PERCEPTION OF PLACE-OF-ARTICULATION CONTRASTS OF ENGLISH WORD-FINAL CONSONANTS IN CONNECTED SPEECH BY JAPANESE ADULT L2 LEARNERS

By

KIKUYO ITO

A dissertation submitted to the Graduate Faculty in Speech-Language-Hearing Sciences

in partial fulfillment of the requirements for the degree of Doctor of Philosophy,

The City University of New York

2012

This manuscript has been read and accepted for the Graduate Faculty in

Speech-Language-Hearing Sciences in satisfaction of the Dissertation requirement for the degree of Doctor of Philosophy.

Date	Winifred Strange, Ph.D. Chair of Examining Committee
Date	Klara Marton, Ph.D. Executive Officer
	Valerie L. Shafer, Ph.D.
	Lisa Davidson, Ph.D.
	Supervisory Committee
	Douglas Whalen, Ph.D.
	Outside Reader

THE CITY UNIVERSITY OF NEW YORK

Abstract

PERCEPTION OF PLACE-OF-ARTICULATION CONTRASTS OF ENGLISH WORD-FINAL CONSONANTS IN CONNECTED SPEECH BY JAPANESE ADULT L2 LEARNERS

by

Kikuyo Ito

Adviser: Dr. Winifred Strange

This study investigated the perception of place-of-articulation contrasts of English word-final stops /p-t-k/, /b-d-g/, and /m-n-ŋ/ followed by a word-initial /p/, /t/, or /k/ in sentences by adult Japanese second language (L2) listeners of English and by native American English (AE) listeners. Minimal triplets differing in place of articulation (e.g., *sip*, *sit*, and *sick*), followed by *positively*, *tauntingly*, or *cautiously* were embedded in a carrier sentence and were recorded in clear and in casual fast speech. Detailed acoustic analysis was carried out, showing that the availability of acoustic cues signaling the place information of the target stops was very consistent as a function of speech mode and type of the target stimuli. Participants listened to sentences, such as, *He said the word sit positively* (or *tauntingly* or *cautiously*), and identified the target words by choosing one of three written options. Identifying place of articulation of word-final oral stops was expected to be challenging for Japanese listeners, especially when the stop is unreleased, because word-final oral stops do not exist in Japanese. Place identification of word-final nasal stops was also expected to be difficult for Japanese listeners because of the place assimilatory nature of Japanese syllable-final nasals that results in the realization of the final nasal as [m], [n] or [ŋ] depending on the place of the following segment.

Japanese listeners' perceptual difficulty was evident in results when word-final oral stop releases were absent/crucially reduced, indicating their heavy reliance on the releases. Japanese

listeners also showed marked difficulty in correctly perceiving word-final nasal stops even in clear speech, contrasting with AE listener's ceiling performance regardless of the speech mode. Positive correlations of the performance by Japanese listeners were seen with their length of residence in English-speaking countries (LOR) and with their English proficiency. Negative correlations were seen with their age of arrival in English-speaking countries (AOA), mainly in fast speech.

Results indicate that L2 perception in context may be considerably improved by clearly articulated speech when the problems are due to reduced availability of acoustic cues and that clearer speech may not be very effective when the problem stems from a language-specific perceptual pattern affected by the phonological rules of listeners' first language.

Dedication

To my late mother, Kumiyo Ito

亡き母、伊藤久美代に捧ぐ

Acknowledgments

I would like to thank my dissertation adviser, Dr. Winifred Strange for her selfless love that she has never stopped giving me. This dissertation would not have been possible without her dedicated support, insightful guidance and wholehearted encouragement throughout the process of writing this dissertation, even after her retirement from the faculty position. I am indebted to her for training me rigorously in her laboratory to be an empirical researcher and for financially supporting me at the same time for many years as well.

I am also grateful for the support of the members of my supervisory committee. I feel very fortunate to have the opportunity to work with such remarkable individuals with profound wisdom and scholarship. I thank Dr. Valerie Shafer for her great insight into this research project. Her careful and thoughtful comments immensely helped improve this dissertation. I thank Dr. Lisa Davidson for her prompt and helpful advice whenever I needed and for her knowledgeable input that always had valuable points that I often overlooked. I would also like to thank Dr. Douglas Whalen, who took over the role of the outside reader at the last moment. His input on the analysis of the speech stimuli was invaluable. My special thanks go to Dr. Klara Marton, the Exective Officer of the Speech-Language-Hearing Sciences program, for always making sure that I had no issues that may prevent me from completing the dissertation.

I am indebted to Dr. Franzo Law II whose help was indispensable for this dissertation, from contributing his speech for the recording of the stimuli through giving statistic advice on the data to giving feedback on an early version of the draft. I would also like to express my sincere thanks to Luca Campanelli whose intense tutorial and helpful advice on carrying out statistic testing using SPSS made it possible for me to complete the statistic analysis part of this research project. I am very grateful to Dr. Marisa Monteleone who was always willing to read

my drafts and kept giving me invaluable advice and comments along with moral support and encouragement. I thank Anthea Vivona for spending considerable time preparing the speech stimuli and for her friendship. My special thanks also go to Monica McIntyre who took time to read the final draft of this dissertation to give her feedback on it.

I am grateful to the "mass mind" of the Speech Perception and Acoustics Laboratory that kept stimulating my research interest, offering lively discussion and creative thinking, and fostering a sense of collaborative community. I would like to convey my gratitude to Dr. Glenis Long who kept giving me a great deal of helpful advice on the organization of the presentation of this project. I am very thankful to Dr. James J. Jenkins for his statistical advice and his kind encouragement. I would also like to express deep appreciation to my friends and family in the United States and in Japan, whose friendship and moral support have been a constant source of strength.

I wish to thank Bruno Tagliaferri for providing his valuable experiment presentation software Paradigm® and Gary Chant for his technical assistance. I also thank Linda Ashour and Loretta Walker for their administrative help and encouragement.

I am grateful to the National Science Foundation for the SBE Doctoral Dissertation Research Improvement Grant, which provided the financial support necessary for this dissertation.

Lastly, I owe my sincerest gratitude to my late mother, Kumiyo Ito, who had always encouraged my pursuit of higher education and had wished my successful completion of the dissertation until the last moment of her passing.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1. Introduction

In speech, gestures of the articulators in the vocal tract are constantly influenced by adjacent phonetic environments. The articulators often do not reach the intended target positions for phonetic segments (target undershoot) because of anticipatory and carry-over effects. This creates temporal overlap of articulatory movements called coarticulation, resulting in remarkable variability of phonetic segments (Daniloff & Hammarberg, 1973; Kent & Minifie, 1977; Repp, 1986; Fowler & Saltzman, 1993). In connected speech, in particular, it is not uncommon that phonetic segments in words are realized quite differently from their canonical forms because of these coarticulatory effects. For example, English word-final /t/ can be produced as released, unreleased, glottalized or flapped in certain contexts (Pickett, 1999; Raphael, Borden & Harris, 2011), and even may have articulatory/acoustic properties very close to /p/ or /k/ in cases of regressive place assimilation (e.g., Barry, 1985; Nolan, 1992). In addition to the influence of phonetic environments, various factors, such as speaking rate, style, and speaker differences may contribute to the variability of phonetic realizations of utterances. In general, fast speech is found to be more divergent from canonical forms than slow speech (e.g., Miller, 1981; Gay, 1981; Hertrich & Ackermann, 1995). Likewise, casual speech is more inclined to deviate from base forms than formal speech (e.g., Manuel, Shattuck-Hufnagel, Huffman, Stevens, Carlson & Hunnicutt, 1992), and male speech tends to have more variability than female speech (e.g., Byrd, 1994). These factors seem to interact with each other.

Although some variations are reported to be less intelligible than their base forms (e.g., Householder, 1956; Halle, Hughes & Radley, 1957; Nolan, 1992; Gaskell & Marslen-Wilson, 2001), native listeners generally seem to be capable of recovering the underlying forms of phonetic segments from coarticulated speech when they are presented in context (e.g., Sumner &

Samuel, 2005; Gow, 2002; Gow, 2003; Manuel, 1995). In the case of non-native perception, however, how well second language (L2) listeners can perceive the underlying representations of phonetic segments in connected speech has not been well explored yet. Furthermore, in spite of the fact that spoken words are usually produced in context in normal communicative situations and that phonetic realization of word-final segments is heavily affected by the following segments, the majority of perception studies on English word-final consonants, by both first language (L1) and L2 listeners, have dealt only with the perception of isolated words (with the exception of the studies on English coronal assimilation). While L2 listeners' perceptual difficulties with English word-final consonants have been documented in isolated speech (e.g., Flege, 1989), their difficulties in connected speech have not been examined. Moreover, little is known about the acoustic properties of word-final consonants as a function of the following phonetic environments. The goal of the present study was to investigate L1 and L2 perception of English word-final consonants followed by a word-initial consonant in connected speech and also to explore their detailed acoustic characteristics. Specifically, the perception of English words ending with /p/, /t/, /k/, /b/, /d/, /g/, /m/, /n/, and /ŋ/ followed by a word beginning with /p/, /t/, /k/ in a meaningful context by adult Japanese L2 listeners of English and by native American English (AE) listeners was examined.

