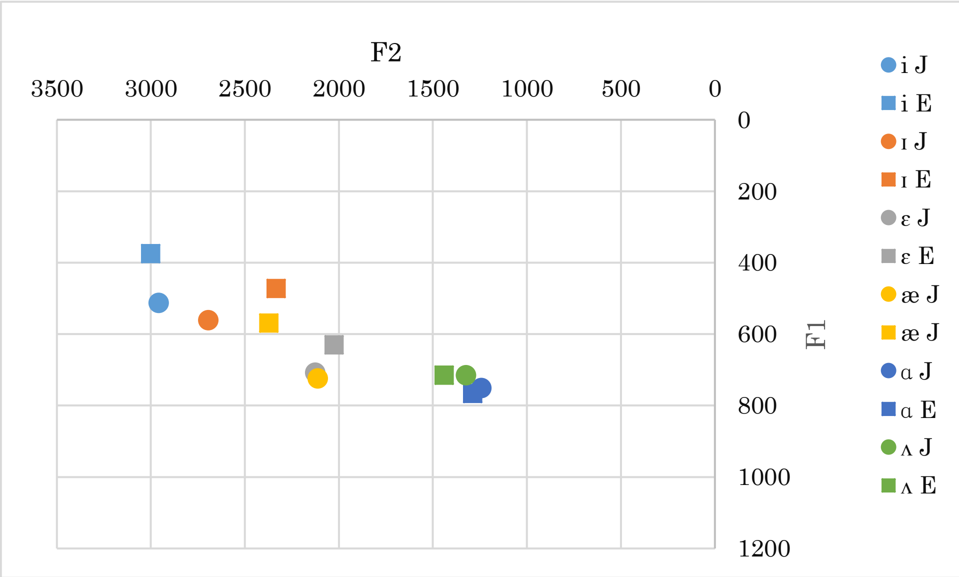

Figure 68. Mean F1 and F2 frequencies of six American English vowels in /pVn/ context uttered by female participants (Repeated after Talker 4).

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 4.

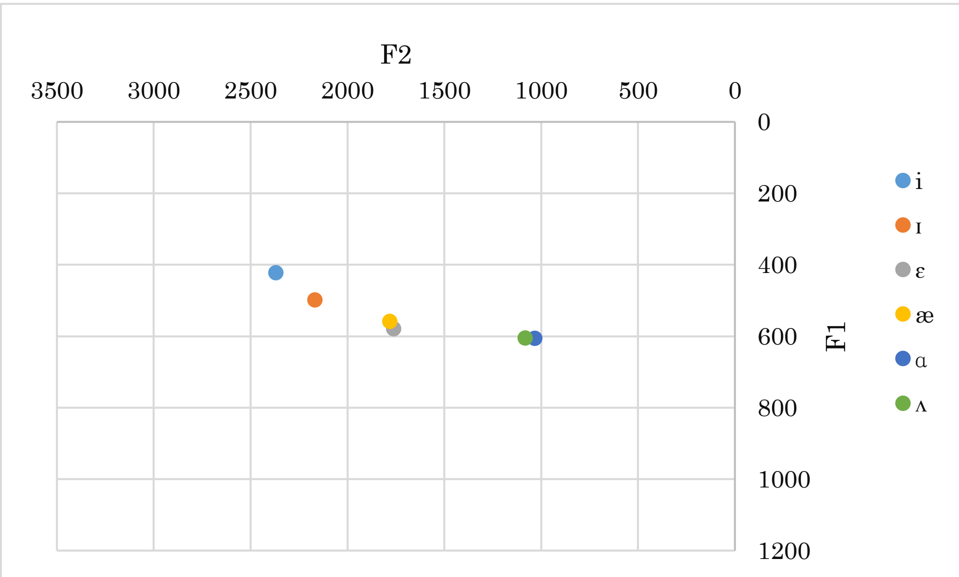

Figure 69. Mean F1 and F2 frequencies of six American English vowels in /pVn/ context uttered by male participants (Repeated after Talker 4)

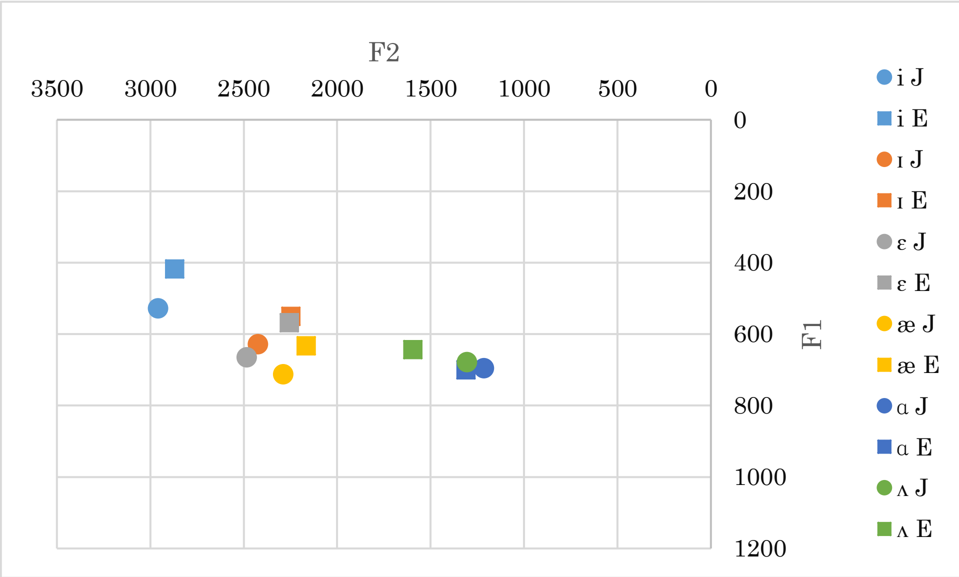

Figure 70. Mean F1 and F2 frequencies of six American English vowels in /pVn/ context uttered by female participants (Repeated after Talker 5).

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 5.

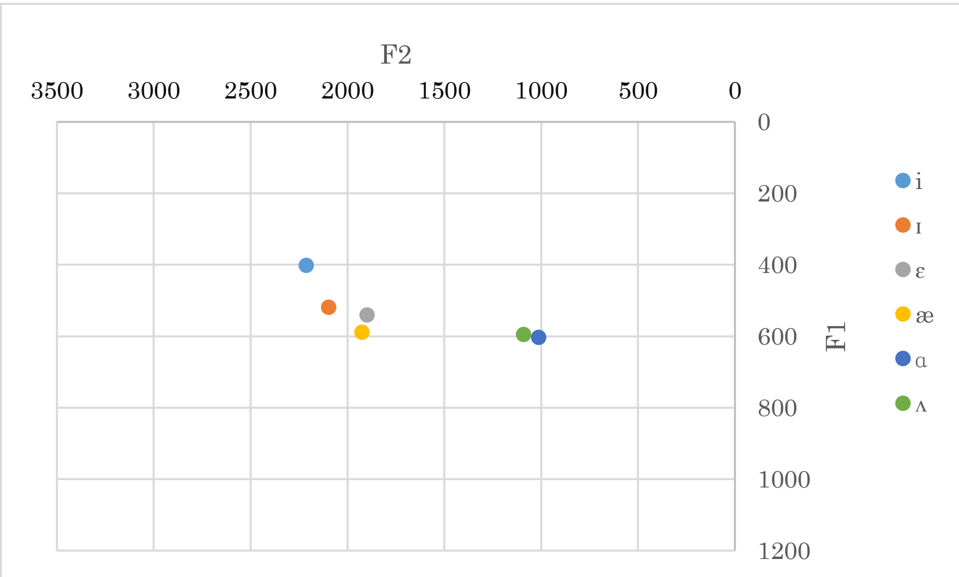

Figure 71. Mean F1 and F2 frequencies of six American English vowels in /pVn/ context uttered by male participants (Repeated after Talker 5)

Vowels are longer in /pVn/ context than in two preplosive contexts, as can be seen in "Read Aloud" condition. In "Read Aloud" condition, /i/ is longer than the other vowels by 150 milliseconds (see Table 39). In “Repetition” condition, durational difference between /i/ and the other vowels is much smaller, and, instead, /æ/ is apparently longer than the other vowels. Relative durational differences on the whole reflect the durational differences among utterances made by the four native speakers whose utterances were repeated by the Japanese participants.

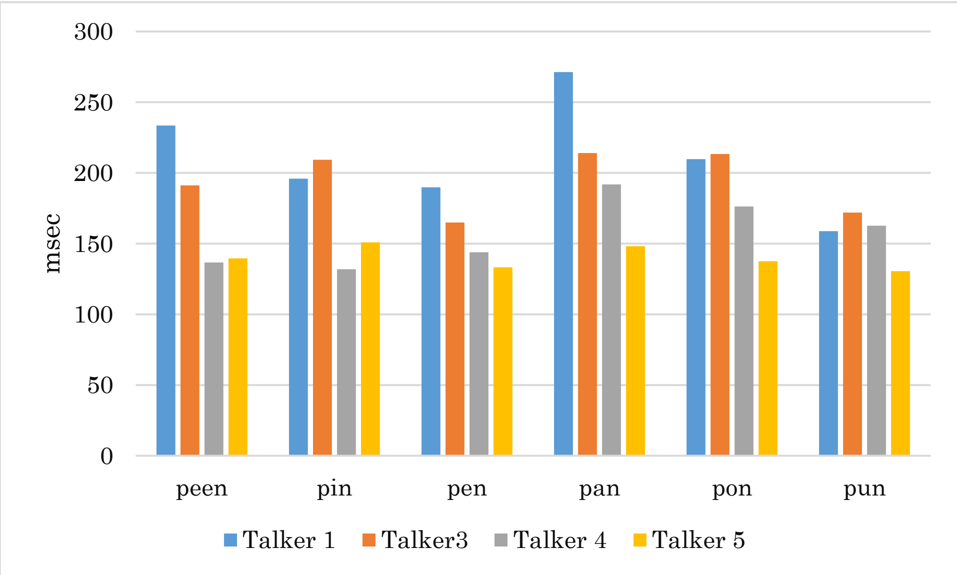

Figure 72. Mean vowel duration in /pVn/ context uttered by Japanese participants in “Repetition” condition

6.2.2.4 /pVl/ context

As with “Read Aloud” condition, the frequencies of F1 and F2 are measured at 25%, 50% and 75% of /Vl/ continua because it is difficult to set the boundary between a vowel and the following /l/. Of these three points, characteristics of each vowel can be more vividly seen at 25%. At backer points, a vowel is more strongly /l/-colored.

Figures 73-77 show mean frequencies of F1 and F2 at the 25% of /Vl/ continua uttered by 28 female participants. The F1 and F2 frequencies of /Vl/ continua uttered by four native speakers are shown in squares. As in the other contexts, native speakers use larger vowel space than do Japanese speakers.

Also similar to the other context, /i/ and /ɪ/ are spectrally more distant than in “Read Aloud” condition, but Japanese speakers’ /i/ and /ɪ/ are not as distant as those uttered by native speakers. Another noticeable difference between Japanese participants and native speakers' utterances is that the vowels that native speakers uttered are generally more retracted than those produced by Japanese participants. The retraction is most visible in /ɪ/ and /ɛ/. These two vowels that native speakers produced are horizontally distant from those Japanese speakers produced.

Japanese speakers' production of /æ/ differs depending on which native speaker they are repeating for the experiment. When they repeated after Talkers 1 and 3, their /æ/ was backer and centralized, but when they repeated after Talkers 4 and 5, their /æ/ was fronted. Japanese speakers' /ɑ/ and /ʌ/ are close, indicating that these two vowels are perceptually close.

A similar tendency is observed in male participants' utterances (Figure 72). Their /ɑ/ and /ʌ/ are even closer than the Japanese speakers.

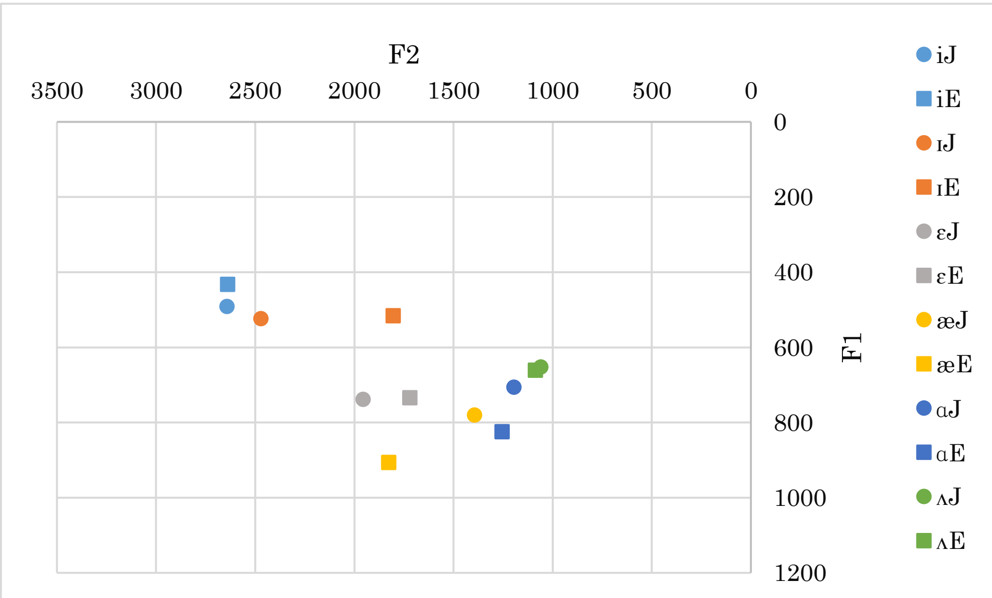

Figure 73. Mean F1 and F2 frequencies at 25% of /pVl/ continua uttered by female participants in "Repetition" condition (Repeated after Talker 1)

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 1.

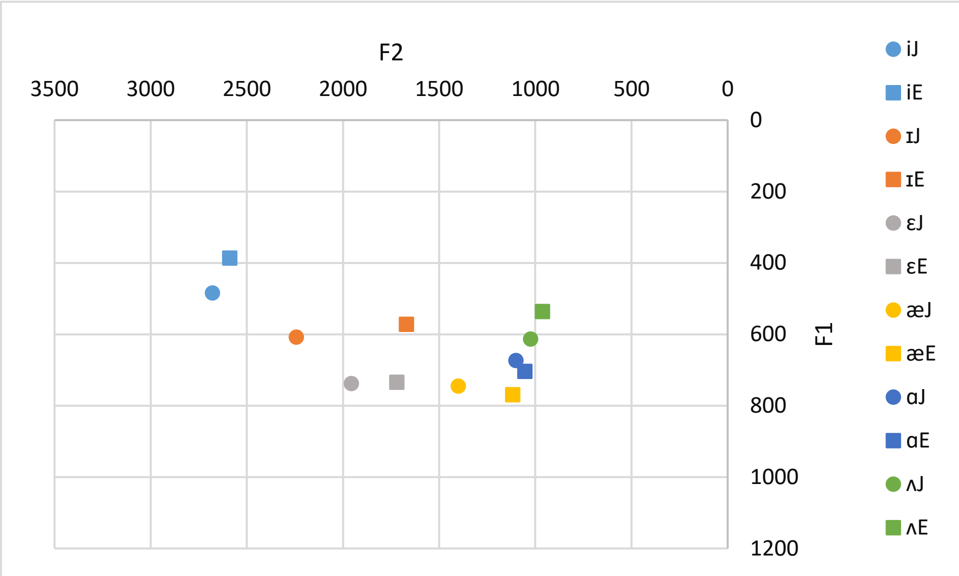

Figure 74. Mean F1 and F2 frequencies at 25% of /pVl/ continua uttered by female participants in "Repetition" condition (Repeated after Talker 3)

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 3.

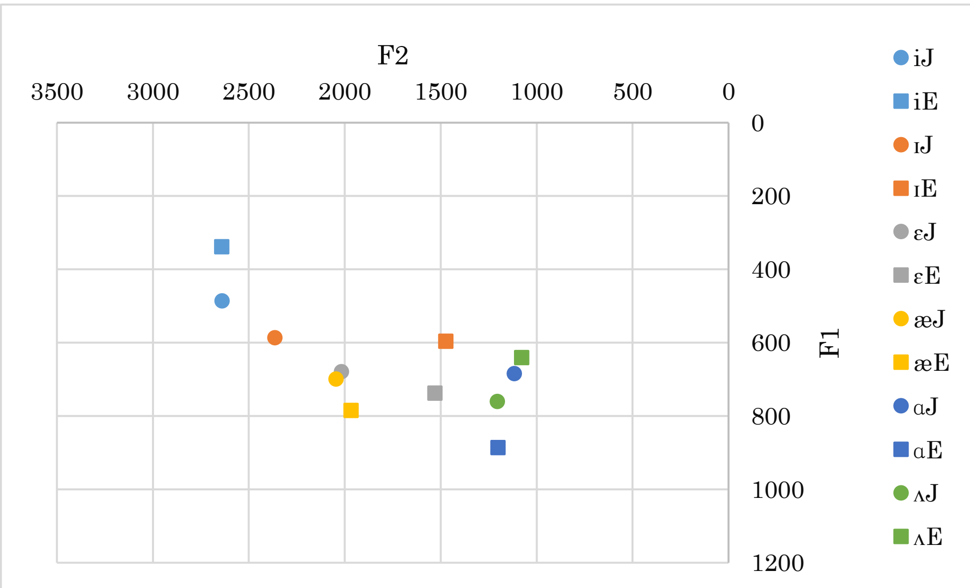

Figure 75. Mean F1 and F2 frequencies at 25% of /pVl/ continua uttered by female participants in "Repetition" condition (Repeated after Talker 4)

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 4.