Japanese syllable structures are predominantly consonant-vowel (CV) and do not allow syllable-final consonants except for a nasal stop. Furthermore, the place of articulation of the Japanese syllable-final nasal consonant (moraic nasal) is unmarked, that is, it can be realized as either labial, alveolar or velar, depending on the place of articulation of the following segment. Thus, for Japanese listeners, correctly perceiving place-of-articulation contrasts in English word-final stops in connected speech is expected to be challenging in two ways; if the word-final stop

is non-nasal (i.e., oral), it will be difficult because final oral stops do not exist in their L1; if the stop is nasal, it will be difficult because of the assimilatory nature of place in syllable-final moraic nasals in their L1. It is likely to be particularly challenging when the surface form is modified from a clearly articulated form due to the coarticulatory effects of the following segments, in such cases as where the stop is unreleased. The difficulty that Japanese listeners may have in perceiving word-final nasal stops is discussed more in the later part of this paper where studies related to the production and perception of the Japanese moraic nasal are reviewed.

The present study has additional variables in its design: two different speech modes – clearly articulated speech and casual fast speech (*Clear Speech* and *Fast Speech*, respectively, hereafter), and Japanese listeners' length of residence in English speaking countries (LOR) as a continuous subject variable. The acoustic analysis discussed later revealed that Clear Speech stimuli and Fast Speech stimuli had durational differences and other acoustic/phonetic differences, such as the extent to which final oral stops had releases in non-geminate contexts. Comparing the differences in acoustic characteristics between the two speech modes and examining the perceptual differences between them by the two language groups gives us an insight into what acoustic/phonetic information is mainly used for place distinction of word-final stops in connected speech by L1 and L2 listeners. In addition, the performance differences among Japanese L2 listeners as a function of LOR indicate whether or not the observed perceptual difficulties can be overcome with increased L2 immersion experience.

1.1. Theoretical Context of the Study

Three theoretical models of L2 perception that can be applied to predict the results of the present study are discussed here: the Speech Learning Model (SLM; Flege, 1995); PAM-L2 (Best & Tyler, 2007), an extended version of the Perceptual Assimilation Model (PAM; Best,

1995); and the Automatic Selective Perception (ASP) model of speech perception (Strange, 2006; Strange & Shafer, 2007; Strange, 2011), a recently developed theoretical model of L2 perception.

Proposed by Flege, the SLM was designed to account for the age-related difficulties in L2 speech production. It is based on the view that L2 learners can, in time, veridically perceive the phonetic properties of L2 speech sounds, in contrast to the arguments of the Critical Period Hypothesis (Lenneberg, 1967). The SLM claims that the mechanisms guiding successful language acquisition, including the ability to form new phonetic categories, remain intact and accessible throughout the lifespan, although the ability to discriminate differences between certain L2 categories decreases with age of L2 acquisition. The SLM proposes that language-specific and position-specific aspects of phonetic segments are specified in perceptual representations called *phonetic categories* stored in long-term memory, and that L2 production is guided by those stored representations. According to the SLM, L2 phonetic categories with relatively large perceptual differences from any existing L1 phonetic category are expected to be formed more easily than those that are more similar to L1 categories. The difficulty in forming a new category arises when an L2 phonetic segment and an L1 segment are perceptually so similar that learners cannot discriminate them. The SLM suggests that when a new L2 category is not formed because it is too similar to an L1 counterpart, the L1 and L2 categories will assimilate, leading to a "merged" L1-L2 category. When a new L2 phonetic category is established, on the other hand, the SLM speculates that the neighboring L1 and L2 categories may dissimilate from each other to preserve phonetic contrast. Regarding the correlations between L2 perception and L2 experience, Flege often sets over three years (e.g., Flege & Liu, 2001) or much longer LOR (e.g., 5 years in Flege, 1988; over 14 years in Flege & Fletcher, 1992) as a criterion for

experienced L2 learners, although his findings tend to indicate that LOR alone is not strongly correlated with the acquisition of L2 speech, compared with the age of arrival in English speaking countries (AOA) and L1/L2 use (e.g., Flege & Liu, 2001).

For the current study, the SLM would predict Japanese listeners' difficulty in correctly perceiving the place of articulation of both word-final oral and nasal stops but with different degrees of difficulties. Since word-final oral stops do not exist in Japanese, the SLM would predict that the place distinction of the English word-final oral stops will be learned as new L2 categories by Japanese L2 learners of English. On the other hand, the SLM would predict that learning the place distinction of English word-final nasal stops will be very hard for Japanese listeners because word-final nasals do exist in Japanese but with total obligatory place assimilation, causing a merged L1-L2 category. Since the SLM claims that L2 categories can be learned eventually, the theory may suggest that for perceptual mastery of English word-final nasals, Japanese listeners will need longer English experience, higher English proficiency, and/or early exposure to the English category than for mastery of easier L2 contrasts.

Although the SLM puts great emphasis on the mutual relationship between perception and production for L2 speech acquisition, it is essentially designed to explicate the process of acquiring L2 speech production. The main focus of the PAM (Best, 1995), on the other hand, is perception by naïve non-native listeners. The PAM's central premise is that unfamiliar non-native phonetic segments, or phones, are perceptually assimilated to the most articulatorily similar L1 phones. The model assumes that these unfamiliar non-native phones will be heard either as good or poor exemplars of a native phonological segment (*Categorized)*, as unlike any native phoneme (*Uncategorized*), or even as a nonspeech sound (*Non-Assimilated*). According to the PAM, successful discrimination of non-native contrasts depends on the goodness of fit of the

non-native phones to L1 categories. The PAM identifies the case where the two members of a non-native contrast correspond to different L1 categories as *two-category assimilation* (TC), the discrimination of which is predicted to be very good. Likewise, the case where the two members assimilate as equally good or poor exemplars of the same L1 category is classified as *single category assimilation* (SC), the perception of which is predicted to be very poor, and the case where the two members assimilate to the same L1 category but one is a better exemplar than the other as *category goodness* difference pattern (CG), the perception of which is predicted to be better than SC contrasts. Furthermore, the case where one nonnative phone is not categorized as a native phoneme while the other is heard as a native speech sound is identified as *Uncategorized-Categorized* (UC) assimilation, the perception of which is predicted to be very good. Finally, the PAM identifies the case where both nonnative phones are not categorized as *Uncategorized-Uncategorized* (UU) assimilation, the perception of which is predicted to be poor to moderate.

Best and Tyler (2007) extended the principles of the PAM to perception by L2 learners who are actively learning a second language (PAM-L2), and make predictions about the L2 learning process over time. In line with the SLM's claim, the PAM-L2 also predicts that there will not be much perceptual learning for L2 phones that are considered good exemplars of an L1 category. Also compatible with the SLM, the PAM-L2 considers that even very difficult L2 contrasts can be perceptually differentiated over time. Unlike the SLM, however, the PAM-L2 suggests that L2 learners develop not only (position-specific) phonetic categories but also (more abstract) phonological categories for L2. In the case of an L2 contrast that shows a CG assimilation pattern, the PAM-L2 predicts that a new L2 category will be formed eventually for the deviant L2 phone, but that no new category is likely to be learned for the L2 phone that is

perceived as a better exemplar of the L1 category. It assumes that learners will categorize the deviant phone as a new L2 phonetic variant of the L1 phonological category before developing a new L2 phonological category. If the members of the contrast show an SC assimilation pattern, the model predicts that L2 listeners will initially have great difficulty discriminating them, and that whether or not the discrimination of SC L2 phones can be learned depends on whether one of them is eventually perceived as a better or poorer exemplar of the L1 phoneme. The PAM-L2 suggests that learners would first have to perceptually differentiate a new phonetic category for at least one of the L2 phones before they could establish a new L2 phonological category. Regarding the cut-off for "experienced" L2 learners, Best and Tyler (2007) suggest that it should be set fairly low, such as 6 to 12 months of LOR, which is much shorter than Flege's settings.

As for the predictions of the results of the present study, the PAM-L2 would also predict Japanese listeners' difficulty in correctly perceiving place of articulation differences of English word-final oral and nasal stops. In the PAM's classification, when the word-final oral stops are not released, they would fall into the UU pattern, which would be quite challenging. This prediction is based on the Japanese phonotactics only allowing prevocalic oral stops where the release is obligatorily. For Japanese listeners, perceiving the existence of a word-final unreleased stop as a consonant is already challenging, let alone their place of articulation. The released final oral stops may be categorized as TC, which would not be very difficult, to the extent that the gestures were similar to word initial oral stops. In this sense, the PAM's prediction of ease or difficulty of the place identification of word-final oral stops would crucially depend on the presence or absence of the stop release. Regarding the perception of word-final nasals, a study by Aoyama (2003) administered a perceptual assimilation test to Japanese listeners using English word-final nasals produced in isolated words. The author found that the /m/-/ŋ/ and /m/-/n/

contrasts were classified as UC type predicting very good perception, but /n/-/ŋ/ contrast was classified as UU type predicting a relatively poor perception. The details of the study are discussed in the later part of this section (1.4.3. Japanese moraic nasals and place assimilation). Although Aoyama (2003) examined word-final nasals in isolated words, not in connected speech, the perceptual assimilation data of English word-final nasals by Japanese listeners are still of great value for the present study. Further consideration to the Aoyama's data in relation to the results of the present study is given in Discussion as well.

Both SLM and PAM-L2 have produced many studies of L2 perception that have supported their predictions about the role of L1/L2 perceptual similarity and relative difficulty. However, the studies examined phonetic contrasts using isolated syllable or word stimuli (e.g., Flege & Liu, 2001; Best et al., 2001). Therefore, the issues of L2 perception related to connected speech, such as the influence of clear and fast speech, have not been explicitly addressed in either of the models. The Automatic Selective Perception (ASP) model of speech perception (Strange, 2006; Strange & Shafer, 2007; Strange, 2011) addresses these issues by considering the influence of stimulus complexity and task demands on L2 perception. It is designed to explicate online processing of acoustic input employed to recognize and perceptually differentiate phonetic sequences. The ASP model predicts that L2 contrasts that represent intra-category (allophonic or free) variations in L1 are particularly challenging for L2 listeners to discriminate.

The ASP model proposes two modes of perceptual processing, a *phonetic mode* and a *phonological mode*. The phonetic mode is a context-specific mode of processing requiring attentional resources, and the phonological mode is a fully automated processing mode used for L1 speech processing that requires minimal cognitive resources. The ASP model claims that the extent to which these modes of processing are used for L2 speech perception is intricately

affected by such factors as listeners' L1 and L2 experience, complexity of the stimuli, and task structure. When stimulus materials and task structure are relatively simple, L2 learners are able to discriminate L2 phonetic contrasts in many cases, using resources in the phonetic mode. As the stimulus materials become more complex and task demands greater, listeners increasingly rely on language-specific patterns of acoustic-phonetic perception, what the ASP model calls *Selective Perceptual Routines* (SPRs) developed in the phonological mode. According to the ASP model, L1 SPRs, which are highly over-learned in adults, are efficient, automatic and highly robust even in non-optimal listening conditions. The model suggests that it is possible for a learner to develop L2 SPRs even in adulthood. It argues that beginning L2 learners are likely to use their automatic L1 SPRs initially, which are often not efficient for processing L2 phonetic segments, causing perceptual difficulties on L2 contrasts. However, because basic auditory sensory capabilities remain intact, perception of L2 contrasts usually improves with experience with the L2 phonological structures over time, by *re-educating* SPRs. Due to the influence of the L1, however, L2 SPRs may not be established in the same way as those of native listeners, even after years of immersion experience. With greater task demands, more complex stimuli, and more challenging listening conditions, L2 speakers' perception worsens more rapidly than that of native speakers, suggesting that L2 SPRs may never be as fully automated as L1 SPRs.

For the present study, the ASP model would predict that even experienced Japanese listeners may have difficulty in correctly perceiving word-final stops in connected speech because the task is more complex than just perceiving final stops of words produced in isolation. The model would also predict that Japanese listeners would show a greater disadvantage than native AE listeners in perceiving fast speech stimuli in which most oral final stops are unreleased (see the acoustic analysis section). Whereas L1 SPRs of AE utilize acoustic information of the

preceding transitional portions of the segments, which is used for the perception of unreleased final stops, L1 SPRs of Japanese are not likely to utilize these cues because Japanese does not have final oral stops. The model would also predict Japanese listeners' difficulty in perceiving the place of articulation of final nasal stops because word-final [m], [n], and [ŋ] are, in a sense, allophones of one phoneme of final nasal /N/ in Japanese. It would also predict a positive correlation between the performance by Japanese listeners and their LOR.

1.2. Production and Perception of Word-final Stops

1.2.1. Production of English word-final stops in context.

Acoustic properties of English word-final stops in connected speech were reported by Crystal and House (1988a) who examined the durational characteristics of stop consonants, using read sentences produced by slow and fast talkers. Their major findings regarding word-final stops are: a) slow talkers released word-final stops more often than fast talkers, b) voiceless stops were released more frequently than voiced stops, and c) velar stops were completely released more frequently than alveolar and labial stops. Their data further indicated that the stress pattern of the syllable and the segment following the word boundary influenced the characteristics of the final stops and that the place of articulation influenced the duration of occlusion and release. Crystal and House (1988b) additionally reported that the vocalic lengthening before a voiced consonant was not observed when preceding non-prepausal consonants. Their findings suggest that even well-observed cues in isolated speech, such as the duration of the preceding vowel for consonant voicing, are not necessarily reliable cues in continuous speech, underscoring the variable realization of word-final consonants in running speech.

Byrd (1993) examined over 2,300 read sentences in the TIMIT database produced by 630 speakers of American English and found different distribution of stops as a function of place of articulation and of voicing. Out of 1,130 sentence-final stops observed, 78% were alveolar, 13% were velar, and 9% were bilabial. He also found that 38% of them were voiced and 62% were voiceless. Presence or absence of release was also examined and the sentence-final stops were found to be released approximately 60% of the time. Bilabial stops were released least frequently (50%), followed by alveolar stops (57%), and velar stops were released most frequently (83%). No difference in the frequency rate of release between voiced and voiceless stops was seen (both 60%).

The presence or absence of final stop releases in the context of place order effect has been discussed from the viewpoint of articulatory overlap (e.g., Surprenant and Goldstein, 1988; Zsiga, 1994, 2000). Based on this notion, stops are expected to be more often released in a front-to-back consonant sequence (e.g., /tk/) than a back-to-front sequence (e.g., /kt/). This occurs because in the front-to-back sequence, the release of the front consonant (e.g., /t/ in /tk/) is expected not to be masked by the closure of the back consonant (e.g., /k/ in /tk/) whereas in the back-to-front sequence, the release of the back consonant (e.g., /k/ in /kt/) is more easily masked by the front closure (e.g., /t/ in /kt/). Zsiga (2000) reported that word-final, phrase medial stops were released 20% of the time when C2 was labial, followed by 38% for alveolars and 39% for velars, being in line with the notion of the place order effect. Davidson (2011) also found the place order effect in stop-obstruent clusters in spontaneous speech from two sources: recorded interviews from the StoryCorps project (Lamothe and Horowitz, 2006) and sentences recorded from a picture description task. The author further found in her data the tendencies of velar stops being released most often, labial stops being released least often, and alveolars being produced

with most variants, indicating that prevalence of the release was heavily affected by phonetic characteristics such as place of the stop and place of the following consonant.

1.2.2. Acoustic cues for word-final stops used by L1 listeners.

It has been consistently observed that unreleased stops are less intelligible for native listeners than released stops (e.g., Householder, 1956; Halle, Hughes & Radley, 1957). Lisker (1999) further investigated L1 perception of English final stops /p/, /t/ and /k/ preceded by different types of vowels, using isolated VC nonwords. The results revealed that the intentionally unreleased stops following monophthongs were perceived more correctly than those following non-monophthongs. Furthermore, the perception of the unreleased /k/ was found to be most affected by the preceding monophthong/non-monophthong vowel quality, compared to the /p/ and /t/, showing that the intelligibility of unreleased stops varied both with the preceding vowel and the place of the stop. Deelman and Connine (2001) also found that release-bearing English word-final consonants were detected faster by L1 listeners than release-deleted counterparts and that perception of voiced consonants was less dependent on the release than voiceless consonants. These effects were evident in non-words but were neutralized in real words, indicating lexical effects on phonetic perception of word-final consonants.

In addition to the acoustic information present in the stop itself, native listeners seem to utilize the anticipatory acoustic information of the preceding segments. A gating study by Warren and Marslen-Wilson (1987), using cross-spliced and unspliced tokens of word-final minimal pairs (e.g., *scoot/scoop*), found that the cross-spliced tokens severely disrupted the identification of final consonants even before the onset of the word-final consonant (alignment point). Warren and Marslen-Wilson (1988) further found that listeners' identification

performance increasingly improved from 50 ms before the onset of the alignment point, indicating listeners' continuous uptake of transitional acoustic-phonetic place information.

1.2.3. Non-native perception of word-final stops.

Several L2 perception studies have pointed out L2 listeners' heavy reliance on the acoustic information in the release bursts for the identification of word-final stops, the degree of which may be influenced by listeners' L1 phonotactics. Flege (1989) reported that when the release burst and closure voicing cues were removed, perception of the English word-final /t/-/d/ distinction by Mandarin listeners, whose L1 does not have word-final consonants, deteriorated significantly while native listeners did not show such a tendency. Furthermore, Flege and Wang (1989) found that native Shanghainese and native Cantonese listeners, whose native languages allow certain types of word-final consonants, were better able than Mandarin listeners to perceive the word-final /t/-/d/ distinction without burst and closure voicing, implying the involvement of listeners' L1 phonotactics in their L2 perception.

Abramson and Tingsabadh (1999) examined the place distinction of Thai final stops /p, t, k, ʔ/, which are never released audibly, by Thai and AE listeners. They found that English listeners' identification of minimal quadruplets of real Thai /CVC/ words was poorer than that of Thai listeners although phonetically trained English listeners performed better than naïve English listeners. The authors suggested that Thai listeners took advantage of the closing gestures of Thai stops including a component that compensates for the absence of release, which was less available to English listeners. Tsukada (2004) also found that bilingual Thai listeners' place discrimination of English final /p, t, k/, the release of which is optional, was as good as their place discrimination of Thai final /p, t, k/, which are never released, while non-Thai-speaking Australian English listeners' place discrimination was good only for English. Tsukada further

found that Thai/English bilinguals' place identification of unknown Korean final /p, t, k/, which are never released, was as good as their performance on English final /p, t, k/, implying positive transfer from their L1. On the other hand, English listeners' place identification deteriorated when listening to unfamiliar Thai and Korean /p, t, k/, suggesting the crucial involvement of L1 phonotactics.