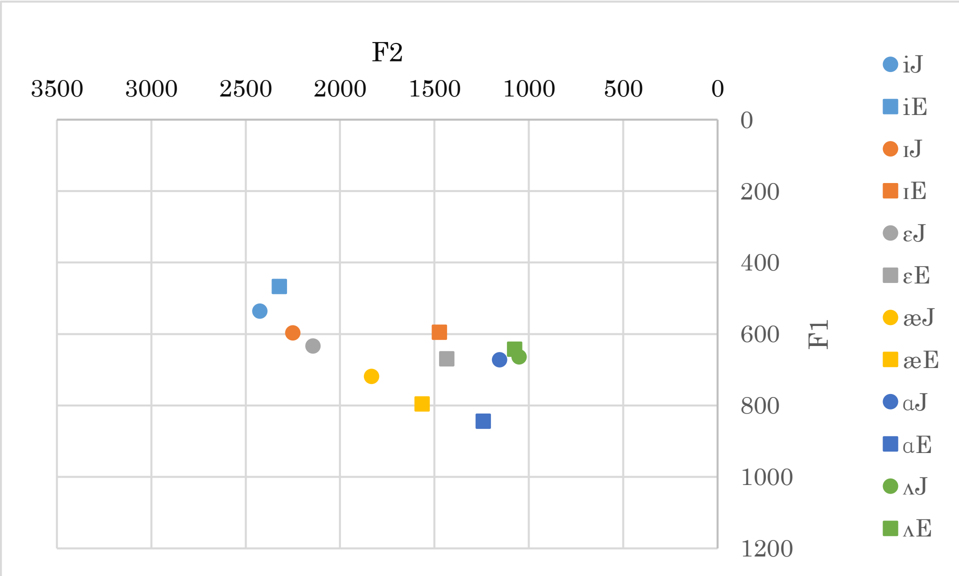

Figure 76. Mean F1 and F2 frequencies at 25% of /pVl/ continua uttered by female participants in “Repetition” condition (Repeated after Talker 5)

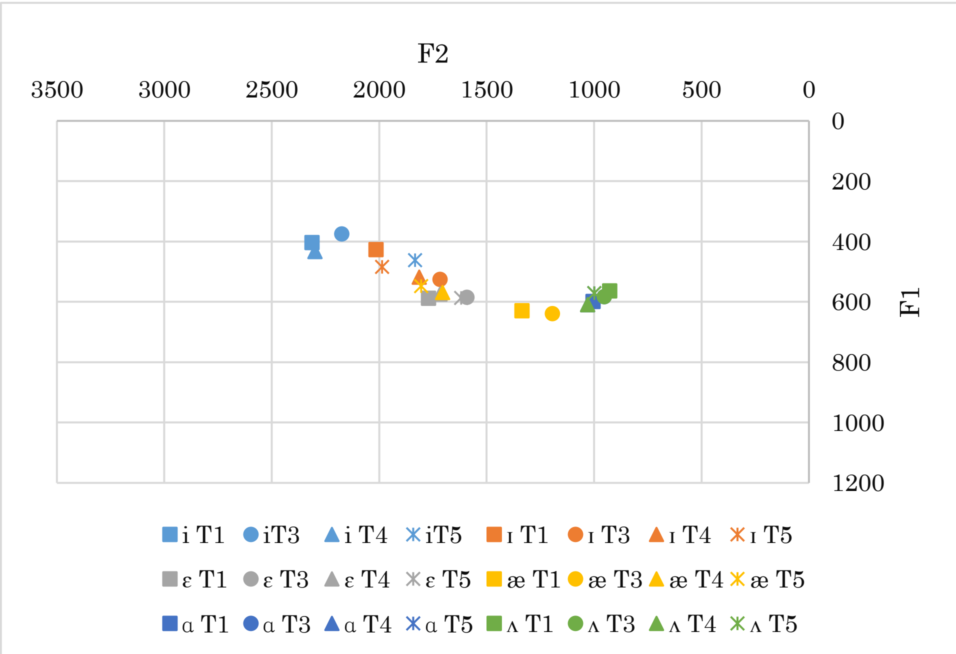

Figure 77. Mean F1 and F2 frequencies at 25% of /pVl/ continua uttered by male Japanese participants in “Repetition” condition

Note. ■: utterances repeated after Talker 1, ●: utterances repeated after Talker 3, ▲: utterances repeated after Talker 4, ＊: utterances repeated after Talker 5)

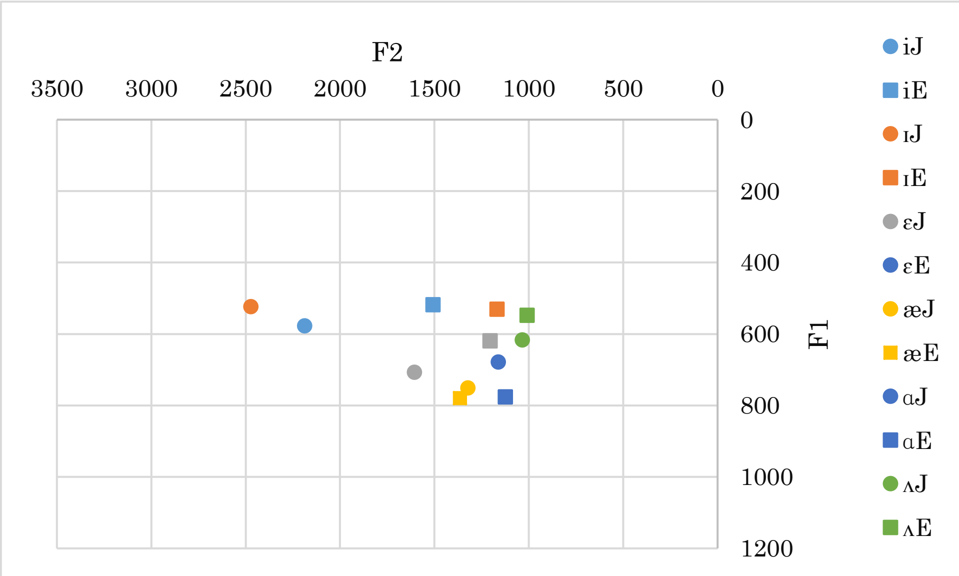

Figure 78. Mean F1 and F2 frequencies at 50% of /pVl/ continua uttered by female participants in "Repetition" condition (Repeated after Talker 1)

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 1.

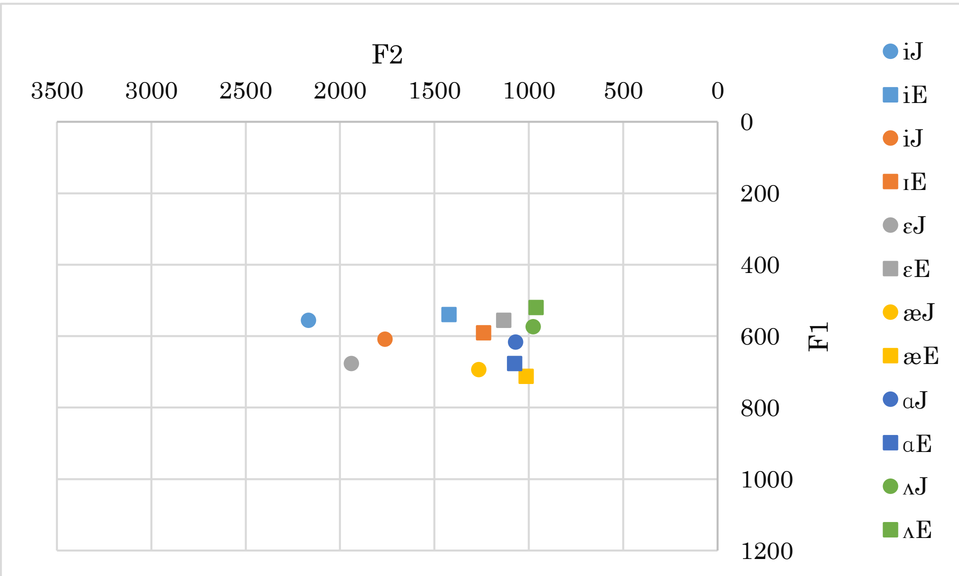

Figure 79. Mean F1 and F2 frequencies at 50% of /pVl/ continua uttered by female participants in "Repetition" condition (Repeated after Talker 3)

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 3.

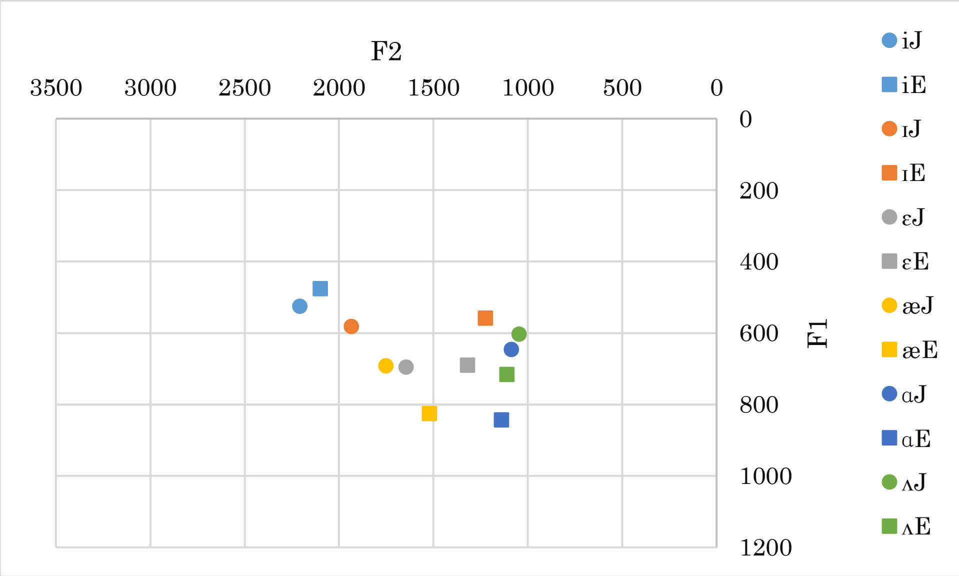

Figure 80. Mean F1 and F2 frequencies at 50% of /pVl/ continua uttered by female participants in “Repetition” condition (Repeated after Talker 4)

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 4.

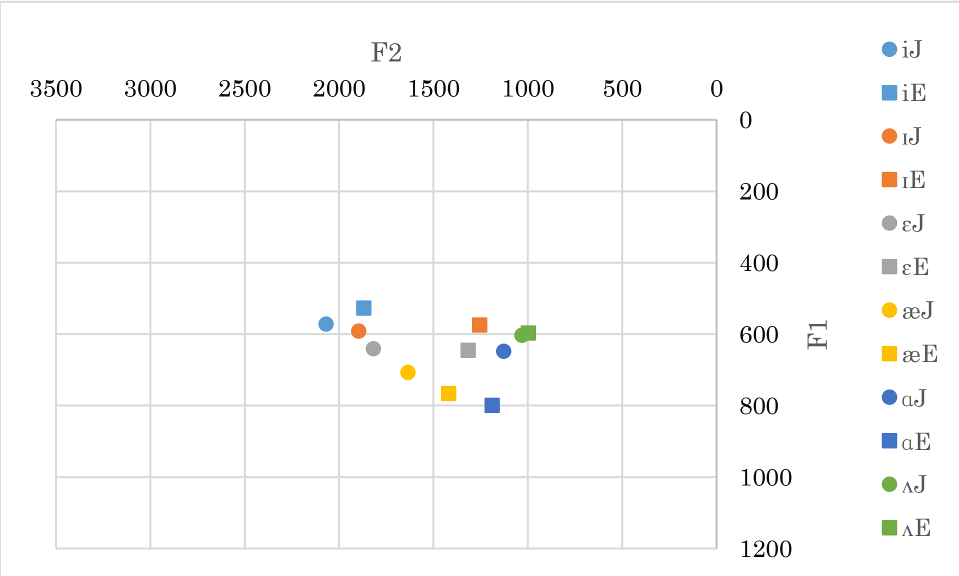

Figure 81. Mean F1 and F2 frequencies at 50% of /pVl/ continua uttered by Japanese participants in “Repetition” condition (Repeated after Talker 5)

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 5.

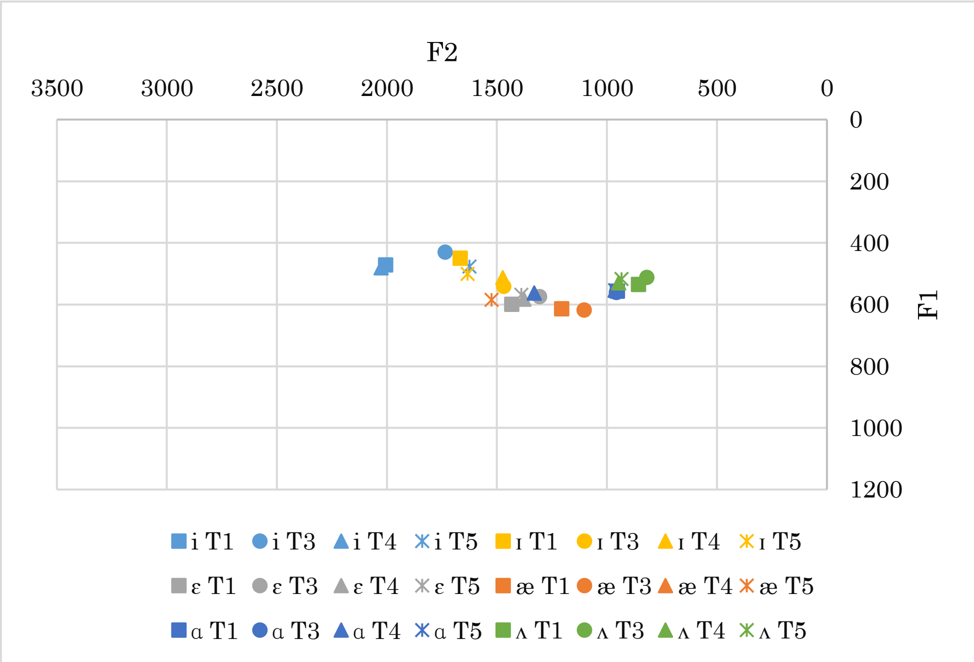

Figure 82. Mean F1 and F2 frequencies at 50% of /pVl/ continua uttered by male Japanese participants in "Repetition" condition (

Note. ■: utterances repeated after Talker 1, ●: utterances repeated after Talker 3, ▲: utterances repeated after Talker 4, ＊: utterances repeated after Talker 5)

At 75% of continua, F2 frequencies of most of native speakers' utterances are around 1000Hz, but F2 frequencies of Japanese participants' /i/, /ɪ/, and /ɛ/ are higher (around 1500hz)

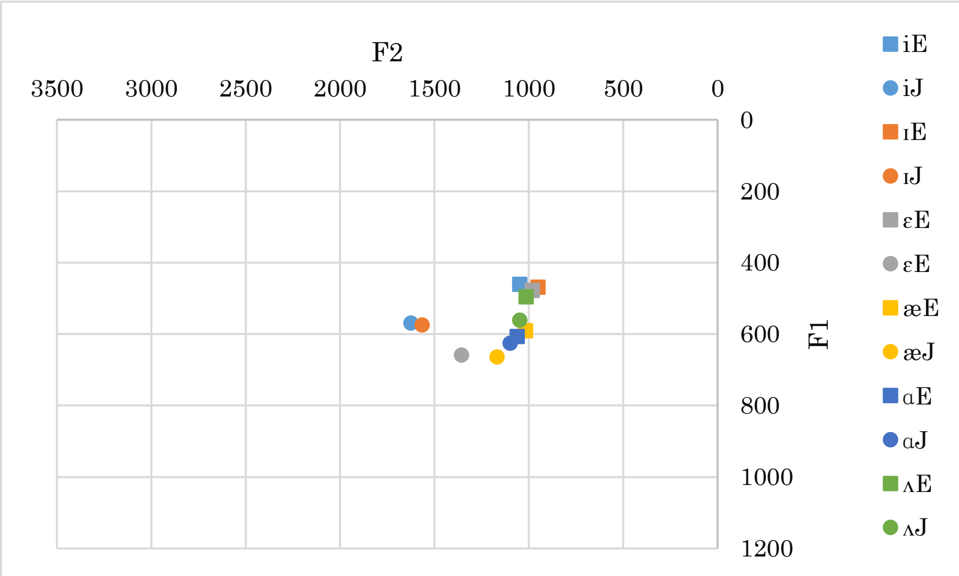

Figure 83. Mean F1 and F2 frequencies at 75% of /pVl/ continua uttered by female participants in "Repetition" condition (Repeated after Talker 1)

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 1.

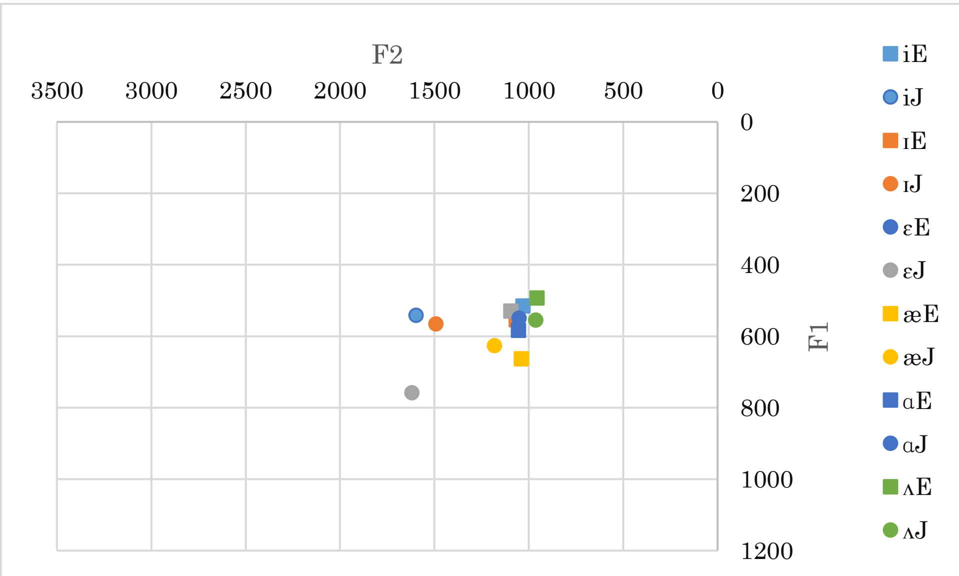

Figure 84. Mean F1 and F2 frequencies at 75% of /pVl/ continua uttered by female participants in "Repetition" condition (Repeated after Talker 3)

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 3.

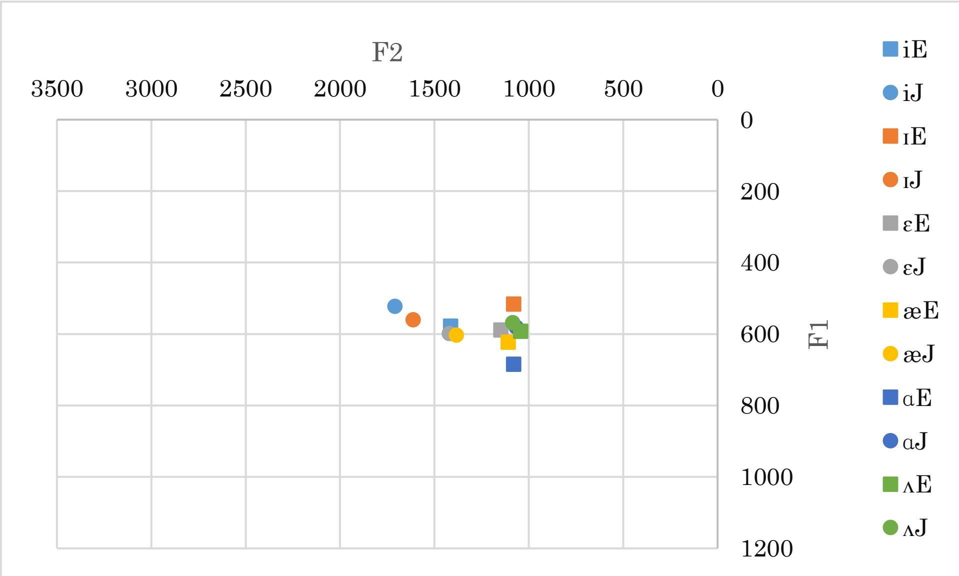

Figure 85 Mean F1 and F2 frequencies at 75% of /pVl/ continua uttered by female participants in "Repetition" condition (Repeated after Talker 4)

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 4.

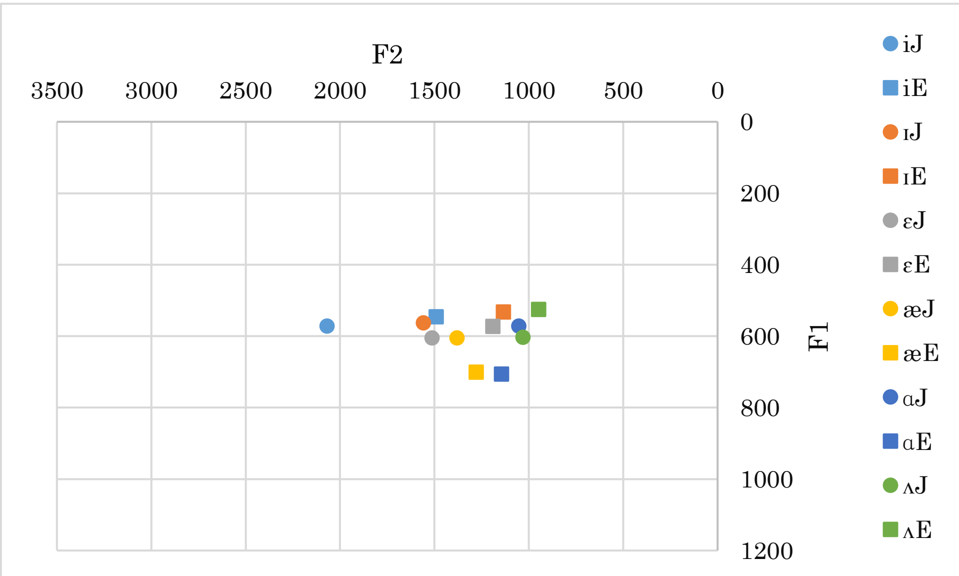

Figure 86. Mean F1 and F2 frequencies at 75% of /pVl/ continua uttered by female participants in "Repetition" condition (Repeated after Talker 5)

Note. ● indicates mean F1 and F2 of vowels uttered by Japanese participants. ■ indicates F1 and F2 frequencies of vowels uttered by Talker 5.