The correlations of L2 listeners' perception of word-final stops with other factors, such as their LOR, AOA, and L1/L2 language use, were investigated by Flege and colleagues. MacKay, Meador and Flege (2001) found that AOA alone did not affect Italian L2 listeners' accuracy in identifying English word-initial and final consonants. However, the performance of early Italian bilingual listeners with low L1 use did not differ significantly from native performance, highlighting the importance of L2 learning at an early age and low L1 use for native-like acquisition of L2 final consonant perception. Flege and Liu (2001) reported that the effect of LOR on the identification of English final stops in noise by Chinese L2 learners was greater when the release was edited out than when the release was present. The authors also found that identification accuracy by a student group improved with longer LOR while a nonstudent group's performance did not change with their LOR, suggesting that the improvement of perceiving English final stops crucially depended on the amount of native-speaker input along with LOR.

1.2.4. Perception of word-final stops in context.

The perception studies of final consonants cited so far all dealt with word or syllable-length speech stimuli produced in isolation. There seems to be extremely limited research on the perception of word-final stops followed by another word, aside from the research dealing with assimilatory effects such as place assimilation, which is discussed later in this review. The study

by Takata and Nábělek (1990) is the only perception study of word-final consonants in sentences in a non-assimilation context. They examined Japanese listeners' perception of English word-initial and word-final consonants in noise, using the Modified Rhyme Test (MRT) that utilizes English word lists consisting of groups of 6 similar-sounding monosyllabic words differing either by initial or by final consonant (House et al., 1965). Using a modified version of the MRT (Kreul et al., 1968), in which the target word was followed by a /p/, as in *You will mark the ____ please*, Takata and Nábělek found that Japanese listeners were more adversely affected by noisy conditions than native listeners, especially when identifying word-final consonants.

While the Takata and Nábělek study is worth noting, considering that the studies examining the perception of word-final stops in sentences are extremely scarce, its main interest is the negative impact of noise on L2 consonant perception, not the L1 and L2 perception of word-final consonants in connected speech itself. Therefore, no detailed information of acoustic properties of word-final consonants followed by another consonant was provided, and the preceding and following phonetic segments were not controlled in such a way that the examination of contextual effects on perception was possible. Studies with detailed acoustic information of stimuli produced in phonetically controlled environments are needed for the investigation of the perception of word-final consonants in connected speech.

1.3. Influence of Speech Style on Acoustic Characteristics and Intelligibility

1.3.1. Acoustic characteristics of clear speech and conversational speech.

As briefly discussed earlier, the phonetic realization of English word-final stops in connected speech is modulated by such factors as speech rate and speech style, among other factors. Since the present study adopted two speech modes, Clear Speech and Fast Speech, what is known about these speech registers has to be taken into consideration as well. There has been a

body of research that examines the acoustic characteristics of conversational and clear speech in conjunction with speech intelligibility. A series of studies by Picheny, Durlach and Braida (1985; 1986; 1989), and Uchanski, Choi, Braida, Reed, and Durlach (1996) investigated the intelligibility differences between clear speech and conversational speech by presenting naturally produced nonsense sentences that were syntactically correct using real words but did not make any sense. Picheny, et al. (1985) found that the intelligibility of clear speech measured by word identification by hearing-impaired listeners was substantially higher than that of conversational speech and that these differences were independent of listener, talker, output level and frequency-gain characteristics. The acoustic analyses by Picheny et al. (1986) further revealed that the speaking rate of clear speech was substantially slower than that of conversational speech due both to pause insertion and lengthening of segments. Greater vowel reduction, fewer released stops, and smaller intensities of stop consonant bursts when they were released were also observed for conversational speech, relative to clear speech. Moreover, Picheny et al. (1989) found that intelligibility of artificially manipulated "fast clear speech" and "slow conversational speech" considerably worsened, regardless of rate, and suggested that speaking rate by itself may not be responsible for the intelligibility difference between clear and conversational speech. The study by Uchanski, et al. (1996) also replicated the Picheny et al. (1989) results, revealing that both time-scaling and pause manipulation reduced the intelligibility even when the manipulation slowed the speaking rate.

Bradlow, Torreta and Pisoni (1996) attempted to identify the acoustic characteristics that make some talkers more intelligible than others, using a multi-talker database. Their analysis revealed that female talkers tended to be more intelligible than male talkers and that fundamental frequency (f0) range correlated positively with higher speech intelligibility but that speaking rate

and average f0 did not. Larger vowel spaces and the precision of articulation were associated with high intelligibility as well, highlighting the importance of a low degree of phonetic reduction and a high degree of articulatory precision for intelligibility.

1.3.2. Increasing speech rate and its consequences.

There has been another line of research examining speech rate and its articulatory and acoustic consequences. The increase of speech rate has been associated with such factors as shortening of segmental duration, reduced articulatory displacement and increased velocity of articulators, and temporal overlap of articulation. However, these effects do not seem to occur linearly. A number of studies have indicated that fast speech involves complex nonlinear transformations produced by a reorganization of speech motor strategies that vary substantially for individual speakers. (e.g., Gay, 1981; Byrd & Tan, 1996)

Studies comparing durational differences between normal-rate speech and fast-rate speech have indicated that shortening of speech segments in fast speech is not proportional, that is, some segments are reduced more than others. For example, vowel durations tend to be reduced more than consonant durations (Gay, 1978; 1981; Max & Caruso, 1997), and durations of stressed syllables are reduced less than those of unstressed syllables (Peterson & Lehiste, 1960; Port, 1981), which may result in a more prominent prosodic pattern than normal-rate speech. It was also found that the extent of consonant reduction does not occur uniformly, showing more reduction in some consonants (e.g., stops, coronals) than others (Byrd & Tan, 1996).

It also has been observed that faster speech induces increased articulatory overlap, although considerable individual variability regarding the degree of overlap was noted in most cases. For example, Munhall & Lofqvist (1992) found completely overlapping gestures in

laryngeal abduction in fast speech, and Davidson (2006) reported more frequent occurrence of schwa elision in fast speech, regardless of phonotactic legality of the resulting surface phonetic sequences. It has been pointed out that the temporal relationships among articulators modulate as a function of speech rate as well (Shaiman, Adams, & Kimelman, 1995; Shaiman, 2001).

However, fast speech is not necessarily less intelligible than normal-rate speech. Matthies et al. (2001) found that compared to normal speech, fast (as well as clear) speech showed higher peak velocity of articulation, indicating more articulatory effort. It has been shown that increased speech rate does not necessarily result in greater undershoot (e.g., Miller, 1981; van Son & Pols, 1990; 1992; Zsiga, 1994). Furthermore, Greisbach (1992) reported that speakers with "precise" articulation were better understood than those with "lax" pronunciation in fast speech, according to impressionistic observations. Thus, the relationship between fast speech and intelligibility is not a straightforward one, under the heavy influence of different articulatory strategies for speeding up adopted by individual speakers.

1.3.3. Intelligibility of clear speech for L1 and L2 listeners.

The intelligibility differences between L1 and L2 listeners were investigated by Bradlow and colleagues by manipulating various factors, including lexical characteristics of words, speaking rate, background noise level, word predictability, and even native- and foreign-accented speech. Bradlow and Pisoni (1999) examined the identification of words by L1 and L2 listeners, using "easy" words (high-frequency words with few phonetically similar sounding "neighbors") and "hard" words (low-frequency words with many phonetically similar sounding "neighbors"). The authors found that the easy words were more intelligible than the hard words across multiple talkers and speaking rates and that fast speech reduced intelligibility but slowing down medium-rate speech did not enhance intelligibility. Their results also indicated that intelligibility

difficulties were reduced as the listeners became familiar with the talker's voice and speech patterns (talker familiarity effect), which was easily transferred to L2 perception. Furthermore, L2 listeners exhibited a strong easy-hard lexical effect even after minimizing the word familiarity factor, suggesting L2 listeners' much greater difficulty in recognizing words when fine phonetic discrimination at the segmental level is required.

Bradlow and Bent (2002) reported that L2 listeners showed a considerably smaller *clear speech benefit* (increased intelligibility from casual to clear speech conditions) than L1 listeners across noise levels and talkers, suggesting that clear speech is essentially native-listener oriented and is only beneficial to listeners with extensive experience with the phonology and phonetics of the target language. The authors argued that while the signal enhancements of clear speech, which make the signal more acoustically salient, may benefit all listeners, the code enhancements of clear speech, which exaggerate the acoustic distance between contrasting categories, are language-specific and beneficial only for well-experienced listeners of the target language. Bradlow and Alexander (2007) controlled word predictability in order to eliminate the semantic influence from their intelligibility study, using four types of sentences: high-probability clear speech, high-probability plain speech[1], low-probability clear speech, and low-probability plain speech. They found that L1 listeners benefited from both semantic and acoustic enhancements whether presented singly or in combination whereas L2 listeners benefited from semantic information only when acoustic enhancement was also present.

The findings of Bent and Bradlow (2003) further highlighted the different patterns of L1 and L2 perception of native accented and foreign accented English sentences. The study revealed that the native talker was most intelligible for L1 listeners, but for L2 listeners, the high-

[1] The term "plain speech" was adopted instead of "conversational speech" in this study.

proficiency (HP) nonnative talker from the same L1 background as the listeners was as intelligible as the native talker, showing a *matched interlanguage speech intelligibility benefit.* Furthermore, for nonnative listeners, a HP nonnative talker from a different L1 background was found to be as intelligible as, or more intelligible than the native talker, indicating a *mismatched interlanguage speech intelligibility benefit*.

1.3.4. Speech register and H&H theory.

Proposed by Lindblom (1990; 1996), the H&H theory claims that the interplay between clear and casual fast speech and their acoustic characteristics are explicable within its framework. The theory argues that speech production is adaptively organized, varying the output along a continuum of hyper- to hypo-speech. It assumes two characteristics of action systems, plasticity and economy, for the control of such adaptation. Plasticity supports "output-oriented" control that tries to meet the intelligibility demands, leading the production to more forcefully articulated "hyper" speech. Economy sets low-cost movements as default for motor activities to achieve minimal expenditure of energy, inducing "hypo" speech. Thus, the H&H theory characterizes speech production as a continual tug-of-war between output-oriented hyper-speech on the one hand and system-oriented hypo-speech on the other. The speaker makes a running estimate of the listener's need for explicit signal information on a moment-to-moment basis, then adapts the production of phonetic forms so that it minimally contains sufficient information for successful recognition.

The arguments of the H&H theory, combining the notion of hyper- and hypo-speech with clear and conversational speech, are as follows. As speech production shifts from conversational to clear speech, that is, from hypo- to hyper-speech, both duration and amplitude of speech tend to increase, whereas the temporal overlap of articulation tends to decrease. The vowels and

consonants of clear speech are expected to be closer than conversational speech to their canonical target values. Thus, the context dependence of articulatory and acoustic patterns is minimal in hyper-speech and maximal in hypo-speech. Coarticulation and reduction are typical of conversational speech, or hypo-speech, for the benefit of production but can never be more extensive than the listeners' perception and comprehension will tolerate.

Moon and Lindblom (1994) compared acoustic characteristics of English front vowels by examining English words produced in isolation at a comfortable rate and vocal effort ("citation-form speech"), and then produced the same words "as clearly as possible" ("clear speech"). The vowel durations in clear speech were longer than those in citation-form speech, and the displacements of F2 values that indicate undershoot effects were smaller in clear speech than in citation-form speech within each speaker. Based on their findings, the authors suggested the involvement of an active output-oriented reorganization of phonetic gestures adaptively tuned for the purpose of compensating for undershoot effects, supporting the aforementioned theoretical applications of clear vs. conversational speech adjustments.

1.4. Place Assimilation as a Product of Coarticulation

1.4.1. Articulatory and acoustic attributes of English coronal place assimilation.

Since the present study concerns place perception of word-final consonants in connected speech, the research dealing with coarticulatory phenomena affecting place of articulation of consonants, namely, the studies on English coronal place assimilation, are reviewed here. Although the acoustic analysis discussed later found no noticeable assimilatory effects in the stimuli of the present study, it is a necessary research area that this study should take into consideration. Reviewing this line of research is also beneficial in another sense in that it closely

examines the place distinction of English word-final stops occurring in context, which has not been comprehensively investigated elsewhere.

English coronal place assimilation (*English place assimilation*, hereafter) is a widely recognized regressive assimilatory phenomenon in which the place of articulation of a final coronal segment (e.g., alveolar stop) approximates the place of the following non-coronal segment (e.g., labial or velar stop). For example, the underlying /t/ in *cat box* may result in the surface form close to [kæp] as if it is produced as *cap box*. Although there has been a tendency to consider place assimilation a dichotomous phonological process (e.g., Chomsky & Halle, 1968), studies examining articulatory and acoustic attributes of English place assimilation have indicated that it is a more complex graded articulatory phenomenon that may be affected by various factors. Several electropalatographic studies have revealed that coronal and non-coronal constrictions may occur simultaneously when underlying coronal segments are produced in a place-assimilated manner (Barry, 1985; Kerswill, 1985; Holst & Nolan 1995). Nolan (1992) argues that the patterns of tongue contact in English place assimilation are a continuum of feature change as a function of speed and style of speech and that even without coronal contact, assimilated segments may have subtle traces of the underlying coronal gesture.

Investigating the acoustic evidence, Zsiga (1994) found place assimilatory effects in the formant transitions of vowels preceding word-final coronal /d/ that was followed by either word-initial /p/, /t/, or /k/ in English sentences[2]. The F2 and F3 transitions of the preceding vowel in real word sequences, such as *bed pan*, *bed tan*, *bed can*, were quantified by measuring the frequencies at the midpoint and at the offset and by calculating the difference between the two values for each token. The tokens produced in sentential context by four talkers at "a

[2] The author used the term "overlap" instead of "place assimilation."

conversational rate" and "a very rapid rate" were acoustically analyzed. The results revealed that the formant transitions differed in the predicted direction depending on the following stop, indicating the coarticulatory (or assimilatory) effects of the following word-initial stop. However, the formant transitions of the rapid-rate speech tokens did not show obvious differences from those of the conversational-rate speech tokens, showing no evidence of a direct relationship between increased speaking rate and increased gestural overlap. Similar analyses were done for unassimilated coronals, place assimilated coronals, and corresponding unassimilated non-coronals (e.g., /t/ in *right dairy*, /t/ in *right berry*, and /p/ in *ripe berry*, respectively) by Gow (2001; 2002; 2003) in his perception studies. He found that F2 and F3 formant transitions of the three types of final stops were different from each other, with the assimilated coronals showing values intermediate between the other two, indicating that the spectral characteristics of place-assimilated coronals differ from both unassimilated coronals and corresponding noncoronals.

1.4.2. Perception of English place-assimilated coronals

While fully assimilated English word-final coronals may assume acoustic-phonetic characteristics almost equivalent to those of underlying non-coronals, the possible phonetic ambiguity of assimilation does not seem to hinder native listeners from recognizing the intended words as long as the assimilation does not create another viable meaning (Gaskell & Marslen-Wilson, 1996, 1998; Gow, 2001, 2002, 2003; Marslen-Wilson, Nix, & Gaskell, 1995; Lahiri & Marslen-Wilson, 1991). However, in the case where the assimilation may potentially generate a semantically ambiguous context, some researchers maintain that it may interfere with speech processing while others argue that it still does not hinder listeners' word recognition. For example, Nix, Gaskell, and Marslen-Wilson (1993) found that sentences such as *They thought the lake cruise was rather boring* caused perceptual ambiguity (i.e., *lake* was often identified as

late). Nolan (1992) also demonstrated perceptual ambiguity of strongly velar-assimilated underlying alveolar stops, using minimal pairs embedded in semantically neutral contexts (e.g., *the road/rogue collapsed*).

A set of cross-modal priming experiments by Gaskell and Marslen-Wilson (2001) revealed that when place assimilation potentially created two viable meanings, as in *I think a quick ru**m** picks you up*, only *rum* but not *run* was primed in a neutral-bias context. When the sentence was preceded by a biasing semantic context favoring *run*, however, there was priming of both *rum* and *run*, suggesting that place assimilation may cause lexical ambiguity depending on the preceding sentential context. The results of a similar study by Gow (2002), however, did not show priming effects for underlying noncoronal forms (e.g., *ripe berry*) when presented fully assimilated underlying coronal forms (e.g., *right berry*) in semantically ambiguous contexts. Gow (2002) suggested that even when fully assimilated, underlying coronals still have acoustic properties that signal their coronal quality, thereby not creating lexical ambiguity under normal conditions. Interestingly, in the Gow (2002) study, when the stimulus sentences with strong assimilation were gated at the offset of the prime (e.g., *This time she tried to get the right //*), the priming effect was seen for both coronal (e.g., *RIGHT*) and non-coronal (e.g., *RIPE*) targets. The author reasoned that listeners needed context information to recover the underlying coronality of the assimilated segment and noncoronality of the subsequent segment, emphasizing the important role that context plays bidirectionally in the disambiguation of place assimilation.

Gow (2003) further found that when a partially place assimilated underlying coronal-ending prime word was cross-spliced onto the following coronal context, that is, for example, when the sentence *The plastic cone bent easily* was gated at the offset of *cone* and was spliced on to *dents easily*, the priming effect was only seen in the target *comb*, but not in *cone*. Furthermore,

he found that a word-initial labial (e.g., /b/ in *bent*) was detected more quickly when preceded by an assimilated coronal (e.g., *cone* from *cone bent*) than when preceded by an unmodified coronal (e.g., *cone* from *cone dents*) or unmodified labials (*comb* from *comb bent*). Based on these findings, Gow argued that a single assimilated segment contains both left-to-right and right-to-left context information that offers simultaneous progressive and regressive context effects, suggesting feature-parsing mechanisms that listeners employ to correctly align recovered features with segments.

1.4.3. Place-assimilatory nature of Japanese moraic nasals and its influence on perception of L2 final nasals.

Another assimilatory effect that may have a great impact on the results of the present study is place assimilation of the Japanese moraic nasal which is an obligatory phonological phenomenon in Japanese. As briefly discussed at the beginning of this paper, this language-specific phonological rule may crucially affect Japanese listeners' place perception of English word-final nasals in connected speech.