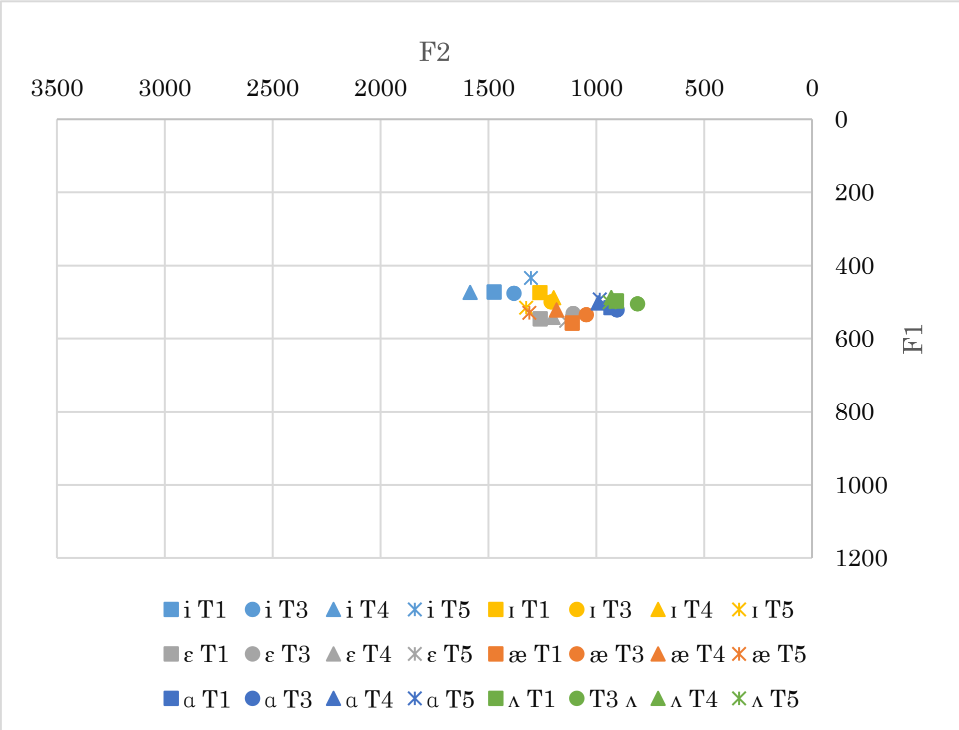

Figure 87: Mean F1 and F2 frequencies at 50% of /pVl/ continua uttered by male participants in "Repetition" condition (■: utterances repeated after Talker 1, ●: utterances repeated after Talker 3, ▲: utterances repeated after Talker 4, ＊: utterances repeated after Talker 5)

7. Japanese vowel categories and the perception of American English vowels

In this chapter, I will discuss how Japanese vowel categories affected the perception of American English vowels by native Japanese speakers by comparing the results of the experiments discussed above.

In Chapter 1, I pointed out that inadequate and insufficient instruction on pronunciation, combined with the massive influx of loanwords from English, together can help form the image of English phones and can hinder the acquisition of authentic pronunciation. In Chapter 6, Japanese participants' production of English vowels was similar to the Japanese adaptation of English vowels. The vowel /i/ was remarkably longer than the other five vowels. The sounds /i/ and /ɪ/ are qualitatively close, and are close to Japanese /i/. /ɛ/ is between Japanese /i/ and /a/, /æ/ and /ʌ/ are in the vicinity of

Japanese /a/, and /ɑ/ is backer and higher than /æ/ and /ʌ /. It can be assumed, then, that Japanese participants' images of English vowels reflect the Japanese adaptation of English vowels.

To verify this assumption, a survey is conducted of university students in Japan on English pronunciation. All the questions and answers are done in Japanese, and 44 students answer the survey. The students responded in their own words. Among the questions is "how do you think the pronunciation of 'beat' is different from that of 'bit'?" All of the students but one refer to the durational difference, and none of them mention qualitative differences of the vowels. In addition to the difference in vowel length, 12 of the students mention the insertion of moraic obstruent /Q/ between the vowel and the syllable-final /t/ in "bit."

Another question in the survey is, "how do you think the pronunciation of 'bat' is different from that of 'but'?" As seen in Chapter 1, both /æ/ and /ʌ/ are adapted as /a/. The answers revealed that nine of the students responded that /æ/ is between Japanese /a/ and /e/, which implies that these students are aware that /æ/ is a front vowel. Another nine students respond either that they have no idea if there is a difference, or that there is no difference between the pronunciations of these two words. The rest of the students respond to the question of how they perceive the differences in these two vowels, but those are irrelevant to this study.

Additionally, the survey includes the question, "how do you think the pronunciation of 'hot' is different from that of 'hut'?" Twenty-seven of the students respond "hot" has [o], and only two of them mention that "hot" has a vowel close to [a]. Three of the students respond that "hut" has a vowel like [u], and four of them respond either that they had no idea or that these two vowels are the same.

Kori (2018) conducts a similar survey of university students in which a multiple-choice format is adopted. In one question, the students choose the closest Japanese vowel to the vowel in "hut," and [a] is the most frequently chosen, but half of them reply that they have no idea as to why. Lexical familiarity may have obscured the results of the two surveys, but Japanese speakers may not have as clear an image of /ʌ/ as that of other English vowels. Previous studies reveal that, in general, Japanese speakers identify /ʌ/

less accurately than other English vowels (Nozawa & Wayland 2012, Nozawa 2016, Nozawa & Cheon 2016).

A previous study by Nozawa (2016) reveals that Japanese speakers identify the American English /i/, /ɪ/, /ɛ/, and /æ/ better than those of New Zealand English, but they identify American English /ɑ/ less accurately than New Zealand English /ɒ/ (New Zealand English equivalent of American English /ɑ/). This occurs even though, overall, these Japanese speakers' identification accuracy of American English vowels is higher than that of New Zealand English vowels. New Zealand English /i/ is diphthongized, and it has a formant contour similar to that of /eɪ/. New Zealand English /ɪ/ is centralized, and it is more distant from Japanese /i/. In New Zealand English, /ɛ/ and /æ/ are both raised to the extent to be [i] and [ɛ]. These two vowels are more distant from Japanese /e/ and /a/. New Zealand English /ɒ/ is higher and closer to Japanese /o/ than is American English /ɑ/. This is yet another example to demonstrate that Japanese speakers expect /ɑ|ɒ/ to have the [o]-like quality.

These results support the hypothesis that Japanese speakers expect English vowels to sound like Japanese adaptations of English vowels. In the following sections, I will demonstrate how Japanese vowel categories and Japanese speakers' image of English vowels affected the results of the experiments described in previous chapters.

7.1 Japanese vowel categories and identification of American English vowels

7.1.1 Identification of /i/

The vowel sound /i/ is usually transcribed as long *ii* /iː/, and /i/ is identified more correctly in a context in which /i/ is equated with *ii*. Moreover, the Japanese speakers in this study produced /i/ as a longer vowel than the other vowels when they read aloud the words on the list. All these results suggest that Japanese speakers have an auditory image that English /i/ sounds like Japanese *ii*; therefore, the central question regarding vowel identification is whether /i/ is more correctly identified if it is more frequently equated with Japanese *ii*. The percentages of correct identification of 16 /i/ tokens (4 talkers × 4 consonantal contexts), and the percentages of occurrences when these 16 /i/ tokens were equated with Japanese *ii* were submitted to Spearman's rank correlation analysis, yielding coefficient ρ=.872 (p<.001). In Figure 88, 16 /i/ tokens are aligned in the order of high percentages of correct identification. Absolute duration of /i/ in each token does not seem to be able to account for the result. Vowels are longer before /n/ than before /t/, but /i/ in the context of /pVn/ are less accurately identified than /i/ in two preplosive contexts (Compare Figures 17-19). It is unclear why /i/ before /n/ is perceived shorter and is less accurately identified. One reason could be the relative duration of the vowel and the following consonant. Another reason could be that, because the vowel is nasalized, Japanese speakers hear /n/ before it really begins, and, consequently, the vowel is perceived shorter. Further research is necessary to investigate the relationship between the actual vowel duration and the perceived vowel duration.

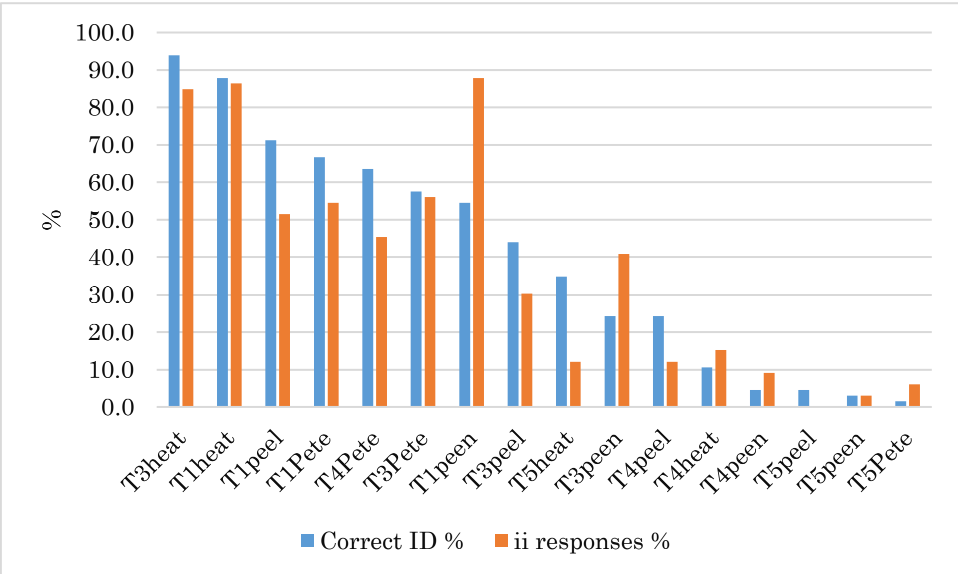

Figure 88. Percentages of correct identification of /i/ tokens (blue bars) and percentages /i/ tokens are equated with *ii* (red bars)

7.1.2 Identification of /ɪ/

The phone/ɪ/ is generally transcribed as short *i*, and it is identified better in two preplosive contexts in which it is equated with *i* more often than in the contexts of /pVn/ and /pVl/, in which /ɪ/ is equated with *i* in less than 50% of instances. Like /i/ tokens, in Figure 81, 16 /ɪ/ tokens are aligned in the order of the high percentage of correct identification along with the percentages of instances that each /ɪ/ token is equated with *i*. The percentages of correct identification of 16 /ɪ/ tokens (4 talkers × 4 consonantal contexts), and the percentages of occurrences when these 16 /ɪ/ tokens were equated with Japanese *i* were submitted to Spearman's rank order correlation analysis, yielding coefficient ρ=.768 (p<.001). So in general, /ɪ/ is more likely to be correctly identified if it is equated with Japanese short *i*

As seen in Table 9, /ɪ/ is perceived as a less ideal exemplar of *i* than /i/ is, and in the /pVn/ and /pVl/ contexts, /ɪ/ is even more distant from the category ideal (as shown in lower category goodness ratings). The two least correctly identified /ɪ/ tokens, Talker 5 and Talker 1's "pin," are the two tokens that received the lowest category goodness ratings (2.8 and 2.7, respectively), which implies not only that fewer Japanese speakers perceived these two tokens as exemplars of Japanese *i*, but even those who equated them as exemplars of Japanese *i* perceived these tokens as more deviant from the category ideal.

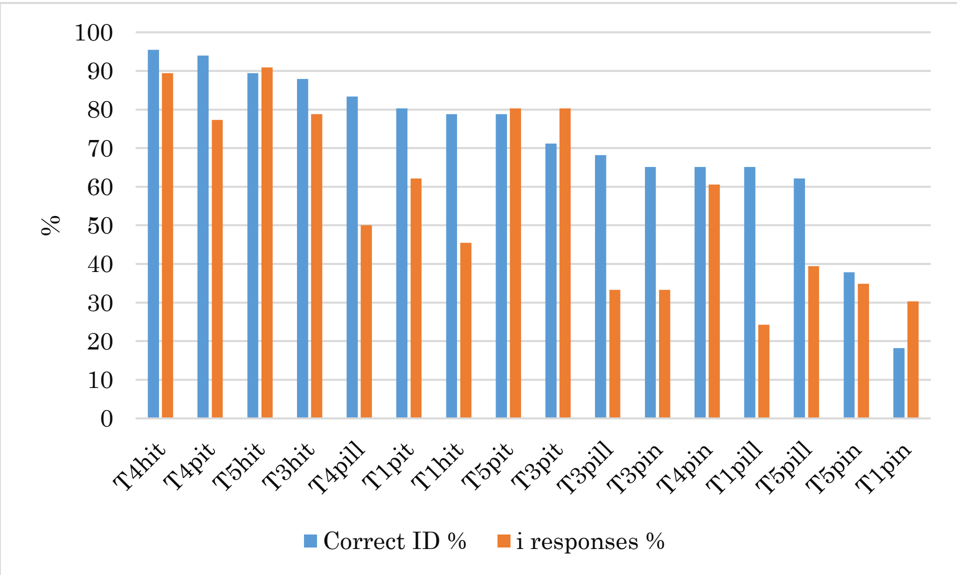

Figure 89. Percentages of correct identification of /ɪ/tokens (blue bars) and percentages /ɪ/ tokens are equated with *i* (red bars)

7.1.3 Identification of /ɛ/

The vowel /ɛ/ is the only vowel whose identification accuracy is not affected by contexts. /ɛ/ is most frequently equated with Japanese *e* in all of the contexts. Only in

/pVl/ context is /ɛ/ equated with *e* in less than 50% of instances, and the mean category goodness rating is the lowest in /pVl/ context. There is a significant difference in identification accuracy between the most accurately identified /ɛ/ token (Talker 3's het, 95.5%) and the least accurately identified /ɛ/ token (Talker 1's "pet," 30.3%). Talker 1's /ɛ/ in /pVt/ context (pet) is backer than that of the other talkers, and it is close to /ʌ/ (see Figure 12). Accordingly, Japanese participants' production of /ɛ/ is backer when they repeated the utterances of Talker 1 (See Figure 55). To see whether /ɛ/ is more correctly identified when it is equated with Japanese *e*, the percentages of correct identification of 16 /ɛ/ tokens, and the percentages that these 16 /ɛ/ tokens were equated with Japanese *e* were submitted to Spearman's rank correlation coefficient, which yielded ρ=0.421 (p=0.105). Thus, there is no significant correlation. The study also shows that some /ɛ/ tokens are frequently equated with long *ee*, most likely because of vowel duration. These tokens may not be equated with *e* because they are perceived longer and equated with *ee*. No American English vowels in this study are typically transcribed as *ee* in Japanese. The vowel /e/ as in *pay* or *cake* is not included. It is unlikely that Japanese participants rely on durational difference to identify /ɛ/, and so the percentages of instances that /ɛ/ tokens are equated with *ee* are included, and Spearman's rank correlation coefficient was calculated, which yielded a significant correlation coefficient ρ=0.615 (p<.05). Figure 82 shows the percentages of correct identification of 16 /ɛ/ tokens and the percentages /ɛ/ tokens are equated with Japanese *e* or *ee*.

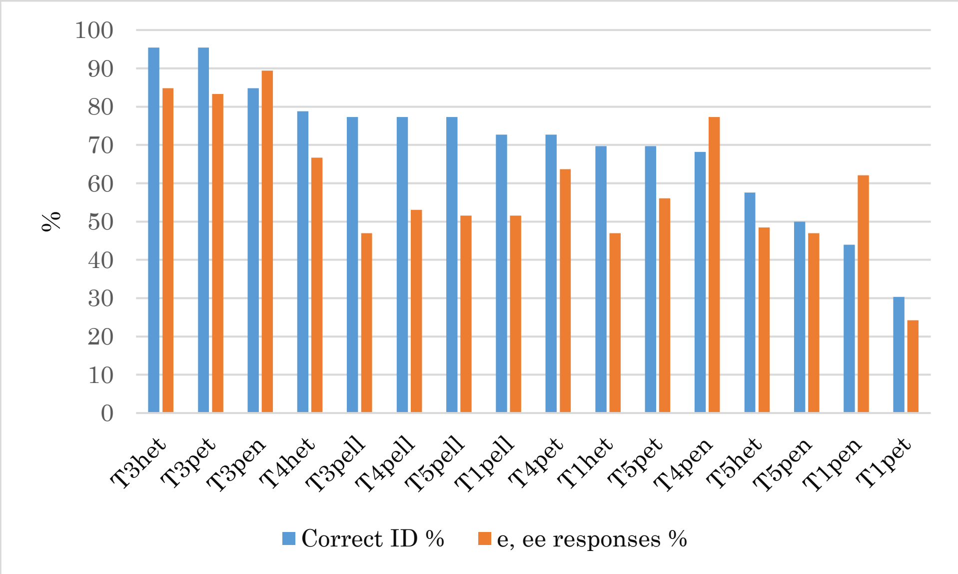

Figure 90. Percentages of correct identification of /ɛ/tokens (blue bars) and percentages /ɛ/ tokens are equated with *e* or *ee* (red bars)

7.1.4 Identification of /æ/

The vowel sound /æ/ is typically transcribed as *a* even though these two vowels are not really spectrally close. Because no English vowel that is commonly adapted as *aa* is included in the study, all of the responses that perceptually assimilated /æ/ to *a* and *aa* are counted as instances in which /æ/ is equated to a Japanese low vowel. Figure 83 shows the percentages of correct identifications of 16 /æ/ tokens and the percentages of incidents where each /æ/ token is classified as Japanese *a* or *aa*. The percentages of correct identification of /æ/ and the percentages of /æ/ being equated to Japanese a or aa were submitted to Spearman's rank correlation coefficient, and it yielded ρ=0.796 (p<.01).

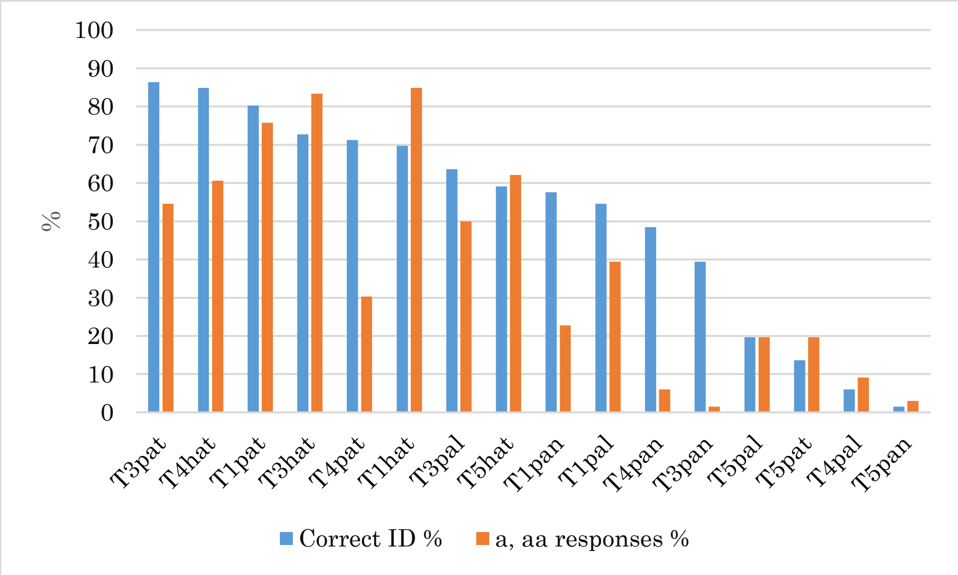

Figure 91. Percentages of correct identification of /æ/ tokens (blue bars) and percentages /æ/ tokens are equated with *a* or *aa* (red bars)

7.1.5 Identification of /ɑ/

The vowel /ɑ/ is commonly adapted as *o*, but it is phonetically closer to *a* than to *o* as acoustic analyses of American English and Japanese vowels show. Studies show that /ɑ/ is perceptually closer to Japanese *a* than to *o* (Strange et al. 1998, 2011, Frieda & Nozawa 2007, Nozawa & Wayland 2012, Nozawa 2018), but studies also imply that Japanese speakers expect /ɑ/ to sound like Japanese /o/ (Nozawa 2016). In this study, at least in /hVt/ and /pVt/ contexts, /ɑ/ is equated with *a* more frequently than with *o*. In /pVl/ context, /ʌ/ is more frequently identified than /ɑ/, and /ʌ/ is also equated with *o* or *oo*. To see whether /ɑ/ is identified more correctly when it is equated with Japanese *o* or *oo*, the percentages of correct identifications of 16 /ɑ/ tokens and the percentages of /ɑ/ being equated with Japanese *o* or *oo* were submitted to Spearman's rank correlation coefficient, and it yielded ρ=0.558 (p<.05).

Figure 92 shows the percentages of correct identifications of /ɑ/ tokens and the percentages of /ɑ/ tokens being equated with Japanese *o* or *oo*. With the exception of Talker 5's "pot" and Talker 3's "pol," /ɑ/ tokens uttered in prenasal contexts are all relatively better identified. Nasality may have affected the perception of vowel height, and in a prenasal context, /ɑ/ was perceived higher, and it equated with Japanese *o* more frequently than in the other contexts. Actually, acoustic analysis revealed that measured F1 frequencies of /ɑ/ are higher in /pVt/ context (see Tables 5-7, Figures 11-13).

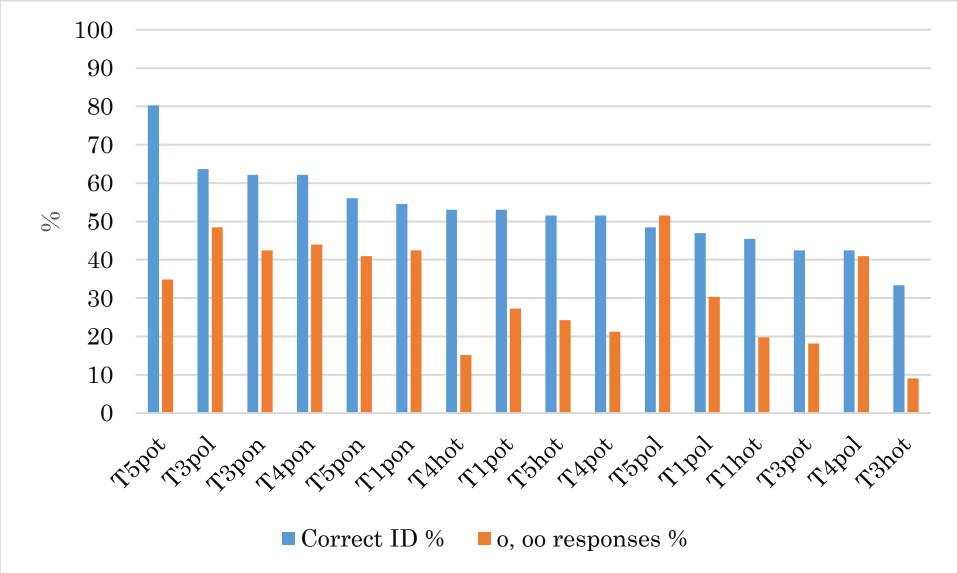

Figure 92. Percentages of correct identification of /ɑ/ tokens (blue bars) and percentages /ɑ/ tokens are equated with *o* or *oo* (red bars)

Although there is no perfect match, the results shown here imply that English vowels are more likely to be correctly identified when they sound like their adaptations to Japanese phonology. Further, the results suggest and Japanese speakers in general expect English vowels to sound like their Japanese adaptations.

7.1.6 Identification of /ʌ/

The vowel /ʌ/ is commonly adapted as *a*, and it is usually perceived as an exemplar of *a* (Strange et al. 1998, 2001; Frieda and Nozawa 2007; Nozawa and Wayland 2012; Nozawa 2018). Nevertheless, according to the survey that I conducted and that I discuss above and to Kori (2018), native Japanese speakers do not seem to be certain what /ʌ/ sounds like. Spearman's rank correlation coefficient was performed between the percentages of correct identifications and the percentages of occurrences in which each /ʌ/ token was perceived as an exemplar of Japanese *a*, but the result shows no significant correlation (ρ=0.178 (p=.51) n.s.). This means that being perceived as an exemplar of *a* has little to do with /ʌ/ being correctly identified. The result of the identification experiment shows that /ʌ/ is most correctly identified in /pVt/ context, and in this context, /ʌ/ is more frequently equated with Japanese *u*. Therefore, Spearman's rank correlation coefficient was performed between the percentages of /ʌ/ tokens being correctly identified, and the percentages of occurrences of /ʌ/ being perceived as an exemplar of Japanese *u*. The result revealed that there is a significant correlation (ρ=0.630 (p<.01)). This is because, first, /ʌ/ is typically spelled as *u* as in *hut*, *cut*, *pun*, and also because English vowels that are commonly adapted as Japanese *u* or *uu* (i.e. /ʊ/ and /u/) are not included in this experiment. Figure 93 shows the percentages of correct identification of /ʌ/ tokens and the percentages of /ʌ/ tokens being equated with Japanese *u*.

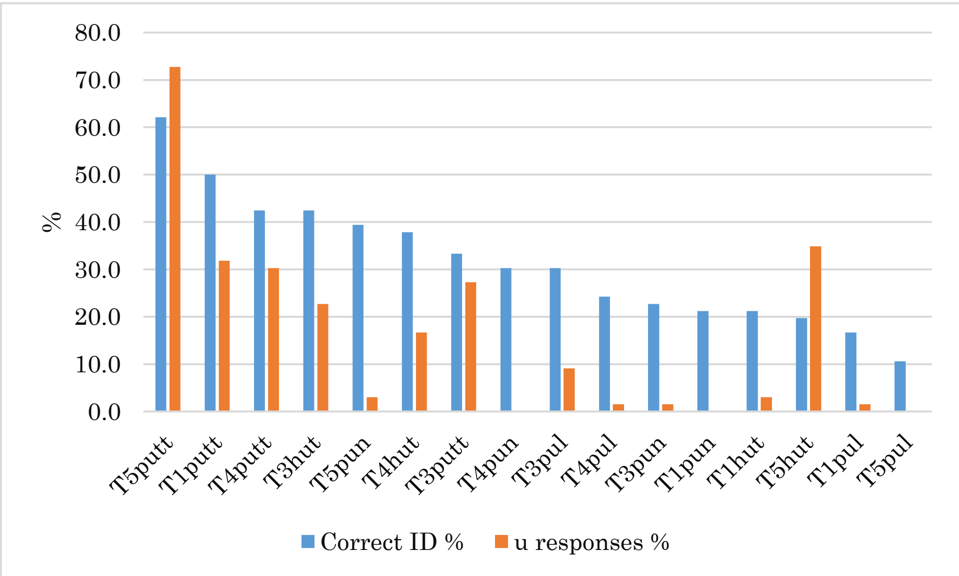

Figure 93. Percentages of correct identification of /ʌ/ tokens (blue bars) and percentages /ʌ/ tokens are equated with *u* (red bars)

7.2 Japanese vowel categories and discrimination of American English vowels

According to Best's Perceptual Assimilation Model (=PAM) (Best 1995), non-native phones that are classified as exemplars of two different native categories are easy to discriminate, but discrimination accuracy is low if non-native phones are classified as exemplars of one native category. If we compare the results of the Perceptual Assimilation Experiment and the Discrimination Experiment, /i/-/ɪ/ is discriminated better in contexts where /i/ is equated with two-mora /iː/, and the discrimination accuracy of /æ/-/ɑ/ is higher in the context when /æ/ is classified as /e/ rather than as /a/. Thus, it seems that perceptual assimilation can predict discrimination accuracy. To verify this, classification overlap scores (Flege & McKay 2004) are calculated. For instance, /i/ in /hVt/ context (heat) is classified as *hii* in 49.6% of instances, as *hi* for 46.6% of instances, as *hie* in 1.9% of instances, and as *hyu* in 0.8% of instances, and /ɪ/ in the same context (hit) is classified as hii in 13.6% of instances, hi in 75.0% of instances, hie in 4.2% of instances, and hyu in 2.3% of instances. /ɪ/ is also classified as he, hei, hia, and bu. Since /i/ is not classified as any of these, they are not included in the calculations. Both /i/ and /ɪ/ are classified as *hii* in 13.6% of instances, as *hi* in 46.6% of instances, as *hie* in 1.9% of instances, and as *hyu* in 0.8% of instances, and the sum of these form the classification overlap score of vowel pair /i/-/ɪ/ in /hVt/ context (Table 17). In the same manner classification overlap scores of all of the vowel pairs in all of the contexts are calculated.

Table 17.

How to calculate a classification overlap score

	heat	hit	overlap
hii	49.6	**13.6**	13.6
hi	**46.6**	75.0	46.6
hie	**1.9**	4.2	1.9
hyu	**0.8**	2.3	0.8
he		2.7	
hei		1.9	
hia		0.4	
		total	**62.9**

Figure 94 shows classification overlap scores of the six vowel pairs in each context in percent. The pairs /i/-/ɪ/ and /ɑ/-/ʌ/ are high in classification overlap scores. The classification overlap scores of /ɛ/-/ɪ/ and /æ/-/ɛ/ are higher in /pVn/ and /pVl/ contexts than two preplosive contexts. On the contrary, the pairs /æ/-/ɑ/ and /æ/-/ʌ/ have higher scores in preplosive contexts than in /pVn/ and /pVl/ contexts. Spearman's rank correlation coefficient was performed in order to determine whether there is a correlation between discrimination accuracy and classification overlap scores, which yielded a negative correlation ρ=-0.632 (p<.01).

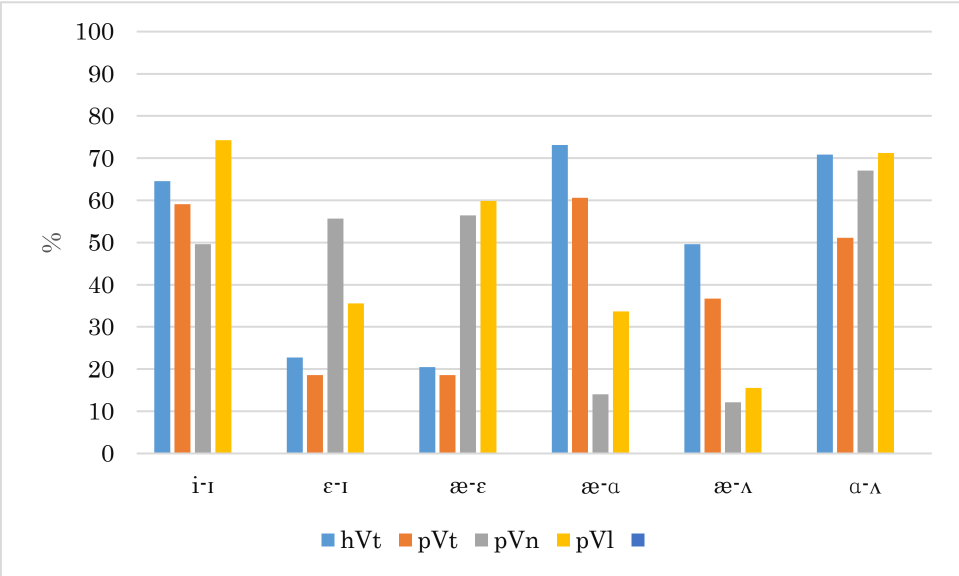

Figure 94. Classification overlap scores

8. Individual differences among Japanese participants

8.1 Phonetic cues to differentiate /i/ and /ɪ/

The results shown so far indicate that /i/ is identified better when it is perceived as a long /iː/. The result of the production experiment also shows that Japanese speakers try to differentiate /i/ and /ɪ/ by duration. However, at least some of the participants seem to be aware of the spectral difference of the two vowels. They differentiate the two vowels by quality as well as by quantity when they read each vowel aloud. Then the question is whether their awareness of qualitative difference between the two vowels facilitates better identification and discrimination of these vowels.

Whether they are aware of the qualitative (=spectral) difference can be assumed by measuring F1 and F2 frequencies of these two vowels produced by the Japanese participants in the “Read Aloud” condition. If they are aware of the difference, then F1 of /i/ should be lower than that of /ɪ/, and F2 of /i/ should be higher than that of /ɪ/. Because the variation range of F2 is wider than that of F1, and because female speech has a wider variation range of formant frequencies, the difference of F2 frequencies can make the difference seem larger than it is. Moreover, female participants may make better use of differences of formant frequencies. Thus the formant frequencies are converted to the Mel scale, which is "perceptual scale of pitches judged by listeners to be equal in distance from one another."[8]

The 33 participants are divided into two groups: those whose /ɪ/ is lower and more centralized than their /i/ in /hVt/ context (Group 1); and the rest of the participants (Group 2). Eighteen participants are classified into Group 1 and the other 15 participants are in Group 2. The mean F1 and F2 frequencies of /i/ and /ɪ/ in Mel scale of all the participants, Group 1 and Group 2, are summarized in Table 18.

Table 18.

Mean F1 and F2 frequencies of /i/ and /ɪ/ uttered by Japanese participants in /hVt/ context in "Read Aloud" condition (in Mel scale), and mean Euclidean distance between the two vowels (E.D.)

	all the participants		Group 1		Group 2	
	F1	F2	F1	F2	F1	F2
/i/	541.1	1836.5	540.5	1865.8	541.8	1801.3
/ɪ/	564.9	1798.8	583.2	1803.3	543.0	1793.3
E.D		44.6		75.6		8.1

If those in Group 1 are aware that /i/ and /ɪ/ are intrinsically different in quality as well as in quantity, and they are more sensitive to qualitative difference than are those in Group 2, then they should perform better in perception experiments.

[8] The following formula is used to convert hertz f into mels m.

$$m = 2595\log_{10}(1 + \frac{f}{700})$$

The mean discrimination accuracy of /i/-/ɪ/ is 76.4% for Group 1, and that of Group 2 is 74.4%. Although Group 1's discrimination accuracy is slightly higher, the mean identification accuracy of /i/ and /ɪ/ of Group 1 is 56.3% and 86.8%, and that of Group 2 is 57.5% and 89.2%. If those in Group 1 know that /i/ and /ɪ/ are spectrally different, they may not be able to use the knowledge to differentiate the two vowels when they hear them. On the other hand, the perception and production may simply be different, or the qualitative difference could serve as a secondary cue, and Japanese listeners may listen for the durational difference as the primary cue to differentiate these two vowels.

In the same manner as in /hVt/ context, all of the participants were divided into two groups based on whether the /ɪ/ they utter is lower and more centralized than /i/ they utter in the /pVt/ context. In this context, 21 participants were classified as Group1, and 12 others were classified as Group 2. Table 19 shows mean F1 and F2 frequencies in Mel scale. Participants in Group 1 correctly identified /i/ and /ɪ/ in 51.8% and 82.7% of instances, while those in Group 2 identified these vowels in 39.6% and 78.1% of instances. As for discrimination, participants in Group 1 discriminated the vowel pair correctly in 74.2% of instances, while those in Group 2 discriminated the vowel pair in 66.0% of instances. Thus, those who differentiate these two vowels qualitatively appear to be more sensitive to acoustic differences between the two vowels.

Table 19.

Mean F1 and F2 frequencies of /i/ and /ɪ/ uttered by Japanese participants in /pVt/ context in "Read Aloud" condition (in Mel scale) and mean Euclidean distance between the two vowels (E.D.)

	all the participants		Group 1		Group 2	
	F1	F2	F1	F2	F1	F2
/i/	534.1	1833.9	526.5	1839.9	547.5	1823.5
/ɪ/	557.2	1796.9	577.8	1782.2	521.0	1822.5
E.D		43.6		72.6		26.6

In /pVn/ context, 16 participants are classified into Group 1, and 17 are classified into

Group 2. Their mean F1 and F2 in the Mel scale are shown in Table 20. The identification accuracy of /i/ and /ɪ/ by participants in Group 1 is 21.9% and 47.7%, respectively, while those in Group 2 correctly identified /i/ and /ɪ/ in 21.3% and 45.6% of instances, respectively. As for discrimination accuracy, participants in Group 1 discriminated the vowel pair in 68.8% of instances while those in Group 2 discriminated the vowel pair in 66.2% of instances. In /pVn/ context, the difference between the two groups is much smaller than in /pVt/ context.

Table 20.

Mean F1 and F2 frequencies of /i/ and /ɪ/ uttered by Japanese participants in /pVn/ context in "Read Aloud" condition (in Mel scale) and mean Euclidean distance between the two vowels (E.D.)

	all the participants		Group 1		Group 2	
	F1	F2	F1	F2	F1	F2
/i/	585.0	1872.8	573.3	1873.9	596.1	1871.7
/ɪ/	603.1	1849.8	634.8	1822.5	573.2	1875.5
E.D.		29.2		80.2		23.2

Table 21.

Mean F1 and F2 frequencies of /pil/ and /pɪl/ at the 25% of /pVl/ continua uttered by Japanese participants in /pVl/ context in "Read Aloud" condition (in Mel scale) and mean Euclidean distance between the two vowels (E.D.)

	all the participants		Group 1		Group 2	
	F1	F2	F1	F2	F1	F2
/i/	577.2	1771.4	558.8	1800.5	602.0	1732.0
/ɪ/	596.9	1708.7	597.8	1714.2	595.8	1701.3
E.D.		65.8		94.7		31.3

In /pVl/ context, 19 and 14 participants are classified into Groups 1 and 2, respectively. The mean F1 and F2 frequencies in Mel scale are shown in Table 21. The identification accuracy of /i/ and /ɪ/ by the participants in Group 1 is 34.9% and 77.0%, respectively, and the identification accuracy by those in Group 2 is 37.5% and 59.8%, respectively. Regarding the discrimination accuracy of the /i/-/ɪ/ vowel pair, the participants in Group 1 discriminated the vowel pair correctly in 61.8% of instances, while those in Group 2 discriminated the pair in 56.0% of instances.

The number of participants who produced /ɪ/ with higher F1 and lower F2 than /i/ in fact differs from one context to another, and, if they do not consistently produce these two vowels in this way, it is doubtful whether participants are really aware of the qualitative difference between the two vowels. Only seven of the participants (J2, J3, J9, J13, J18, J25, J30) differentiated these two vowels qualitatively in all of the four consonantal contexts. These seven participants' mean F1 and F2 frequencies of /i/ and /ɪ/ in the four different consonantal contexts and mean Euclidean distance between the two vowels are shown in Table 22. These participants produced /i/ and /ɪ/ with larger qualitative differences than those who did not make qualitative differences in all the four consonantal contexts.

Table 22.