Japanese syllable-final nasals are realized as moraic and have very different phonological characteristics from non-moraic initial nasals preceding a vowel[3]. Japanese moraic nasals, represented as /N/, are the only consonant that can occur word-finally in Japanese. Whereas a non-moraic initial nasal needs a following vowel to constitute a mora, a moraic nasal stands alone as one mora. Furthermore, the place of Japanese moraic nasals is unmarked whereas the place of Japanese non-moraic nasals is always marked. That is, the moraic nasal /N/ is articulated as uvular utterance-finally or before a pause, but can be articulated at many other places because of their obligatory place assimilation to the following segment, resulting in several allophonic

[3] The mora is a subsyllabic unit most of which are open-syllable such as V and CV with the exception of moraic consonants in Japanese.

variations including [m], [n] and [ŋ] (Amamura et al., 1983; Vance, 1987; Nakajo, 1990). These unique features of Japanese syllable-final moraic nasals seem to create very distinctive perceptual patterns of syllable-final nasals by Japanese listeners.

Otake, Hatano, Cutler and Mehler (1993) found that the detection of CVN targets (e.g., *tan*) in CVNVCV words (e.g., *tanishi*) is extremely difficult for Japanese listeners whereas English and French listeners did not show such a tendency. Cutler and Otake (1994) further found that Japanese listeners detected syllable-final nasals significantly faster than syllable-initial nasals even in English words (e.g., /n/ was detected faster in *candy* than in *canopy*), whereas English listeners detected the target equally rapidly, suggesting that Japanese listeners applied language-specific mora-based segmentation procedures to English words irrespective of their appropriateness.

Otake, Yoneyama, Cutler and van der Lugt (1996) found that when asked to detect a nasal sound represented by the letter "N", Japanese listeners detected moraic nasals (e.g., *ka<u>n</u>pa*) faster and more accurately than nonmoraic nasals (e.g., *ka<u>n</u>agu*) in Japanese real words irrespective of phonetic realization of the moraic nasal ([m], [n̪], [n], or [ŋ]), whereas Dutch listeners with no knowledge of Japanese did not show such a tendency. Moreover, Dutch listeners failed to respond to bilabial moraic nasals realized as [m] (e.g. , *to<u>n</u>bo*) when /n/ was the target but did respond when /m/ was the target, in line with the categories of their native phonology. When moraic nasals were cross-spliced across following place contexts, the place mismatch between the moraic nasal and the following context did not prevent Japanese listeners from responding to the targets but caused some detectable response delays, indicating that Japanese listeners utilized the place information of the moraic nasal as anticipatory information for the following consonant. The authors argued that Japanese listeners rapidly abstracted a

unitary representation of Japanese moraic nasals from various phonetic realizations, suggesting language-specific perceptual patterns of moraic nasals by Japanese listeners.

Cutler and Otake (1998) further consolidated their argument about Japanese listeners' phonological abstraction of syllable-final nasals. In their study, listeners were asked to blend phonologically legal pseudo-word pairs containing a syllable-final nasal in Japanese and in Dutch (e.g., *ranga-serupa* in Japanese and *lempost-duidel* in Dutch) and to produce the made-up word orally (e.g., *ranpa* and *lemdel*, respectively, for the above examples). The authors found that Japanese listeners produced more assimilated than unassimilated nasals for both Japanese materials (e.g., [rampa] for *ranpa*) and Dutch materials (e.g., [lendel] for *lemdel*) while Dutch listeners produced more unassimilated than assimilated nasals for both materials. The authors interpreted the results as indicating that obligatory place assimilation in Japanese nasal-stop sequences led Japanese listeners to represent nasals as unmarked for place of articulation, which in turn led them to assimilate the nasals to the place of the following stop. For Dutch data, the authors suggested that the subjects, whose L1 allows nasal place assimilation only optionally, preserved the original place of articulation. The authors concluded that obligatory versus optional constraints in native-language phonology led to different representations of spoken-language input, whether the input is in the native or a foreign language.

Japanese listeners' perceptual confusion over the place of articulation of English word-final nasals was investigated by Aoyama (2003). The author adopted a two-forced-choice identification task to examine the perception of the word-final /m/, /n/, and /ŋ/ by native speakers of Japanese and by those of Korean, using minimal pairs such as *same-sane* and *sin-sing* produced in isolation. Aoyama found that Japanese listeners made more errors than English listeners for all final nasal contrasts (/m/-/n/, /m/-/ŋ/ and /n/-/ŋ/) and more errors than Korean

listeners for the /n/-/ŋ/ contrasts. The author also found that more errors were seen in Japanese performance for the /n/-/ŋ/ than for the /m/-/ŋ/ and /m/-/n/ contrasts, indicating Japanese listeners' considerable difficulty in distinguishing /ŋ/ from /n/ syllable-finally. Further inspection revealed that Japanese listeners misheard /ŋ/ as /n/ more than vice versa, showing perceptual bias toward alveolar for the /n/-/ŋ/ distinction.

In order to explore the explanation for the reason why the /n/-/ŋ/ contrast was particularly difficult for Japanese listeners in the framework of the PAM (Best, 1995), Aoyama (2003) additionally conducted a perceptual assimilation task in which Japanese listeners were asked to write the words they heard in the Katakana orthography, using the same stimuli. Aoyama found that /m/ was assimilated to a Kana character transcribed as /mɯ/ more than 90% of the time and /mɯ/ was not used for either /n/ or /ŋ/ whereas neither /n/ nor /ŋ/ was consistently classified with one L1 category and that the same L1 Kana characters, transcribed as /N/ and /Ngɯ/, were used for both /n/ and /ŋ/. Adopting the PAM's framework, Aoyama discussed that both /m/-/ŋ/ and /m/-/n/ contrasts were classified as UC type, the discriminability of which is predicted to be very good, but /n/-/ŋ/ contrast was classified as UU type, the discriminability of which in this case was expected to be relatively poor because of the confounding use of the same labels for the /n/ and /ŋ/. The author concluded that the finding that Japanese listeners had difficulty in correctly perceiving the word-final /n/-/ŋ/ contrast was consistent with the PAM prediction.

The above studies indicate that the language-specific perceptual patterns of Japanese syllable-final moraic nasals are likely to be applied to non-native inputs irrespective of the efficiency and that Japanese listeners have difficulty in identifying the place of articulation of word-final nasals even when the words are clearly produced in isolation. Thus, it is not hard to imagine the situation where Japanese listeners have severe problems in identifying English word-

final nasals in connected speech where the final nasal is followed by a word-initial consonant with a different place of articulation. Since this problem stems from the phonological rule of their L1, clearly articulated speech may not help improve their perception effectively and the difficulty may persist across a large range of LOR.

1.5. Design of the Study

The present study investigated L1 and L2 perception of place contrasts in English word-final oral and nasal stop consonants using a computerized forced-choice identification task. Stimulus sentences (e.g. "He said the word *sit* cautiously.") were presented one at a time and minimal triplets of target words differing only in the place of articulation of the final stop (e.g., *sip*, *sit*, and *sick*) served as the response choices. The study examined the effects of speech mode (Clear Speech *vs.* Fast Speech) in two independent groups of Japanese L2 learners of English with native speakers of American English as control groups. The type of stop (*contrast type*, hereafter: Voiceless /p/-/t/-/k/, Voiced /b/-/d/-/g/, and Nasal /m/-/n/-/ŋ/) was a repeated measures variable within each listener group. In addition, target words were embedded in sentences in which the following word began with /p/, /t/, or /k/ in all possible combinations. Thus, the effects of the following place context on the place perception of the target final stops were also examined as a repeated measures variable.

In addition to the main experiment mentioned above, a word familiarity rating task for both language groups and an English proficiency test for the Japanese groups were administered in order to investigate their correlations with the performance. The word familiarity rating task let listeners rate their familiarity with the words in written form on a 7-level scale, and the English proficiency test evaluated the Japanese listeners' spoken English proficiency using a standardized telephone test. More detailed information is described in the following Methods.

The present study chose real words, rather than non-words, for target words in order to approximate the conditions of the experiment to conversational situations in real life as closely as possible while controlling the phonetic environments of the speech stimuli. While being aware of the lexical aspect that the real words inherently bear, adopting real words was desirable for the study because it intends to investigate the use of fine-grained acoustic-phonetic information by L1 and L2 listeners as a part of the process of recognizing words in running speech. Furthermore, it was practically impossible to construct a set of minimal triplets consisting of simple monosyllabic CVC non-real words while controlling the adjacent phonetic environments in the way that the present study did. The study therefore opted to leave the lexical aspect of the target words as they are and to report the lexical effects on the phonetic perception of the word-final stops afterwards, if any, by comparing the results of the word identification with those of word familiarity ratings and with the word frequencies of the corresponding target words. Perception studies have indicated that the lexical influence on word identification is minimal in forced-choice identification tasks with closed-set response options when sufficient acoustic-phonetic information is given. For example, in the aforementioned gating studies by Warren and Marslen-Wilson (1987; 1988), word frequency effects were minimal, influencing the identification of the final consonants only when the available acoustic information was sufficiently ambiguous. Results of Ito and Strange (2009), adopting a 2 -choice identification task, also indicated that word familiarity affected Japanese listeners' English word identification only when listeners were less able to take advantage of the acoustic-phonetic information. Since the present study also adopted a forced-choice identification task, the results were expected to follow these studies.

The study also opted to minimize the control of producing the stimuli, the releasing/unreleasing of the final oral stops in particular, because one of the objectives of the

current study was to report acoustic characteristics of English word-final stops in connected speech as they were. While the speaker was asked to keep the intensity and speech rate for each speech mode as consistent as possible, he was not asked to release or not to release the final oral stops of the target words. Although it resulted in having unbalanced stimuli in terms of oral stop releasing in the design, it was a necessary decision for this study because the acoustic properties of English final stops as a function of the following context or as a function of speech mode had not been well documented. Since the present study intended to investigate the perception of English final stops in context in a listening condition close to real life situations, the stimuli were not edited for the perception experiment either.

The lexical effects were analyzed by examining the correlation of performance on the main experiment with scores on the word familiarity task mentioned above, and with word frequency of the target words based on the $SUBTLEX_{US}$ corpus (Brysbaert & New, 2009). The $SUBTLEX_{US}$ corpus is a word frequency measure newly introduced as an improved alternative to the widely used but old Kučera and Francis's frequency norms (Kučera & Francis, 1967) for American English. The $SUBTLEX_{US}$ frequency norms are based on 51millon words from subtitles obtained from American films and television episodes whereas the Kučera and Francis's norms are based on only 1million words from adult reading materials. The $SUBTLEX_{US}$ norms are reported to predict lexical decision times quite consistently, especially for short words (Brysbaert & New, 2009). This study uses the $SUBTL_{WF}$ score that indicates the word frequency per million words based on the $SUBTLEX_{US}$ corpus.

The correlations between perceptual performance of Japanese listeners and their L2 experience, English proficiency, and L2 use were also examined, adopting the following measurements for the subject variables: length of residence (LOR) and age of arrival (AOA) in

English-speaking countries as L2 experience; scores of a standardized English proficiency test called the Versant ^{TM} English Test (see Methods for details) as English proficiency; and proportion of L1 and L2 language use in their daily life as L2 use.

1.6. Hypotheses

The hypotheses of the present study were as follows:

1.6.1. Main effects of language group (AE *vs.* Japanese) and interactions with consonant type.

Perception of place of articulation of English word-final oral and nasal consonants followed by another consonant in connected speech (*in context*, hereafter) will be poorer overall for Japanese L2 listeners of English (*Japanese listeners*, hereafter) than for native American English speaking listeners (*AE listeners*, hereafter) in both Clear and Fast Speech conditions. Americans are expected to show ceiling effects in the Clear Speech condition. For Japanese listeners, correctly perceiving place of articulation of English word-final nasal stops in context will be especially hard in both conditions due to the interference from the obligatory phonological assimilation rule for place of final nasals to following stops in Japanese.

1.6.2. Main effects of speech mode (Clear *vs.* Fast) and interaction with language group.

For both Japanese and AE listeners, word-final stops in Clear Speech will be perceived more accurately overall than those in Fast Speech. The expected perceptual advantage of Clear Speech over Fast Speech is called *Clear Speech benefit*, hereafter. The Clear Speech benefit will be larger for Japanese listeners than for AE listeners because Japanese listeners are more disadvantaged by the limited availability of acoustic information in Fast Speech than are AE listeners. In other words, there will be an interaction of language group and speech mode.

For Japanese listeners, place perception of word-final oral stops will be much more difficult in Fast Speech, where most oral stops were unreleased (see Acoustic Analysis), resulting in large clear speech benefits. In contrast, Japanese listeners will show little or no Clear Speech benefit in perception of nasal stops in context. Since the expected perceptual difficulty of nasal place contrasts by Japanese listeners is L1-phonology-based, their performance will be very poor even in the Clear Speech condition.

1.6.3. Effects of following context on perception of final stops.

In Clear Speech, Japanese listeners' place identification of word-final oral stops will be poorer when followed by a stop with the same place of articulation where the final stops are unreleased. AE listeners' perception will not be affected by the following context because they will be able to tap into acoustic information available in the preceding segments even without stop releases (Warren & Marslen-Wilson, 1987; 1988).

1.6.4. Differences in perceptual accuracy across types of stimuli.

AE listeners' perceptual accuracy of word-final stops in Clear Speech will not show differences across the types of the stimuli because of ceiling effects. AE listeners' perception of word-final stops in Fast Speech will be better on nasal than on oral stops because of different levels of availability of acoustic information between nasal and oral stops. Nasal murmurs in the occlusion portion of nasal stops are likely to remain even in Fast Speech whereas the deletion or decreased intensity of stop releases in Fast Speech will reduce place information for oral stops. AE listeners will perceive voiced oral stops more accurately than voiceless stops because of less dependence on release burst information for perceiving voiced final consonants (Deelman & Connine, 2001).

Japanese listeners' perception of word-final stops in Clear Speech will be poorer on nasal than on oral stops because of the L1 phonological rule. In Fast Speech, the differences in perceptual accuracy on nasal and oral stops will be reduced because perception of both voiced and voiceless oral stops will be less accurate due to the reduction of information provided by stop releases.

1.6.5. Correlations with demographic variables.

Performance of Japanese listeners in both Clear Speech and Fast Speech will be positively correlated with their L2 experience and proficiency and will be negatively correlated with their AOA.

Chapter 2. Methods

2.1. Participants

Forty eight adult native speakers of Japanese living in the U.S. (41 females, 7 males) and a control group of 36 adult native speakers of AE (27 females, 9 males) participated in the experiment. The Japanese participants were recruited from a wide range of LOR in order to examine the correlation between their performance in the experiment and language immersion experience. Their AOA was, however, restricted to 14 years of age or older to focus on the examination of late learners of English. Their LOR ranged from 1 month to 17 years and 9 months and their AOA ranged from 14 to 40 years of age. The mean age of the Japanese group was 33 years (range: 22 to 44 years). The Japanese participants were quasi-randomly assigned to one of two test conditions: (N = 24), Clear Speech or Fast Speech. Care was taken so that the participants in the two conditions had approximately equivalent distributions in terms of LOR.

The AE participants were monolingual or near-monolingual speakers with no exposure to other languages in their early life (10 years of age or earlier). Out of 36 AE participants, only 12 were randomly assigned for the Clear Speech condition because of the expected ceiling effects in their performance. The remaining 24 were assigned to the Fast Speech condition.

Language background information about the participants was collected by means of a questionnaire. The questionnaires for Japanese and for AE are attached as Appendix B (Japanese) and C (AE). Only the subjects who passed a hearing screening test (ANSI standards 25 dB HL at 500, 1000, 2000 and 4000 Hz) participated in the experiment. The information of Japanese and AE participants is presented in Table 1.

Table 1: Biographical information for Japanese and AE participants

Language Group	Test Type	Gender Distribution	Mean Age (Range)	Mean LOR (Range)	Mean AOA (Range)
Japanese (N = 48)	Clear Speech (N = 24)	3 males 21 females	32.3 (22 – 44)	5y 8m (1m – 17y 8m)	25.5 (17 – 34)
	Fast Speech (N = 24)	4 males 20 females	33.5 (23 – 43)	5y 8m (3m – 16y)	27.5 (14 - 40)
AE (N = 36)	Clear Speech (N = 12)	4 males 8 females	30 (23 – 44)		
	Fast Speech (N = 24)	5 males 19 females	30 (22 – 41)		

2.2. Stimulus Materials

A total of 54 target words, constituting 18 monosyllabic CVC minimal triplets, each of which differed in only the place of articulation of the word-final stops, such as *sip-sit-sick*, *lab-lad-lag*, and *sum-sun-sung*, were constructed. (The places of articulation of the target words, Labial, Alveolar, and Velar, are called *target place*, hereafter.) These 18 minimal triplets were subdivided into three groups, according to the types of contrast of the final stops (*contrast type*, hereafter): Voiceless contrasts (/p/-/t/-/k/), Voiced contrasts (/b/-/d/-/g/), and Nasal contrasts (/m/-/n/-/ŋ/). Thus, each contrast type included six word-final minimal triplets, making 18 individual words per type. The vowel of the target words in each group was controlled in such a way that two triplets contained /ɪ/, two triplets contained /æ/, and the other two triplets contained either /ʌ/ or /ɑ/. The minimal triplets were chosen out of words with a monophthong only, because of Lisker's (1999) finding that unreleased stops followed by a non-monophthong were less intelligible (although the finding was regarding isolated VC syllables).