Mean F1 and F2 frequencies of /i/ and /ɪ/ uttered by 7 participants who consistently differentiate /i/ and /ɪ/qualitatively in "Read Aloud" condition (in Mel scale) and mean Euclidean distance between the two vowels (E.D.)

	hVt		pVt		pVn		pVl	
	F1	F2	F1	F2	F1	F2	F1	F2
/i/	518.4	1840.5	496.6	1841.6	552.3	1879.9	551.5	1790.6
/ɪ/	567.5	1775.2	558.7	1767.6	615.3	1806.7	594.0	1685.6
E.D.		81.7		96.6		96.6		113.2

However, the results of the identification and discrimination experiments reveal that these participants were not more successful in identifying and discriminating /i/ and /ɪ/ than the rest of the participants. In Table 23, the mean identification accuracy of /i/ and /ɪ/ by the seven participants is compared to that of the rest of the participants. In /hVt/

and /pVl/ contexts, the seven participants' identification accuracy of /i/ and /ɪ/ is higher than that of the rest of participants, but in the /pVt/ context, their identification accuracy of /ɪ/ is lower than that of the rest of the participants, and, in the /pVn/ context, the seven participants identified /i/ slightly less accurately than the rest of the participants.

Table 23 shows the mean discrimination accuracy of the /i/-/ɪ/ vowel pair by the seven participants and by the rest of the participants. Like identification accuracy, the seven participants' discrimination accuracy is higher than that of the rest of the participants in the /hVt/ and /pvl/ contexts, but, in the /pVn/ context, the opposite is true. Moreover, in /pVt/ context the seven participants discriminated the vowel pair better than the rest of the participants, but the difference is just 1.1%. This seems unexpected, considering that the seven participants differentiated /i/ and /ɪ/ with larger Euclidean distance in /pVt/ and /pVn/ contexts than in /hVt/ context.

Table 23.

Mean identification accuracy in percent of /i/ and /ɪ/ by the seven participants and the rest of the participants in the four contexts

	hVt		pVt		pVn		pVl	
	/i/	/ɪ/	/i/	/ɪ/	/i/	/ɪ/	/i/	/ɪ/
seven participants	58.9	92.9	53.6	80.4	21.4	51.8	39.3	78.6
the rest of participants	56.3	86.5	45.7	81.3	21.6	45.2	35.1	67.3

Table 24.

Mean discrimination accuracy in percent of /i/-/ɪ/ vowel pair by the seven participants and the rest of the participants in the four contexts

	hVt	pVt	pVn	pVl
seven participants	83.3	71.9	60.7	64.3
the rest of participants	73.4	70.8	69.2	58.0

Each of the seven participants' responses and errors was examined. In the identification task, one of the seven participants, J30, correctly identified /ɪ/ in /pVt/

context just in 50% of all the eight trials. This participant's low identification accuracy lowers the mean identification accuracy of the seven participants (Group 1). This participant misidentified /ɪ/ for /i/. Euclidean distance between this participant's /i/ and /ɪ/ is 99.8, so her /i/ and /ɪ/ are not particularly close. In /pVn/ context, /i/ is least accurately identified among all the four contexts (See Table 14 in Chapter 4), so mean identification accuracy is lower than in the other three contexts, but four participants, J2, J3, J9 and J30, lower the mean identification accuracy of /i/ by the seven participants in /pVn/ context. The Euclidean distance between /i/ and /ɪ/ uttered by these four participants is respectively 95.7, 61.9, 48.9 and 44.3, so the distance between the two vowels is relatively smaller. These four participants identified /i/ correctly in 0%, 25%, 12.5%, and 0% of eight trials. J2, J3 and J9 predominantly identified /i/ for /ɪ/. J30, on the other hand, identified /i/ for lower vowels /ɛ/, /æ/, /ɑ/ and /ʌ/ as well as for /ɪ/. This participant identified /ɪ/ for /ɛ/, /æ/ and /ɑ/ as well.

As for discrimination, four participants, J2, J3, J13, and J25, discriminated the vowel pair less accurately than the average. These four participants' Euclidean distance between the two vowels is respectively 95.7, 61.9, 223.2 and 73.2, so except J13 /i/ and /ɪ/ these participants produced are closer than the group's average.

From all of these together, it can be assumed that, in general, those who differentiate /i/ and /ɪ/ qualitatively in production are likely to be more sensitive to qualitative difference between the two vowels in perception.

In the Production Experiment in Chapter 6, it was found that /i/ and /ɪ/ are more distant from each other when Japanese participants repeated after native speakers, but that the /i/ and /ɪ/ that the Japanese participants produced are still closer to each other than those that native speakers produced. Here I would like to examine how the sensitivity to discern spectral or qualitative differences between /i/ and /ɪ/ facilitates better identification and discrimination of these vowels.

Tables 23-26 show mean F1 and F2 frequencies of /i/ and /ɪ/ in Mel scale as uttered by Japanese participants in the "Read Aloud" condition and the "Repetition" condition. Mean F1 and F2 frequencies were compared for the two vowels uttered in the two conditions by measuring Euclidean distance (E.D. in Tables). T1, T3, T4, and T5 in

Tables each denote Talker 1, Talker 3, Talker 4, and Talker 5, after whom Japanese participants repeated. In general, /ɪ/, uttered in the Repetition condition, is more distant from /ɪ/ uttered in the Read Aloud condition than is the /i/ uttered in the Repetition condition from the /i/ in the Read Aloud condition. In other words, spectral difference between /i/ and /ɪ/ is widened mostly by moving /ɪ/ backward and downward. Euclidean distance between the two condition is generally larger for /ɪ/than for /i/. Equally noticeable is that Euclidean distance between vowels uttered in two different conditions is larger in /pVn/ and /pVl/ contexts than in /hVt/ or /pVt/ contexts. Japanese participants "modified" their production more in /pVn/ and /pVl/ contexts to make their production sound closer to that of native speakers.

As for the significance of differences among Talkers, the vowel sound /i/ following Talker 1 is generally closest to /i/ in the "Read Aloud" condition, and /i/ after Talkers 4 and 5 are rather distant from /i/ in the Read Aloud condition. In the Production Experiment, /i/ and /ɪ/ following Talker 4 are close to each other (Figures 50 and 51), and this can be confirmed in Table 25.

Table 25.

Mean F1 and F2 frequencies in Mel scale of /i/ and /ɪ/ uttered in /hVt/ context by Japanese participants in "Read Aloud" condition and "Repetition" condition, along with Euclidean distance between vowels uttered in the two conditions

	read aloud		after T1		after T3		after T4		after T5	
/i/	541.1	1836.5	537.3	1836.5	524.7	1822.2	558.8	1775.0	553.6	1798.3
E.D.				3.7		21.7		64.0		40.2
/ɪ/	564.9	1798.8	604.3	1743.0	603.8	1736.0	587.5	1744.4	611.1	1729.5
E.D.				68.3		73.9		58.9		83.3

Table 26.

Mean F1 and F2 frequencies in Mel scale of /i/ and /ɪ/ uttered in /pVt/ context by Japanese participants in "Read Aloud" condition and "Repetition" condition along with Euclidean distance between vowels uttered in the two conditions

	read aloud		after T1		after T3		after T4		after T5	
/i/	534.1	1833.9	535.8	1832.7	511.8	1841.4	524.5	1802.7	555.2	1800.2
E.D.				2.1		23.6		32.7		39.8
/ɪ/	557.2	1796.9	582.1	1725.9	577.9	1746.6	563.8	1750.9	584.2	1738.8
E.D.				75.2		54.4		46.5		64.0

Table 27.

Mean F1 and F2 frequencies in Mel scale of /i/ and /ɪ/ uttered in /pVn/ context by Japanese participants in "Read Aloud" condition, and "Repetition" condition along with Euclidean distance between vowels uttered in the two conditions

	read aloud		after T1		after T3		after T4		after T5	
/i/	585.0	1872.8	582.7	1878.8	613.4	1838.3	602.0	1828.1	610.7	1820.3
E.D.				6.5		44.6		47.8		58.4
/ɪ/	603.1	1849.8	655.1	1772.7	697.2	1695.4	650.4	1743.2	702.7	1659.8
E.D.				93.0		180.8		116.6		214.5

Table 28.

Mean F1 and F2 frequencies in Mel scale of /i/ and /ɪ/, uttered at 25% of /pVl/ continua by Japanese participants in "Read Aloud" condition and "Repetition" condition along with Euclidean distance between vowels uttered in the two conditions

	read aloud		after T1		after T3		after T4		after T5	
/i/	577.2	1771.4	578.8	1758.4	572.1	1743.9	584.6	1736.8	630.4	1650.8
E.D.				13.1		28.0		35.4		131.8
/ɪ/	596.9	1708.7	617.8	1661.8	681.8	1602.4	659.8	1627.3	667.2	1609.2
E.D.				51.4		136.0		102.8		121.8

In each consonantal context, it can be seen that spectral differences between /i/ and /ɪ/ is discerned, but some participants were sensitive to the spectral differences they heard while others were not. As in the Read Aloud condition, participants were divided into two groups: those who differentiated /i/ and /ɪ/ spectrally in the right direction, with F1 frequency of /ɪ/ higher and F2 frequency of /ɪ/ lower than those of /i/ (Group 1); and those who did not (Group 2), regardless of whether they differentiated these two vowels in the Read Aloud condition. Tables 29-32 show the number of participants in each group, along with the percentages of correct identification of /i/ and /ɪ/, the percentages of correct discrimination of /i/-/ɪ/, and the mean Euclidean distance between /i/ and /ɪ/ uttered by participants in Group 1

Table 29.

The number of participants classified into Group 1 and Group 2 based on the utterance they repeated after 4 talkers in /hVt/ context.

	Group 1	Group 2		Group 1	Group 2
After T1	N=25	N=8	After T3	N=22	N=11
/i/	92.0	75.0	/i/	97.7	86.4
/ɪ/	80.0	75.0	/ɪ/	84.1	95.5
/i/-/ɪ/	78.0	67.7	/i/-/ɪ/	78.4	69.7
E.D.	154.4		E.D.	187.1.	
After T4	N=11	N=22	After T5	N=21	N=12
/i/	13.6	9.1	/i/	38.1	29.1
/ɪ/	90.1	97.2	/ɪ/	88.1	91.7
/i/-/ɪ/	78.0	67.7	/i/-/ɪ/	77.8	71.5
E.D.	170.7		E.D.	163.7	

Note. /i/ and /ɪ/ denote the percentages of correct identification of these vowels, /i/-/ɪ/ denote the percentages of correct discrimination of the vowel pair, and E.D. denotes the mean Euclidean distance between /i/ and /ɪ/ uttered by participants in Group 1

In /hVt/ context, the participants in Group 1 outnumber those in Group 2, except when they repeated after Talker 4. This explains why the /i/ and /ɪ/ vowel sounds that participants uttered after Talker 4 are spectrally close. Generally, those in Group 1 outperform those in Group 2 in identifying /i/ and discriminating /i/-/ɪ/, but this is not the case with the identification of /ɪ/. Those in Group 2 performed better than those in Group 1. Participants were not told what vowel they would hear when they worked on the Production Experiment. They tried to reproduce a vowel as close as possible to the vowel they heard. Some of the participants discerned spectral differences between /i/ and /ɪ/, and they produced /ɪ/ with higher F1 and lower F2 than /i/, but it may be the case that participants in Group 1 were not always certain that the vowel they heard was /ɪ/. Some Group 1 members misidentified /ɪ/ for /ɛ/, and this lowers the identification accuracy of /ɪ/, even though they discerned spectral differences between /i/ and /ɪ/. Some of the participants in Group 2 also misidentified /ɪ/ for /ɛ/, but less often than those in Group 1. Participants in Group 1's identification accuracy of the /i/ uttered specifically by Talker 4 and Talker 5 is lower than i/ that was uttered by Talker 1 and Talker 3, even though they discerned spectral difference /i/ and /ɪ/. This can be attributed to durational difference. For Talker 1 and Talker 3, /i/ is longer, and that vowel sound is more likely to be identified correctly. Spectral information is not enough for Japanese listeners to correctly identify /i/. It has to be noticeably longer than /ɪ/. This supports the claim that the image of English vowels that Japanese speakers hold strongly affects Japanese speakers' perceptions of English vowels. Thus, /i/ has to sound like two-mora *ii*.

Table 30: *The number of participants classified into Group 1 and Group 2 based on the utterance they repeated after 4 talkers in /pVt/ context.*

	Group 1	Group 2		Group 1	Group 2
After T1	N=19	N=14	After T3	N=24	N=9
/i/	68.4	64.3	/i/	58.3	55.6
/ɪ/	81.6	78.5	/ɪ/	68.8	77.8
/i/-/ɪ/	75.4	66.6	/i/-/ɪ/	74.3	68.0
E.D.	204.2		E.D.	149.6.	
After T4	N=19	N=14	After T5	N=16	N=17
/i/	68.4	57.1	/i/	3.3	0
/ɪ/	100	97.2	/ɪ/	86.7	66.7
/i/-/ɪ/	72.4	67.7	/i/-/ɪ/	73.4	69.7
E.D.	170.7		E.D.	177.3	

Note. /i/ and /ɪ/ denote the percentages of correct identification of these vowels, /i/-/ɪ/ denote the percentages of correct discrimination of the vowel pair, and E.D. denotes the mean Euclidean distance between /i/ and /ɪ/ uttered by participants in Group 1

Figure 22 shows that the most successful participant correctly identified /i/ in 87.5% of instances, while the least successful participant correctly identified /i/ just in 12.5% of instances. Similarly, Figure 27 shows that, while at least one participant discriminated /i/-/ɪ/ perfectly, the least successful participant actually discriminated the vowel pair correctly in 33.3% of instances. Note, for example, J8 and J11, whose identification accuracy of /i/ is 12.5% and 25%, respectively. These two participants belong to in Group 2 except when they repeated the utterances of Talker 5. Then, J12, whose discrimination accuracy is the lowest, is in Group 2. Those who did not perform well in identification and discrimination tasks are less sensitive to spectral differences between /i/ and /ɪ/.

Similar tendencies were observed for /pVt/ context. Talker 3's /ɪ/ was more frequently misidentified as /ɛ/ by participants in Group 1. Talker 5's /i/ is a bit lower and backer in the vowel space (see Table 6) than are /i/ tokens uttered by the other three Talkers, and it is the shortest of all of the four /i/ tokens. Consequently, Talker 5's /i/ was predominantly misidentified as /ɪ/. If spectral differences are discerned between Talker 5's /i/ and /ɪ/, this is because /ɪ/ is misidentified as /ɛ/.

From Figure 22, it can be seen that some of the participants did not identify /i/ in /pVt/ context at all. J9 and J12's identification accuracy of /i/ was 0%, and J2, J7, J8, and J23 correctly identified /i/ in just 12.5% of instances. Participant J30 correctly identified /i/ in 25% of instances. These participants predominantly misidentify /i/ for /ɪ/. As for discrimination, J7 and J8's discrimination accuracy was 33.3%.

In /pVn/ context, /i/ and /ɪ/ were least accurately identified (see Table 14). No participant's identification accuracy of /i/ was particularly low (Figure 23). In fact, seven participants did not identify /i/ at all (J2, J12, J21, J23, J29, J30, J34), and with the exception of J7, who correctly identified /i/ in all the trials, the participants' identification accuracy was 37.5% or lower. As for /ɪ/, two participants (J20, J34) correctly identified /ɪ/ in none of the trials (0%). One participant correctly identified /ɪ/ in one of the trials (12.5%). Seven other participants (J7, J13, J14, J21, J23, J24, J30) identified /ɪ/ in two of the trials (25.0%). As for discrimination, four participants' discrimination accuracy was lower than 50% (J19: 33.3%; J10, J25, J28: 41.7%).

More importantly, participants in Group 1 identified /i/ and /ɪ/ less accurately than did those in Group 2 regardless of which native speaker they repeated after, but participants in Group 1 discriminated the vowel pair more accurately than did those in Group 2. This is largely because /i/ was frequently misidentified as /ɪ/, and, in turn, /ɪ/ was frequently misidentified as /ɛ/. Participants in Group 1 may be more sensitive to spectral differences, and this is why /i/ and /ɪ/ were mistakenly identified as lower vowels. Participants in Group 2, who are presumably less sensitive to spectral differences between /i/ and /ɪ/, may have chosen /ɪ/ primarily based on perceived vowel duration.

Table 31.

The number of participants classified into Group 1 and Group 2 based on the utterance they repeated after 4 talkers in /pVn/ context.

	Group 1	Group 2		Group 1	Group 2
After T1	N=22	N=11	After T3	N=25	N=8
/i/	47.7	68.2	/i/	18.0	43.8
/ɪ/	15.9	22.7	/ɪ/	62.0	75.0
/i/-/ɪ/	74.4	53.8	/i/-/ɪ/	69.7	60.4
E.D.	210.5		E.D.	243.7	
After T4	N=19	N=14	After T5	N=22	N=11
/i/	2.8	6.7	/i/	0	9.1
/ɪ/	61.1	70.0	/ɪ/	36.4	40.9
/i/-/ɪ/	72.4	62.2	/i/-/ɪ/	69.3	63.6
E.D.	177.8		E.D.	285.3	

Note. /i/ and /ɪ/ denote the percentages of correct identification of these vowels, /i/-/ɪ/ denote the percentages of correct discrimination of the vowel pair, and E.D. denotes the mean Euclidean distance between /i/ and /ɪ/ uttered by participants in Group 1.

As argued earlier, English vowels are most likely to be correctly identified when the vowels sound like the auditory image that Japanese speakers hold. In /pVl/ context, because F2 continues to descend, /i/ cannot keep its steady-state high front position, which is likely to cause the vowel sound to be distant from Japanese speakers' image. This led to the low identification accuracy of /i/ in this context (Table 15). F1 and F2 frequency transition of /i/ can be seen in Table 8 and in Figures 14-16. /ɪ/ is more largely retracted (Figures 73-77), so the difference in formant frequencies between /i/ and /ɪ/ is larger in /pVl/ context. It is, therefore, surprising that some of the participants did not seem to discern the spectral differences between the two vowels even though more than half of the participants differentiated the two vowels in the right direction in production. One thing that has to be pointed out here is that, unlike in the other three consonantal contexts, the boundary between the vowel and the following /l/ is unclear, so F1 and F2 frequencies were measured at 25%, 50%, and 75% of /Vl/ continua. F1 and F2 frequencies at 25% of the continua might not be a fair representation of vowel quality. Further research is necessary to investigate this. Without such research, we cannot predict identification and discrimination accuracy of /i/ and /ɪ/ based on the mean F1 and F2 frequencies of /i/ and /ɪ/ that the Japanese participants produced for this study.

Table 32.

The number of participants classified into Group 1 and Group 2 based on the utterance they repeated after 4 talkers in /pVl/ context.

	Group 1	Group 2		Group 1	Group 2
After T1	N=18	N=15	After T3	N=22	N=11
/i/	88.9	50.0	/i/	43.2	22.7
/ɪ/	58.3	73.3	/ɪ/	68.2	34.1
/i/-/ɪ/	58.3	60..6	/i/-/ɪ/	59.1	59.8
E.D.	227.4		E.D.	261.0	
After T4	N=21	N=12	After T5	N=16	N=17
/i/	19.0	33.3	/i/	2.9	6.3
/ɪ/	81.0	87.5	/ɪ/	58.8	65.6
/i/-/ɪ/	61.1	56.3	/i/-/ɪ/	55.4	63.5
E.D.	210.8		E.D.	109.5	

Note. /i/ and /ɪ/ denote the percentages of correct identification of these vowels, /i/-/ɪ/ denote the percentages of correct discrimination of the vowel pair, and E.D. denotes the mean Euclidean distance between /i/ and /ɪ/ at the 25% of /pVl/ continua uttered by participants in Group 1

8.2 Perception and production of two vowels adapted as Japanese *a*

As pointed out in the Introduction, both /æ/ and /ʌ/ are commonly adapted as *a* in Japanese. Thus, both *track* and *truck* are *torakku* in Japanese. Japanese speakers seem to believe that /æ/ is the closet English vowel to Japanese *a*, most likely because /æ/ is usually represented by the letter a (Nozawa, 2018). The results of the Production

Experiment in the "Read Aloud" condition agree with this view. However, some of the participants produced /æ/ as a front vowel, and they seemed to be aware of phonetic differences between English /æ/ and Japanese *a*.

Here I would like to examine whether Japanese participants who produce English /æ/ as a front vowel are more sensitive to phonetic differences between /æ/ and /ʌ/. The sensitivity will be measured by identification and discrimination accuracy.

It is impractical to divide all of the participants into two groups: those who produce /æ/ as a front vowel and those who do not. Therefore, I decided to select participants who are likely to produce /æ/ as a front vowel constantly, and then choose those who are likely to produce /æ/ as non-front vowel constantly. Those who are likely to produce /æ/ as a front vowel are defined as those whose frequencies of /æ/ in the contexts of /hVt and /pVt/ in "Read Aloud" condition are higher than 1600Hz (for male participants, higher than 1500Hz), and those who are likely to produce /æ/ as a non-front vowel are defined as those whose F2 frequencies of /æ/ in /hVt/ and /pVt/ contexts in "Read Aloud" condition are lower than 1400Hz (for male participants, lower than 1300Hz).

Nine participants are included in the Front-vowel Group (J1, J4, J5, J8, J18, J19, J25, J26, and J32), and all members of the group are female. Eight participants are included in Non-front-vowel Group (J2, J7, J10, J12, J14, J16, J21, J22), two of whom (J2, J22) are male. The results are summarized in Table 33. Large differences between the two groups can be seen in the identification accuracy of /ʌ/, and the discrimination accuracy of /æ/-/ʌ/, rather than the identification accuracy of /æ/. This is probably because when those in the Front-vowel group hear a Japanese *a*-like vowel, they can rule out the possibility that it is /æ/. The difference between the two groups is smaller in the /pVt/ context, most likely because in this /pVt/ context, /ʌ/ is most frequently equated with the Japanese *u*, and so phonetic differences between /æ/ and /ʌ/ may be easier to detect.

Of the eight participants in the Non-front vowel Group, four participants "modified" their pronunciation when they repeated after native speakers, and then produced /æ/ as a front vowel. The other four (J2, J7, J10, J14), however, continued to produce /æ/ as a non-front vowel even when they repeated after native speakers.

Table 33.

Mean identification accuracy of /æ/ and /ʌ/ and mean discrimination accuracy of /æ/-/ʌ/ in /hVt/ and /pVt/ context (in %) by Front-vowel Group and Non-front vowel Group

	/hVt/			/pVt/		
	/æ/	/ʌ/	/æ/-/ʌ/	/æ/	/ʌ/	/æ/-/ʌ/
Front vowel Group	77.8	51.4	84.3	66.7	59.7	69.4
Non-front V Group	70.3	28.1	78.1	67.2	53.1	50.0

In /pVn/ context, 15 participants (J1, J4, J5, J8, J11, J13, J15, J17, J18, J19, J25, J26, J29, J31, J32) produced /æ/ as a front vowel and 11 participants (J2, J6, J9, J10, J12, J14, J16, J21, J23, J24, J28) produced it as a non-front vowel. Of the Front vowel Group, only J13 is male, and J2, J23 and J24 in Non-front vowel Group are male. The identification accuracy of /æ/ and /ʌ/ by the participants in the two groups is compared. Also compared is the discrimination accuracy of /æ/-/ɛ/ and /æ/-/ʌ/ by the two groups of participants. Because /æ/ is raised and fronted in /pVn/ context, it is closer to /ɛ/ than in /hVt/ or /pVt/ contexts. The results are summarized in Table 34. Front vowel Group outperformed Non-front vowel Group both in identification and discrimination accuracy. Like in /hVt/ and /pVt/ context, the difference is largest in the identification accuracy of /ʌ/. When they heard vowels that sound like Japanese *a*, participants in Front vowel Group can rule out the possibility that the vowel is /æ/, and they are also more sensitive to phonetic differences between /æ/ and /ɛ/ than are participants in Non-front vowel Group.

Table 34.

Mean identification accuracy of /æ/ and /ʌ/ and mean discrimination accuracy of /æ/-/ɛ/and /æ/-/ʌ/ in /pVn/ context (in %) by Front-vowel Group and Non-front vowel Group

	/æ/	/ʌ/	/æ/-/ɛ/	/æ/-/ʌ/
Front vowel Group	44.2	38.3	62.2	86.1
Non-front vowel group	37.5	19.3	50.8	84.8

Almost all the participants produced /æ/ as a front vowel when they repeated the utterances of the native speakers, but a few participants still produced /æ/ as a non-front vowel even when they repeated after native speakers. J10 produced /æ/ as a non-front vowel when she repeated the speech of Talkers 1 and 4. J11 produced /æ/ as a non-front vowel when she repeated after Talkers 1, 3 and 4, even though she produced the vowel as a front vowel when she read aloud words from the word list. J30 also produced /æ/ as a non-front vowel when she repeated after Talker 1. These examples indicate that knowing what the phonetic features of a vowel are like does not always facilitate the correct perception of the phonetic features of the vowel. Furthermore, the examples indicate that there is an individual difference as to the perception of phonetic features of non-native vowels even if learners know how to pronounce the vowel correctly.

8.3 Perceptual assimilation and the pronunciation of /ɛ/

As shown in Figure 90, Talker 1's production of /ɛ/, as in "pet," was the least accurately identified of all the 16 /ɛ/ tokens. This particular /ɛ/ token is backer than the other /ɛ/ tokens, and it is equated to Japanese *u* more frequently than to *e*. Mean F1 and F2 frequencies of /ɛ/ uttered by 33 Japanese participants when they repeated after Talker 1 are 720 and 1372 (in Mel) respectively. Eight participants correctly identified this /ɛ/ token in two trials in the Identification Experiment, and mean F1 and F2 frequencies of /ɛ/ uttered by these eight participants when they repeated after Talker 1 are 734 and 1421 (in Mel). Nine participants, on the other hand, misidentified this /ɛ/ token as backer vowels /ɑ/ or as /ʌ/ in two trials. Mean F1 and F2 frequencies of /ɛ/ uttered by these participants are 715 and 1325 (in Mel), respectively.

In the Perceptual Assimilation Experiment, just three participants equated this /ɛ/ token to Japanese *e* in two trials. Mean F1 and F2 frequencies of /ɛ/ uttered by these three participants, when they repeated after Talker 1, are 685 and 1527 (in Mel). Sixteen participants equated this /ɛ/ token to Japanese vowels other than *e*, and mean F1 and F2 frequencies of /ɛ/ uttered by these participants are 727 and 1314 (in Mel).

These results imply that there is the possibility of individual differences of an L1 vowel category boundary although the specific causes of these differences is beyond the scope of this study.

8.4 Perceptual assimilation and the pronunciation of /ɑ/

The Low back vowel /ɑ/ is closer to the Japanese *a* than to *o*, but it is commonly adopted as *o* in Japanese. This discrepancy makes the identification of this vowel by native Japanese speakers difficult. An acoustic analysis revealed that Japanese participants who perceptually assimilated /ɑ/ to Japanese *a* or *aa* in fact produced /ɑ/ as a lower vowel, while those who equated /ɑ/ with Japanese *o*, *oo*, or *ou* produced it as a back but higher vowel. Neither group produced /ɑ/ as a low back vowel. Examples are shown in Table 33. Regardless of whether /ɑ/ is equated to Japanese *a* or *o*, the vowel that Japanese participants identified is higher (indicated by lower F1 frequencies) than the native speaker's vowel that they repeated.

Table 35.

Mean F1 and F2 frequencies of /a/ uttered by Japanese participants who equated /a/ to Japanese a and o

	native speakers		those who equated to *a*		those who equated to *o*	
	F1	F2	F1	F2	F1	F2
T3pon	799	1128	796	1138	763	1080
T1pot	1063	1215	815	1133	773	1113
T4hot	954	1244	843	1196	804	1169

8.5 Linguistic experience and the perception of English vowels

As shown earlier in Table 9, the linguistic experience and English proficiency of these Japanese participants are not uniform. To determine how large are individual differences in identification accuracy, the percentages of correct identification of all of the vowels in all the consonantal contexts are submitted to Multidimensional Scaling (ALSCAL). Zero, where X-axis and Y-axis cross, is the average of all participants. The numbers in the X- and Y-axes indicate the standard deviation. J 25 is the most distant from the group (-2.9 in Dimension 1). This subject was born in Malaysia and has spent more time abroad than in Japan. The subject received the TOEIC ® L & R Score of 990. Her linguistic experience has certainly made her more sensitive to English vowel categories.

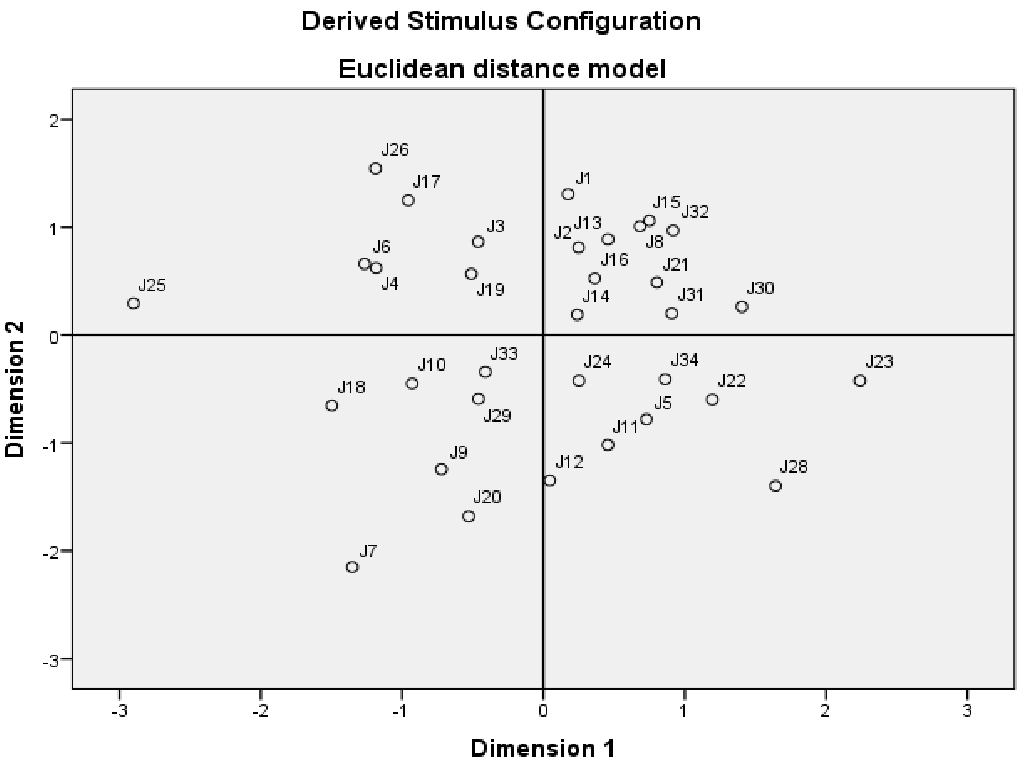

Figure 95. Result of MDS (Identification Data)

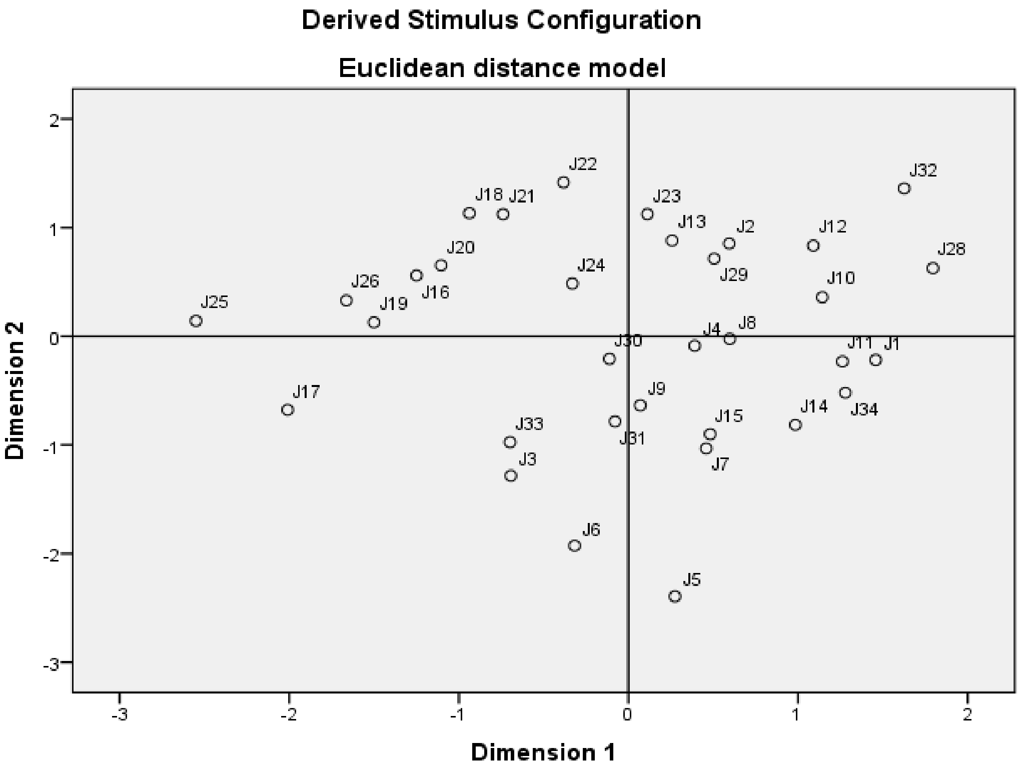

Figure 96. Result of MDS (Discrimination Data)

Similarly, the percentages of correct discrimination of all the vowel pairs in all of the consonantal contexts are submitted to Multidimensional Scaling (ALSCAL). The result is shown in Figure 96. Compared with the result of the identification aspects of the study, individual differences are small, but, still, J25 is the most distant from the group average.

This participant is by far the most proficient of all. Some other participants have lived overseas, but, judging from the results of the experiments, their perceptions of English are not so drastically different from those who studied English in Japan.

9. Conclusion

The results of this study have revealed that in general, postvocalic /n/ and /l/ decrease the identification and discrimination accuracy of American English vowels by Japanese speakers. This is not merely because vowels become less distinctive in these contexts, but also because American English vowels become more deviant from the images that

Japanese speakers hold.

These images can largely be attributed to the Japanese adaptation of English phones, i.e. *katakana* transcription of loanwords from English. Previous researches suggest that this convention has been strongly influenced by Japanese scholars in the 19th Century, who in turn, were influenced by foreign advisors from Britain. Thus, /ɑ/ as in *hot* or *pot* is adapted as *o* (*hotto*, *potto*). /i, ɪ, ɛ, æ, ɑ, ʌ/ are commonly transcribed as *ii*, *i*, *e*, *a*, *o*, *a* respectively. Both /æ/ and /ʌ/ are transcribed as *a*, but because /æ/ is usually represented by the letter *a,* Japanese speakers seem to believe that /æ/ is closer to the Japanese *a*. Japanese speakers do not seem to have a clear image of /ʌ/. These images exercise a strong influence on the perception of English vowels by native Japanese speakers.

Simply put, Japanese speakers are most likely to identify English vowels when English vowels sound like the images they hold. For instance, even in the same consonantal context, /i/ tokens that are equated with long *ii* are more likely to be correctly identified than those that are not. English vowels uttered before /n/ and /l/ are more distant from the images held by Japanese speakers. The difference in syllable-structure may also come into play. Japanese has a very simple syllable structure, and most of the syllables are open syllables, due to which vowel quality is hardly affected by a coda consonant. American English /n/ and /l/ in the coda position change the vowel quality in a way that is not possible in Japanese.

What explains this strong influence of the image of English vowels on the perception of English vowels by Japanese speakers? Limited exposure to authentic English and a flood of loanwords from English may be the answer. As I pointed out in the introduction, pronunciation is not taught sufficiently or adequately, and young children learn English words as loanwords rather than as English vocabulary.

The results of this study also revealed some individual differences that exist among Japanese participants. Some participants produced /æ/ as a front vowel while others did not. Those who produced /æ/ as a front vowel are more sensitive to differences between /æ/ and /ʌ/, and they discriminated /æ/-/ʌ/ better, also identifying /ʌ/ better than those who did not produce /æ/ as a front vowel. Individual differences were also found with regard to the sensitivity to qualitative differences between /i/ and /ɪ/, but those who were sensitive to the difference did not always identify /i/ and /ɪ/ better. It seems that they still respond based on durational difference to correctly identify these two vowels.

How can Japanese learners of English improve their identification and discrimination accuracy of English vowels? One important factor may be the quality and quantity of input. It is difficult to drastically increase the amount of input in an EFL environment where English is not used extensively, but both educators and learners have to make a conscious effort to increase the amount of exposure to English. A study by Flege, Frieda and Nozawa (1997) on Italian immigrants to Canada revealed that those who used English less often retained a heavier Italian accent than those who used English more often. The quality of input is also important. Most textbooks use one or two talkers' utterances, but to be resilient to talker difference and differences caused by the consonantal context, one must be exposed to multiple talkers' utterances in various contexts. A study by Lively, Logan and Pisoni (1993) on native Japanese speakers' identification of English /l/ and /r/ demonstrated that Japanese speakers trained using multiple talkers' tokens were able to generalize to novel words and an unfamiliar talker's utterances, while those who trained using a single talker's utterances could not generalize to a new talker. Further research is necessary, but a similar result is expected from vocal identification training, and one needs to be exposed to varied speeches uttered by multiple talkers in order to be a better communicator in English.

References

Aoyama, K., Flege, J., Guion, S., Yamada, T., & Akahane-Yamada, R. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /r/ and /l/. *Journal of Phonetics,* 23, 233-250.

Beddor, P. S. (1993) The Perception of nasal vowels, In Huffman, M. K. & Krakow, R. A. (eds.) *Nasals, Nasalization, and the Velum Volume 5 in Phonetics and Phonology* (pp. 171-196). Cambridge, MA: Academic Press.

Beddor, P. S., Karakow, R. A. & Goldstein, L. M. (1986) Perceptual constraints and phonological change: a study of nasal vowel height. *Phonology Yearbook* 3: 197-217

Best, C. (1995) A direct realist view of cross-language speech perception. In Strange, W., (ed.) S*peech Perception and Linguistic Experience* (pp. 171-204). Baltimore, MD: York Press.

Boberg, C. (2015) North American English. In Reed M. & Levis, J. M. (eds.) *The Handbook of English Pronunciation* (pp. 230-250). West Sussex: Wiley Blackwell.

Boddernberg, M. (2011) *English Loanwords in the Japanese Language*. GRIN Verlag.

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. & Steinlen, A K. (2003) Consonantal Context affects Cross-Language Perception of Vowels, *the Proceedings of the 15th International Congress on Phonetic Sciences* (ICPhS 2003). pp. 2289-2292.

Bradlow, A. R. (1993a) *Language-Specific and Universal Aspects of Vowel Production and Perception: A Cross-Linguistics Study of Vowel Inventories*. Ithaca, NY: DMLL Publications, Cornell University.

Bradlow, A. R. (1993b) A Cross-language comparison of vowel production and perception: language-specific and universal aspects, *Working Papers of the Cornell Phonetics Laboratory,* 8:29-85.

Bundgaard-Nielsen, R., Best, C. T., & Tyler, M. D. (2011) Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics,* 32, 51-67.

Chomsky, M. & Halle, M. (1968) *The Sound Pattern of English*. The MIT Press.

Cruttenden (2014) *Gimson's Pronunciation of English*. Routledge.

Crystal, D. (1997) *A Dictionary of Linguistics and Phonetics,* 4th Edition. Oxford: Blackwell Publishers.

Davenport, M. & Hannahs, S. J. (2005) *Introducing Phonetics and Phonology.* Second Edition. London: Hodder Education.

Deterding, D. (2015) Segmentals. In Reed M. & Levis, J. M. (eds.) *The Handbook of English Pronunciation* (pp.69-84). West Sussex: Wiley Blackwell.

Edwards, H. T. (2003) *Applied Phonetics: The Sounds of American English* Third Edition. Clifton Park, NY: Thomson Delmar Learning.

Flege, J. E. (1995) Second language speech learning theory, findings, and problems. In Strange, W., (ed.) *Speech Perception and Linguistic Experience* (pp. 223-277). Baltimore, MD: York Press.

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997) Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, 25, 169-186.

Flege, J. E. & McKay, I. (2004) Perceiving vowels in a second language. S*tudies in Second Language Acquisition,* 26, 1-34.

Frieda & Nozawa (2007) You are what you eat phonetically: The effect of linguistic experience on the perception of foreign vowels. In Bohn, O. –S. & Munro, M. J. (eds) *Language Experience in Second Language Speech Learning in Honor of James Emil Flege* (pp. 79-96). Amsterdam: John Benjamins Publishing.

Fujimura, O. (1972) *Onseikagaku (Phonetic Science)*. University of Tokyo Press.

Giegerich, H. J. (1992) *English Phonology An Introduction.* Cambridge University Press.

Harnsberger (2001) On the relationship between identification and discrimination of non-native nasal consonants, *Journal of the Acoustical Society of America,* 110, 489-503.

Hillenbrand, J. & Clark, M. J. (2001) Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America*, 109: 749-863.

Hillenbrand, J., Getty, L. A., Clark, M, J. & Wheeler, K. (1995) Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97:3099-

3111.

Hirata Y. & Tsukada, K. (2003) The Effects of Speaking Rates and Vowel Length on Formant Movements in Japanese. P*roceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in Speech Production and Perception*. pp. 73-85.

Hirata, Y. & Tsukada, K. (2009) Effects of Speaking Rate and Vowel Length on formant frequency displacement in Japanese. *Phonetica*, 66:129–149.

Homma, Y (1969) Nichieigo Boin no Onkyo Onseigakuteki Kosatsu (Thoughts on Vowels of Japanese and English Based on Acoustic Phonetics). *Shuryu*, 31: 25-41, English Literary Society of Doshisha University. (reprinted on Homma, Y. (ed.) (1992) *Acoustic Phonetics in English & Japanese* (enlarged and revised edition, pp. 1-21). Kyoto: Yamaguchi Shoten.

Imanishi, M & Miwa, J. (1989) Boin no Onkyoteki Tokucho (Acoustic characteristics of vowels) in Sugito, M. (ed.) *Koza Nihongo to Nihongo Kyoiku vol. 2 Nihongo no Onsei On-in (Japanese and Japanse Education vol. 2 Japanese Phonetics and Phonology)* (pp. 85-108).

Ingram, J.　C. L. and Park. S. G. (1998) Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners. *The Journal of the Acoustical Society of America*, 103: 1161-1174.

Inui, T. (2010) *John Manjiro no Eikaiwa (Japanese First Manual of English Conversation by John Mung).* Tokyo: J Research Shuppan.

Irwin, M. (2011) *Loanwords in Japanese (Studies in Language Comparison Series 125)*. Amsterdam: John Benjamins Publishing.

Jespersen, O.　(1926) *Lehrbuch der Phonetik*. Berlin: Teubner.

Johnson, K. (2012) Acoustic and Auditory Phonetics Third Edition. West Sussex: Wiley-Blackwell.

Kasahara, K., Machida, N., Osada, E., Takahashi, T. & Yoshizawa, S. (2012) Vocabulary Knowledge of 5th and 6th Graders at Elementary School: Connection between sound, meaning and spelling. *JES Journal*, 12: 90-101.

Kobayashi, Y. (2005) *Nichiei Gairaigo no Hatsuon (Pronunciation of Loanwords in Japanese and English).* Hiroshima: Keisuisha.

Kori, S. (2018) *Nihongo o Bogoto suru Daigakusei ga motsu Eigo no hatsuon no imeeji to Nihongo ni tsuiteno tadashii hatsuonkan (Image of English Pronunciation that University Students whose L1 is Japanese, and how they View the Correct Pronunciation of Japanese). A Paper Presented at the June Meeting of Kinki Society of Phonetic Science.*

Krakow (1993) Nonsegmental influneces on velum movement patterns: Syllables, sentences, stress, and speaking rate, In Huffman, M. K. & Krakow, R. A. (eds.) N*asals, Nasalization, and the Velum Volume 5 in Phonetics and Phonology* (pp. 87-116). Cambridge, MA: Academic Press.

Labov, W. (1994) *Principles of Linguistic Change Internal Factors*. Oxford: Blackwell.

Labov, W. (2010) *Principles of Linguistic Change Cognitive and Cultural Factors.* Malden, MA: Wiley-Blackwell.

Labov, W., Ash, S. & Boberg, C. (2006) A*tlas of North American English: Phonetics, Phonology & Sound Change.* Mouton de Gruyter.

Ladefoged, P. & Johnson, K. (2011) *A Course in Phonetics International Edition* 6th Edition. Stamford, CO: Wadworth Cengage Learning.

Ladefoged, P. (1999) American English, The International Phonetic Association (ed.) *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet* (pp.42-44). Cambridge: Cambridge University Press.

Ladefoged, P. (2001) *Vowels and Consonants: An Introduction to the Sounds of Languages*. Malden, MA: Blackwell Publishing.

Ladefoged, P. (2003) *Phonetic Data Analysis: an Introduction to Fieldwork and Instrumental Techniques.* Malden, MA: Blackwell Publishing.

Lee, B., Guion, S. G. & Harada, T. (2006) Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition*, 28, 487-513.

Lengeris, A. (2009) Perceptual assimilation and L2 learning: Evidence from the perception of southern British English vowels by native speakers of Greek and Japanese, *Phonetica*, 66: 169-187.

Liljencrants & Lindblom (1972) Numerical stimulation of vowel quality systems: the role of perceptual contrast. *Language*, 54: 839-62.

Lindsey, G. (1990) Quantity and quality n British and American vowel systems. In Ramsaron, S. (ed.) *Studies in the Pronunciation of English: A Commemorative Volume in Honour of A.C. Gimson* (pp. 106-118). London: Routledge.

Lively, S. E., Logan, J. S. & Pisoni, D. B. (1993) Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, 94, 1243-1255.

Morrison, G. S. (2002) Japanese Listeners' Use of Duration Cues in the Identification of English High Front Vowels, *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*. pp. 189–200.

Mourlton, W.C. (1990) Some vowel systems in American English. In Ramsaron, S. (ed.) *Studies in the Pronunciation of English: A Commemorative Volume in Honour of A.C. Gimson* (pp. 119-136). London: Routledge.

Mutsukawa, M. (2009) *Japanese Loanword Phonology*. Kyoto: Hituzi Syobo Publishing

Nishi, Strange, Strange, W., Akahane-Yamada, Kubo & Trent-Brown (2008) Acoustic and perceptual similarity of Japanese and American English vowels. J*ournal of the Acoustical Society of America*, 124: 576-588.

Nozawa, T. (2015) Effects of Attention and Training Method on the Identification of American English vowels and coda nasals by native Japanese listeners. In the *Proceedings of The 18th International Congress on Phonetic Sciences (ICPhS 2015)*. pp. 1-5.

Nozawa, T. (2016) The Effects of Prenasal Raising/Tensing of American English /æ/ on the perception of American English vowels by native Japanese listeners. *New Sounds 2016 Books of Abstracts.* p. 149.

Nozawa, T. (2018) How Native Speakers of Japanese and American English label vowels of each other's L1 in terms of their L1 vowel categories. Research on Phonetic Language, 12, 39-54. Graduate School of Language and Culture, Osaka University.

Nozawa, T & Cheon, S. Y. (2011) How Listeners' L1 Affects Perceived Similarity Of American English, Japanese and Korean Vowels. *Proceedings of the 17th International*

Congress on Phonetic Sciences (ICPhS 2011). pp. 1514-1517.

Nozawa, T & Cheon, S. Y. (2016) The effects of prenasal raising of American English /æ/ on the identification of American English vowels by native Japanese and Korean listeners. *Journal of the Acoustical Society of America*, 140, 3334-3335.

Nozawa, T. & Flege, J. E. (2001) Nihonjin Eigogakushusha no Eigo Boin Chikaku (Perception of English Vowels by Japanese Learners of English) *Journal of the Japan Society of Speech Sciences* 1: 17-27

Nozawa, T. & Frieda, E. (2002) Effects of perceptual assimilation of the perception of English vowels by native Japanese speakers. *Journal of the Acoustical Society of America,* 112, 2250 (abstract).

Nozawa, T. & Wayland, R. (2012) Effects of consonantal contexts on the discrimination and identification of American English vowels by native speakers of Japanese. *Journal of the Japan Society of Speech Sciences,* 13, 19-39.

Ohara, J. J. & Jaeger, J. J. (1986) *Experimental Phonology*. Orlando: Academic Press.

Olive, J. P., Greenwood, A. & Coleman, J. (1993) *Acoustics of American English Speech - A Dynamic Approach*. New York: Springer.

Ota, K. (2012) Nihon no Eigokyoiku ni okeru Onseishido no Genjo: Shoki Eigokyoiku ni okeru Onseishido no Donyu oyobi sono Kyojuho no Kakuritsu o Mezashite (Current status of voice instruction in English education in Japan: Toward the introduction of Pronunciation Instruction in Early English Education and the Establishment of its Teaching Method) *Kyushu International University Institute journal of social science and the humanities,* 69:53-73.

Ota, K. (2013)Nihon no Eigokyoiku ni okeru moten: Eigokyoiku no Genjo to Kadai (The blind spot of English education in Japan: Current status and problems of English education). *Kyushu International University Kokusai kankeigaku bulletin*, 8, 37-70.

Pennington, M. C. (1996) *Phonology in English Language Teaching: An International Approach*. London: Routledge.

Peternon, G. E. & Lehiste, E. (1960) Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32: 6893-703.

Picket (1957) Perception of vowels heard in noises of various spectra. *Journal of the*

Acoustical Society of America, 29: 613-620.

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

Pols., L. C. W., van der Kamp, L, J, Th., & Plomp, R. (1969) Perceptual and physical space of vowel sounds. *Journal of the Acoustical Society of America*, 46, 458-467.

Preston, D. R. & Yamagata, A. (2004) Katakana representation of English loanwords: Mora conservation and variable learner strategies. *Journal of Sociolinguistics*, 8: 359-379.

Quackenbush, E. (1974) How Japanese English words. *Linguistics-An Interdisciplinary Journal of the Language Sciences*, 131: 59-85.

Rogen-Revell, P. (2011) English Phonology and Pronunciation Teaching. London: Bloomsbury.

Rosner, B.S. & Pickering, J.B. (1994) *Vowel Perception and Production*. Oxford: Oxford Science Publications.

Saito, S. (1997) *Nihongo Onseigaku Nyumon (An Introduction Japanese Phonetics)* Tokyo: Sanseido.

Saji, R. & Saeki, Y. (2012) A study on word comprehension and its familiarity for Japanese 6th graders of elementary school. *JES Journal*, 12: 115-124.

Schouten, M. E. H. & van Hessen, A, J. (1992) Modeling phoneme perception. I: Categorical perception. *Journal of the Acoustical Society of America*, 95, 1842-55.

Shibata, Y., Yokoyama, S. & Tara, S. (2006) Eigo Hatsuon Shido ni kansuru kyoin no jittaichosa (Survey on teachers' awareness on pronunciation teaching). *the Proceedings of 6th Meeting of Kyushu-Shikoku Branch of the English Phonetic Society of Japan*. pp. 49-55.

Snodgrass, J. G., Levy-Berger, G. & Haydon, M. (1985) *Human Experimental Psychology.* New York: Oxford University Press.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S.A., & Nishi, K.(2001) Effects of Consonantal Context on Perceptual Assimilation of American English Vowels by Japanese Listeners, *Journal of the Acoustical Society of America,* 109, 691-1704.

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.

Sugito, M. (1996) *Nihongo no Oto (Sounds of Japanese)*. Osaka: Izumishoin.

Takebayashi (1998) *Eigo Onseigaku (English Phonetics).* Tokyo: Kenkyusha.

Teshima, M. (2011) On Teaching English Pronunciation in Japanese Secondary Education: How It Is and How It Should Be. *Journal of the Phonetic Society of Japan*, 15: 31-43.

Vance, T. (1986) *An Introduction to Japanese Phonology (Suny Series in Linguistics).* Buffalo: State University of New York Press.

Vance, T. (2008) *The Sounds of Japanese*. Cambridge: Cambridge University Press.

Wada, A. (2015) Gimukyoiku Dankai ni okeru Eigo Onsei Shido no Genjo to Kadai (The current condtion and problems of English sound instruction in compulsory education) *Bulletin of the Graduate School of Education, the University of Tokyo,* 55: 417-424.

Wiik (1965) *Finnish and English vowels*. Anales Universitatis Turkuesis, B, 94 Turku. cited in Lindsay (1990).

Yavas (2006) *Applied English Phonology*. Malden, MA: Blackwell Publishing

Zhang, Chen, Yan, Wang, Shi & Ng (2016) The Influence of language experience on the categorical perception of vowels: evidence from Mandarin and Korean. *Proceedings of Interspeech 2016*. pp. 873- 877.

Appendix A

Results of identification experiment sorted by four talkers (correct responses in bold)

/hVt/		responses in %					
	talker	heat	hit	het	hat	hot	hut
heat	T1	**87.9**	12.1	0	0	0	0
	T3	**93.9**	4.5	1.5	0	0	0
	T4	**10.6**	84.8	3.0	0	0	1.5
	T5	**34.8**	60.6	4.5	0	0	0
hit	T1	19.7	**78.8**	1.5	0	0	0
	T3	7.6	**87.9**	4.5	0	0	0
	T4	1.5	**95.5**	3.0	0	0	0
	T5	0	**89.4**	9.1	0	0	1.5
het	T1	0	4.5	**69.7**	6.1	9.1	10.6
	T3	0	0.0	**95.5**	0.0	0	4.5
	T4	0	7.6	**78.8**	6.1	0	7.6
	T5	1.5	31.8	**57.6**	4.5	1.5	3.0
hat	T1	7.6	0	4.5	**69.7**	9.1	9.1
	T3	1.5	0	3.0	**72.7**	13.6	9.1
	T4	0	0	0	**84.8**	3.0	12.1
	T5	0	1.5	19.7	**59.1**	4.5	15.2
hot	T1	7.6	0	0.0	28.8	**48.5**	15.2
	T3	0	0	1.5	56.1	**30.3**	12.1
	T4	0	0	1.5	37.9	**40.9**	19.7
	T5	0	1.5	0	18.2	**53.0**	27.3
hut	T1	1.5	0	0	24.2	53.0	**21.2**
	T3	4.5	0	0	40.9	19.7	**34.8**
	T4	0	0	0	21.2	47.0	**31.8**
	T5	0	0	3.0	16.7	33.3	**47.0**

Results of identification experiment sorted by four talkers (correct responses in bold)

/pVt/		responses in %					
		Pete	pit	pet	pat	pot	putt
Pete	T1	**66.7**	30.3	3.0	0	0	0
	T3	**57.6**	39.4	3.0	0	0	0
	T4	**63.6**	33.3	3.0	0	0	0
	T5	**1.5**	98.5	0	0	0	0
pit	T1	6.1	**80.3**	7.6	0	1.5	4.5
	T3	12.1	**71.2**	15.2	0	0	1.5
	T4	3.0	**93.9**	3.0	0	0	0
	T5	3.0	**78.8**	16.7	0	0	1.5
pet	T1	0	0	**42.4**	10.6	12.1	34.8
	T3	0	0	**95.5**	1.5	0	3.0
	T4	1.5	1.5	**72.7**	22.7	0	1.5
	T5	3.0	16.7	**69.7**	1.5	1.5	7.6
pat	T1	1.5	0	1.5	**80.3**	9.1	7.6
	T3	0.0	0	0.0	**86.4**	1.5	12.1
	T4	3.0	0	13.6	**71.2**	1.5	10.6
	T5	1.5	4.5	65.2	**13.6**	3.0	12.1
pot	T1	0	0	0	39.4	**53.0**	7.6
	T3	0	0	0	48.5	**42.4**	9.1
	T4	0	0	0	31.8	**51.5**	16.7
	T5	0	1.5	0	7.6	**80.3**	10.6
putt	T1	0	1.5	0	4.5	43.9	**50.0**
	T3	0	0	0	42.4	24.2	**33.3**
	T4	0	0	1.5	12.1	43.9	**42.4**
	T5	0	0	1.5	0.0	36.4	**62.1**

Results of identification experiment sorted by four talkers (correct responses in bold)

/pVn/		responses in %					
	talker	peen	pin	pen	pan	pon	un
peen	T1	**54.5**	43.9	0	0	0	1.5
	T3	**24.2**	74.2	0	0	1.5	0
	T4	**4.5**	89.4	4.5	1.5	0	0
	T5	**3.0**	93.9	1.5	0	0	1.5
pin	T1	65.2	**18.2**	12.1	3.0	1.5	0
	T3	1.5	**65.2**	30.3	3.0	0	0
	T4	3.0	**65.2**	27.3	3.0	0	1.5
	T5	4.5	**37.9**	51.5	3.0	1.5	1.5
pen	T1	7.6	4.5	**56.1**	27.3	1.5	3.0
	T3	3.0	4.5	**84.8**	7.6	0	0
	T4	1.5	3.0	**68.2**	9.1	3.0	15.2
	T5	0	48.5	**50.0**	1.5	0	0
pan	T1	12.1	1.5	27.3	**57.6**	0	1.5
	T3	7.6	6.1	40.9	**39.4**	0	6.1
	T4	1.5	1.5	31.8	**48.5**	4.5	12.1
	T5	0	21.2	74.2	**1.5**	0	3.0
pon	T1	0	0	3.0	27.3	**54.5**	15.2
	T3	0	0	0	16.7	**62.1**	21.2
	T4	0	0	0	30.3	**62.1**	7.6
	T5	0	1.5	0	19.7	**56.1**	22.7
pun	T1	0	0	0	33.3	45.5	**21.2**
	T3	0	1.5	0	39.4	36.4	**22.7**
	T4	0	0	3.0	18.2	48.5	**30.3**
	T5	0	0	0	21.2	39.4	**39.4**

Results of identification experiment sorted by four talkers (correct responses in bold)

/pVl/		responses in %					
	talker	peel	pill	pell	pal	pol	pul
peel	T1	**71.2**	24.2	3.0	0	0	1.5
	T3	**43.9**	51.5	3.0	0	0	1.5
	T4	**24.2**	71.2	3.0	0	0	1.5
	T5	**4.5**	89.4	3.0	0	0	3.0
pill	T1	31.8	**65.2**	3.0	0	0	0
	T3	1.5	**68.2**	28.8	1.5	0	0
	T4	3.0	**83.3**	13.6	0.0	0	0
	T5	1.5	**62.1**	30.3	1.5	0	4.5
pell	T1	4.5	3.0	**74.2**	4.5	4.5	9.1
	T3	3.0	4.5	**77.3**	9.1	0.0	6.1
	T4	3.0	9.1	**77.3**	4.5	0.0	6.1
	T5	1.5	19.7	**77.3**	1.5	0.0	0.0
pal	T1	6.1	0	7.6	**54.5**	10.6	21.2
	T3	0	0	9.1	**63.6**	9.1	18.2
	T4	0	0	90.9	**6.1**	0.0	3.0
	T5	0	7.6	68.2	**19.7**	0.0	4.5
pol	T1	0	0	1.5	34.8	**45.5**	18.2
	T3	0	0	0	12.1	**60.6**	27.3
	T4	0	0	1.5	22.7	**47.0**	28.8
	T5	0	0	0	27.3	**40.9**	31.8
pul	T1	0	0	0	12.1	69.7	**18.2**
	T3	0	0	0	10.6	60.6	**28.8**
	T4	0	0	6.1	12.1	54.5	**27.3**
	T5	0	0	0	9.1	80.3	**10.6**

Appendix B

Results of perceptual assimilation experiment sorted by four talkers

		Responses						
		hii	*hi*	*hie*	*hu*	*ho*	*hyu*	*haa*
heat	T1	**57**	7	1	1			
		5.0	4.6	1.0	5.0			
	T3	**56**	4	1		1		
		4.9	5.0	6.0		4.0		
	T4	10	**53**	1			2	1
		4.2	4.9	5.0			3.5	2.0
	T5	8	**56**	2				
		3.6	4.1	3.0				

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses						
		hii	hi	hie	he	hei	hia	hyu
hit	T1	24	**32**	6		2	1	1
		4.1	3.3	3.0		3.0	1.0	1.0
	T3	10	**51**	3	1			1
		3.4	3.5	3.3	4.0			2.0
	T4	2	**55**	2	3	2		1
		3.5	4.4	4.0	1.7	2.0		2.0
	T5		**60**		3	1		3
			3.5		3.7	6.0		2.7

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses													
		he	*hea*	*hee*	*hie*	*hi*	*hei*	*ha*	*haa*	*hia*	*hya*	*ho*	*hou*	*hu*	*hyu*
het	T1	**29**	3	2	2	1		19	1	1	4	2	1	1	
		3.7	2.3	2.0	2.5	1.0		3.4	2.0	5.0	5.5	2.5	3.0	1.0	
	T3	**50**	1	7	1	2	2	1		1	2				
		4.0	1.0	3.6	4.0	2.0	4.0	3.0		5.0	5.5				
	T4	**44**	3		0	10	3	1			1				
		4.0	3.3			3.7	4.0	1.0			3.0				
	T5	**32**			3	26	2				2				1
		3.7			3.3	3.6	2.0				2.5				5.0

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses										
		ha	*haa*	*hya*	*he*	*hee*	*hua*	*hoa*	*hea*	*hi*	*hia*	*ho*
hat	T1	18	**38**	2		2	2	1	4			
		3.8	4.0	5.0		2.5	2.0	2.0	2.8			
	T3	**29**	28	1		2		1	4	1		
		3.5	3.9	4.0		1.5		4.0	3.0	4.0		
	T4	**37**	3	9	2	1	2	2	8		2	
		3.3	2.0	4.1	2.5	1.0	3.5	1.0	4.1		3.5	
	T5	**37**	3	4	9	1	3	3	3		2	1
		3.0	2.3	4.3	3.4	5.0	3.0	2.7	2.7		1.5	1.0

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses								
		ha	haa	ho	hoa	how	hoo	hua	hia	hea
hot	T1	10	**36**	8	4	2	4	1	1	
		3.4	4.3	3.8	2.3	3.0	3.5	2.0	2.0	
	T3	18	**28**	7	3	1	2	3		5
		4.1	4.1	5.1	2.3	3.0	4.0	2.0		1.8
	T4	**39**	3	13	7	1		2		1
		4.1	3.0	4.2	3.0	1.0		1.0		3.0
	T5	**44**	1	17	2	1		1		
		3.9	2.0	4.0	1.5	1.0		2.0		

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses															
		ha	hu	ho	hoa	hou	haa	hie	he	hua	huu	hoo	hya	hi	hea	hie	hyo
hut	T1	**36**	2	20	5	3											
		4.1	2.0	3.9	2.4	3.0											
	T3	**28**	1	6	3	2	21	1	1	1	1	1					
		3.5	3.0	4.8	2.7	3.0	3.6	4.0	2.0	2.0	3.0	5.0					
	T4	**30**	11	20	1	4											
		3.0	3.2	3.7	1.0	1.8											
	T5	9	**23**	19	5	3							2	1	2	1	1
		3.9	2.8	2.9	2.8	2.0							3.0	2.0	3.0	2.0	2.0

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses										
		pii	*pi*	*pu*	*pyu*	*pya*	*pe*	*pei*	*pa*	*pie*	*pou*	*poo*
Pete	T1	**36**	23	1	3	1	1	1				
		4.6	5.0	4.0	4.3	2.0	6.0	1.0				
	T3	**37**	28						1			
		4.9	5.5						6.0			
	T4	30	**33**				1			1	1	
		4.5	4.6				3.0			2.0	3.0	
	T5	4	**61**				1					1
		5.5	5.5				1.0					6.0

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses										
		pi	*pi*	*pe*	*pie*	*pyu*	*pu*	*pei*	*po*	*poo*	*paa*	*pya*
pit	T1	**42**	3	9		4	4	3				
		3.7	4.0	3.6		4.0	3.8	2.3				
	T3	**54**	2	5	2			1	1	1		
		4.1	4.0	3.6	4.5			1.0	3.0	3.0		
	T4	**52**	4	3		2		2			1	2
		3.4	3.5	4.0		6.0		2.5			3.0	4.0
	T5	**56**	2	2			4		2			
		4.2	3.0	6.5			2.5		5.0			

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses													
		pe	*pee*	*pea*	*pa*	*pa*	*pu*	*pua*	*puu*	*po*	*poa*	*pa*	*paa*	*pi*	*pyu*
pet	T1	15	1	3	6	1	**19**	4	4	9	1	1	1	2	1
		3.9	4.0	3.0	2.2	3.0	3.2	3.5	3.0	3.0	2.0	2.0	3.0	2.5	2.0

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses														
		pe	pee	pea	pa	pu	pua	po	pi	pyu	pyo	pie	pei	pya	pia	pii
pet	T3	**44**	11	2	1			1		1	1	2	1			
		4.1	2.3	2.5	3.0			6.0		5.0	1.0	4.0	4.0			
	T4	**37**	5		11	4	3	3	1			1		1		
		4.1	3.0		3.4	2.8	4.0	2.3	5.0			6.0		5.0		
	T5	**37**			1	10		2	9			2	2	1	1	1
		3.9			2.0	3.4		2.5	2.4			2.0	4.0	5.0	2.0	3.0

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses													
		pa	*paa*	*pe*	*pea*	*pee*	*poa*	*pya*	*pua*	*po*	*pia*	*pu*	*pie*	*pei*	*pou*
pat	T1	**27**	25	1	1	2	2	2	5	1					
		4.9	4.6	3	3	2	2.5	5	3.2	3					
	T3	**25**	11	4	5	1	3	6	7	1	1	1			
		3.6	3.2	4	2.8	1	1	3.7	3.4	4	2	2			
	T4	15	5	1	**19**	3	2	8	5	2	6				
		3.3	2.2	4	4	2.3	2	3.8	3	3	3.2				
	T5	15		**33**	1	1	1	2	3	3	1	3	1	1	1
		2.3		4.1	2	3	1	2	1.7	2.3	3	2.7	3	5	2

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses											
		pa	*paa*	*po*	*poo*	*pua*	*pi*	*poa*	*pou*	*pia*	*pu*	*pea*	*pe*
pot	T1	19	**21**	8	10	3	2	3	1				
		4.2	4.2	4.1	4.0	4.3	3.5	1.7	3.0				
	T3	**31**	13	10	2	3	2	4		1			
		5.2	5.2	5.1	3.0	2.0	2.0	1.8		1.0			
	T4	**34**	7	10	3	7	2	1			1		
		4.8	4.1	4.7	3.0	2.4	3.5	1.0			3.0		
	T5	**26**	1	24		6	1				1	1	1
		4.2	4.0	4.3		2.5	1.0				7.0	1.0	3.0

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

putt		Responses													
		pa	*pu*	*pua*	*po*	*pou*	*pi*	*paa*	*puu*	*pea*	*poa*	*pe*	*pee*	*poo*	*pyu*
	T1	20	**21**	8	11	6	1								
		4	3.8	2.9	4.3	2.3	5				0				
	T3	17	**19**	9	2	1	1	2	2	4	2	1	1	1	
		3.4	3.8	3	5	3	1	1.5	2.5	4	1.5	3	1	3	
	T4	14	**20**	7	16	3	1	1		1	1			1	1
		2.9	3.2	3.3	3.8	2	7	2		2	2			4	2
	T5	4	**47**	2	8	1	1		1		1				1
		4.3	4	3.5	2.8	2	1		2		3				1

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses					
		pii	*pi*	*pie*	*pei*	*pyu*	*pea*
peen	T1	**58**	7	1			
		4.9	5.7	4.0			
	T3	27	**36**	1	2		
		4.0	4.7	7.0	3.0		
	T4	6	**58**			1	1
		3.5	4.4			3.0	4.0
	T5	2	**64**				
		2.0	4.9				

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses										
		pi	*pii*	*pe*	*pea*	*pei*	*pee*	*pia*	*pie*	*pyu*	*pu*	*pya*
pin	T1	**25**	18	6	2	3	9	1	4	1		
		2.9	4.2	3.3	1.5	2.3	2.9	1.0	2.3	2.0		
	T3	23	1	**30**	1	4	1	1	5			
		3.7	4.0	3.6	4.0	3.0	3.0	1.0	3.0			
	T4	**39**		20	1	3	1	1	1			
		3.7		3.7	3.0	2.0	2.0	4.0	3.0			
	T5	23		**32**		4	1		3		2	1
		3.1		4.4		2.8	2.0		3.7		1.0	4.0

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses												
		pe	*pee*	*pea*	*pa*	*paa*	*pie*	*pua*	*po*	*poa*	*pi*	*pu*	*pei*	*pii*
pen	T1	**28**	13	3	10	7	2	1	1	1				
		5.0	3.2	2.0	4.6	3.7	2.0	1.0	6.0	3.0				
	T3	**55**	4	1	6									
		4.1	3.0	1.0	3.2									
	T4	**50**	1	2	1				1		3	5	3	
		4.6	3.0	3.5	1.0				3.0		2.7	2.4	4.0	
	T5	**31**					5				24		4	1
		3.8					0.3				2.9		3.3	1.0

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses														
		pa	paa	pe	pea	pee	pei	pie	pya	pua	pyu	po	poo	pii	pia	pi
pan	T1	6	9	9	12	**17**	2	6	1	2	1	1	1			
		3.3	4.1	3.9	4.0	3.5	4.0	3.7	6.0	2.0	2.0	4.0	3.0			
	T3	1	1	16	10	**21**	3	7	3	2				1		
		2.0	3.0	3.1	4.2	3.2	4.0	4.9	3.3	2.0				3.0		
	T4	5		17	**20**	5		7	5	4	1				1	
		2.2		4.4	2.8	3.2		3.6	2.6	2.3	3.0				4.0	
	T5	2		**48**	2		3	5			1				1	4
		2.5		3.7	1.5		3.0	3.4			3.0				1.0	2.3

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses									
		pa	*paa*	*po*	*poo*	*pou*	*pua*	*pi*	*pia*	*poa*	*pea*
pon	T1	8	**22**	9	19	3	3	1	1		
		4.0	5.0	5.2	4.5	3.0	4.3	2.0	6.0		
	T3	4	**22**	9	21	2	6	1		3	
		2.0	4.2	4.3	4.8	4.5	2.2	6.0		2.7	
	T4	**18**	10	17	9	2	5			1	1
		4.4	3.8	4.2	3.4	2.0	3.2			6.0	6.0
	T5	**32**	3	26	1	2				2	
		4.2	2.3	4.1	2.0	1.5				4.0	

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses												
		pa	paa	pua	po	pou	poo	pea	poa	pu	pya	puu		pe
pun	T1	**27**	5	6	21	4	4	1	1					
		5.1	5.2	2.3	4.8	3.5	3.5	7.0	3.0					
	T3	**33**	9	5	11	2	4		1	1				
		4.9	3.3	1.4	4.5	4.0	3.8		5.0	2.0				
	T4	**30**	1	5	22	3	2		2		1			
		4.6	2.0	2.2	4.0	4.3	3.5		3.0	0.0	5.0			
	T5	**35**	1	2	20		1			3	1	1		2
		3.6	1.0	2.0	3.0		2.0			2.0	3.0	2.0	2.0	2.0

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses										
		pii	*pi*	*pyu*	*pyo*	*pie*	*pia*	*po*	*pei*	*paa*	*pu*	*pya*
peel	T1	**34**	11	15	1	2	2	1				
		4.4	4.2	4.3	3.0	3.5	2.5	1.0				
	T3	**20**	19	16	1	3	4	1	1	1		
		4	4	4	1	3	3	3	2	2		
	T4	8	**28**	14	4	3	7			1	1	
		4.1	4.4	3.3	2.8	1.7	2.4			1.0	1.0	
	T5		**41**	13	3		6				1	2
			3.9	4.4	3.3		3.2				3.0	2.0

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses													
		pi	*pii*	*pie*	*pyu*	*pei*	*po*	*paa*	*pa*	*pya*	*pu*	*pe*	*pea*	*pee*	*pia*
pill	T1	17	**18**	3	22	2	1	1	1						1
		3.9	68.0	2.7	3.5	4.0	1.0	1.0	1.0						3.0
	T3	**22**	5	3	16	7				2	1	7	2	1	
		3.3	3.4	1.3	3.2	2.3				4.0	1.0	2.4	3.0	3.0	

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses													
		pi	*pii*	*pie*	*pyu*	*pei*	*po*	*pa*	*pya*	*pe*	*pia*	*poa*	*pou*	*pyo*	*pua*
pill	T4	**33**		2	14	3	1		1	7	3	1	1		
		3.5		1.0	3.3	1.0	2.0		2.0	2.0	1.7	1.0	2.0		
	T5	**32**	1	1	16	1	1	1		6	2	1		3	1
		3.4	2.0	3.0	2.6	1.0	1.0	1.0		1.3	4.0	1.0		2.3	1.0

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses														
		pe	pee	pea	pei	paa	pi	pie	pyu	poa	pou	piu	poo	pia	pyo	pya
pell	T1	**19**	15	6	4	2	1	3	3	1	10	1				
		4.0	2.7	2.3	4.8	2.5	1.0	2.3	1.0	5.0	3.0	5.0				
	T3	**27**	2	6	5	1	4	6	3	1	1		1	3	1	2
		3.2	2.0	1.7	2.2	2.0	2.3	2.8	1.7	2.0	2.0		1.0	2.0	5.0	2.5

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses												
		pe	*pee*	*pea*	*pei*	*paa*	*pi*	*pie*	*pyu*	*pou*	*pia*	*po*	*pyo*	*pya*
pell	T4	**34**	2	8	3	1	1	4	4		1	2		4
		3.5	2.0	1.3	2.0	1.0	5.0	1.8	1.8		1.0	2.5		3.5
	T5	**33**	1	5	1		9	3	10	1	1	1	1	
		2.6	2	2	1		2.8	3	2.7	2.0	1.0	4.0	1.0	

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses													
		pa	*paa*	*pe*	*pea*	*pei*	*pee*	*poa*	*pou*	*pi*	*pia*	*pie*	*pua*	*poo*	*pyu*
pal	T1	7	**19**	5	4	1	9	1	7	2	1	4	5	1	1
		4.1	2.9	3.6	2.3	4.0	4.2	2.0	2.5	3.0	1.0	2.3	4.0	6.0	1.0

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses													
		pa	*paa*	*pe*	*pea*	*pee*	*poa*	*pou*	*pia*	*pie*	*pua*	*poo*	*pyu*	*pya*	*pu*
pal	T3	**18**	15	5	4	4	2	7	1	1	5	1	1	2	2
		3.4	3.1	1.8	2.0	1.8	6.0	2.5	2.0	1.0	3.4	2.0	2.0	2.0	1.5

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses														
		pa	*paa*	*pe*	*pea*	*pei*	*pee*	*poa*	*pou*	*pi*	*pia*	*pie*	*pua*	*pyu*	*pya*	*po*
pal	T4	5	1	**33**	12		3	1		1	1	4	1	1	3	
		2.2	2.0	3.7	3.2		3.0	2.0		3.0	1.0	2.3	1.0	1.0	2.0	
	T5	13	1	**26**	6	2	1		2	2	1	2	2	3	2	2
		3.1	4.0	3.0	1.7	1.5	3.0		1.5	4.0	2.0	2.0	2.0	1.3	2.5	4.5

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses											
		po	*pou*	*poo*	*poa*	*pa*	*paa*	*pua*	*pya*	*pi*	*pea*	*pu*	*pe*
pol	T1	3	**25**	15	3	4	12	2	1				
		4.0	5.0	4.5	2.7	2.5	4.3	2.5	5.0				
	T3	5	22	**26**		2	8	2		2			
		3.4	4.6	4.3		3.0	2.8	1.5		2.0			
	T4	**17**	11	8	7	14	2	3		1	2	1	
		3.6	3.5	3.0	1.9	3.1	4.0	2.7		3.0	2.5	1.0	
	T5	**21**	12	11	3	12	4	2					1
		4.0	3.3	3.6	1.7	4.3	2.3	2.0					3.0

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

		Responses														
		pa	pyo	pu	po	pou	poo	pei	pyu	puu	pi	pie	pe	pea	paa	poa
pul	T1	2	1	1	5	30	26	1								
		3.5	3.0	3.0	3.6	4.9	4.5	1.0								
	T3			4	15	12	31		1	2	1					
				2.8	4.9	4.3	4.5		1.0	4.0	3.0	2.0				
	T4	6		1	33	15	3		1			1.5	2	1	1	1
		2.7		6.0	3.6	4.2	3.0		2.0				3.0	1.0	1.0	5.0
	T5	1			38	12	12						1			2
		6.0			4.4	3.9	4.3						5.0			3.0

Note. The numbers in upper stand imply the number of responses.

The numbers in lower stand show mean category goodness rating

APPENDIX C

Language Background Questionnaire

Subject No: ________

Language Background Questionnaire

Name____________________________(please print) Today's Date ___________

Gender: male / female

Date of Birth ____________________

Place of Birth ________________________________ (都道府県、市)

Native Language ______________________

How would you rate your English proficiency?

(Very poor 1--2--3--4--5--6--7--8--9--10 Very good)

Your TOEIC score_____________ TOEFL___________________

Do you speak other foreign languages? (Yes / No)

If **YES** which one(s)? ______________________

How would you rate your speaking skills for this language on a scale from 1-10?

(1= very poor; 10=very good)

What are your parents' birthplaces (都道府県、市)?

Mother ____________________ Father ______________________

What are your parents' native languages?

Mother ____________________ Father ______________________

Have you ever lived outside of JAPAN? Y / N

If **YES**, for how long and where? __________________________________

APPENDIX D

Survey on English pronunciation

大学入学前に発音記号を習いましたか？

習った

多分習った

どちらともいえない

多分習っていない

習っていない

あなたは発音記号をどの程度読めますか？

ほぼ完璧に読める

ある程度は読める

どちらともいえない

あまり読めない

ほとんど読めない

英語学習において発音はどれほど重要だと思いますか？

とても重要

ある程度重要

どちらともいえない

あまり重要ではない

まったく重要ではない

あなたの英語の発音はどれくらいだと思いますか？

まったくよくない

あまりよくない

よくわからない

どちらかといえばいい

とてもいい

beat と bit の発音はどう違うと思いますか？

bat と but の発音はどう違うと思いますか？

hot と hut の発音はどう違うと思いますか？

think の最初の音はどう発音すると思いますか？

right と light の発音はどう違うと思いますか？