The target words were followed by one of the following three adverbs: *positively*, *tauntingly*, or *cautiously*, creating a total of 162 two-word sequences (54 target words × 3 adverbs). The onset of these adverbs was either /p/, /t/, or /k/, and all adverbs had the –ly suffix at the end. The following vowels were the same or phonetically similar to each other (i.e., /ɑ/ or /ɔ:/) that assumed the primary stress. The two-word sequences, therefore, contained 27 types of consonantal sequences at the word boundary: 9 final stops of the target words (/p/, /t/, /k/, /b/, /d/, /g/, /m/, /n/, and /ŋ/) × 3 initial stops of the following adverbs (/p/, /t/, and /k/). In addition, six CVC minimal triplets (e.g., *pick-tick-kick*, and *bet-debt-get*) and three minimal pairs (e.g., *map-nap*), each of which differed in the place of articulation of the word-initial stops, were chosen to construct a total of 72 filler sequences by combining them with the same three adverbs used in the experimental sequences[4]. The filler sentences constituting word-initial minimal triplets/pairs were added in order to force listeners to attend not only to the word-final segment but also to the word-initial segment of the target word. It was intended to bring the listening conditions of the experiment closer to a real-life listening condition than just attending to the word-final segment of the target word. All two-word combinations including the fillers are presented in Appendix A.

All two-word sequences were preceded by "He said the word …" in the recording, such as, "He said the word *sit* cautiously," making syntactically and semantically viable English sentences. The stimulus sentences were produced in two modes of speech, Clear Speech and Fast Speech, by a 29-year-old phonetically trained male native speaker of AE from the New York area. The speaker was instructed to read the sentences "as if speaking to a non-native English listener or speaking in a noisy environment" for Clear Speech production, and "as if speaking to

[4] Since word-initial velar nasal /ŋ/ is not allowed in English, only word-initial minimal pairs are possible for the /m/-/n/-/ŋ/ contrast. Thus, the words beginning with voiced oral stops were substituted to constitute the minimal triplet options (e.g., *map-nap-gap*).

a native-English-speaking friend in a hurry" for Fast Speech production. Each sentence was produced three times for each speech mode, using randomized lists. All sentences were digitally recorded in a sound-attenuated room at a sampling rate of 22,050 Hz, monaural, with 16 bit resolution, using a microphone (SHURE SM 48), using SOUND FORGE 4.5 software. The individual sentences were copied from the original recording and stored in separate files, using SOUND FORGE 4.5.

Out of three tokens of each stimulus sentence, two tokens were chosen for use in the experiment according to judgments of two native speakers of AE. In order to keep the stimuli consistent, most tokens were chosen from two sets, out of three, of the recordings, except for the tokens that were judged as disfluent or distorted by at least one of the AE listeners. Those tokens were replaced with the corresponding tokens from the third list after making sure that the replacements sounded fluent with no distortion. The chosen tokens were normalized for root mean square (RMS) amplitude to median RMS value of one third of the tokens randomly chosen from the entire sentence stimuli.

2.3. Acoustic Analysis

Acoustic analysis was carried out to evaluate the distinctive acoustic information for place contrasts available for listeners as well as to document some global differences between utterances produced in the two speech modes. The following measurements were obtained, using Sound Forge 10.0 Pro and MultiSpeech software: 1) durations of the stimulus sentences; 2) transitions of the first three formants of vowels preceding the target stops (*preceding vowels*, hereafter) measured by establishing values at the vocalic mid-point and at the offset; 3) presence or absence of releases of the word-final oral target stops (*stop releases*, hereafter); 4) magnitudes of stop releases measured by their durations and amplitudes; 5) spectral shapes of stop release

bursts; 6) formant frequencies of the preceding vowel /æ/ in the Nasal contrast stimuli. In addition, acoustic characteristics of stop occlusion illustrating the differences between voiceless, voiced and nasal stops, which were observed in the acoustic analysis, are described in the sixth subsection. Sentence durations were measured to examine overall temporal differences between the tokens produced in Clear Speech and Fast Speech. Formant transitions preceding stop closure, especially F2 and F3 transitions, provide distinctive information for place of articulation of oral released and unreleased stops as well as that of nasal stops. It is expected that the perception of word-final oral stops, especially that of non-native listeners, will be crucially affected by the presence or absence, as well as the magnitude, of the stop releases. Spectral shapes of stop release bursts in Clear Speech were examined to investigate whether they provided distinctive place information as proposed by Stevens and his colleague (Stevens & Blumstein, 1978; Blumstein & Stevens, 1979). The formant frequencies of the preceding vowel /æ/ in the Nasal stimuli were examined to see if there was a raising or "tensing" of the /æ/ preceding a /ŋ/, which could signal a cue for the place (i.e., velar) of the target nasal. Finally, the description of acoustic characteristics distinctly present during stop occlusions of word-final voiceless, voiced and nasal stops may give the bases of explanations for perceptual patterns observed in the results.

2.3.1. Durations of stimulus sentences.

The durations of the stimulus sentences, including fillers, were measured by waveforms and illustrated in bar charts in Figure 1. As can be seen, the durational differences between Clear Speech tokens and Fast Speech tokens were very large with very little variability across contrast types within each speech mode (mean proportion of clear to fast speech duration = 1.44). Thus, there were clear differences produced by the speaker as a function of instructions, and utterances

within each mode showed consistent rhythmic production of the sentences. The following sections focus on the analysis of the target stimuli.

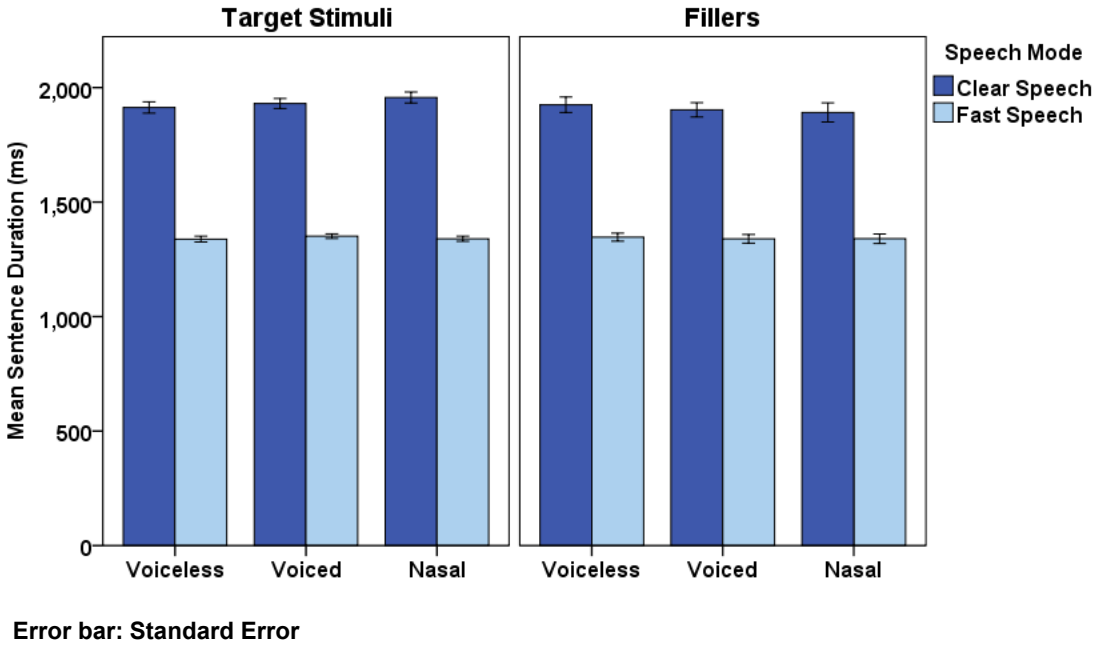


Figure 1: Mean Sentence Duration of Stimuli

2.3.2. Formant transitions of preceding vowel.

The formant transitions of preceding vowels were analyzed by measuring the first three formants of the vowel at the midpoint and at the offset point. Using MultiSpeech software, the selected area containing approximately four to five phonemes of the critical area (e.g., /sɪtkɔ:/ in *sit cautiously*) of each stimulus was displayed on the screen in temporally aligned waveforms and wideband spectrograms. The analysis size of the spectrogram was set at 125 points (258.40 Hz) with a 0-4,000 Hz display in Hamming window. For the measurements at the midpoint, a 15-ms window was located so that the peak amplitude of the central pitch period of the vowel was centered. For the measurements at the offset, the window was placed at the last 15 ms of the

vowel prior to the onset of closure. The formant frequencies were examined in a display of a narrowband Fast Fourier Transform (FFT) spectrum with the analysis size of 512 points with 0 pre-emphasis superimposed by a linear predictive coding (LPC) spectrum using autocorrelation with a 15-ms frame length and the filter order of 30. Figure 2 illustrates how the formant frequencies at the mid point and at the offset point of the preceding vowel were measured. The numerical values of the three formant frequencies and their bandwidths estimated by the LPC analysis were recorded after making sure that the superimposed LPC envelope coincided with the spectral peaks in the FFT spectrum.

The F1, F2 and F3 frequencies at the mid point and at the offset point were averaged across the stimuli sharing the same target stop consonant and the same preceding vowel within each speech mode[5]. The line graphs sorted by contrast type, vowel type, and speech mode are presented in Appendix D. The lines indicating the F1, F2 and F3 transitions of the three target places (Labial, Alveolar, and Velar) were superimposed in each graph, showing the different patterns of formant transitions between the target places. Consistent patterns were observed within the same vowel types even across different contrast types and across speech modes. The near convergence of the formant frequencies at the mid point was seen for all three formants in almost all graphs, showing relatively little vowel-to-consonant coarticulation at the midpoint of the syllables. The divergence of the formant structures as a function of stop place (differences in slope and direction of transitions from midpoint to stop closure) was most notable in the F2 frequencies and least noticeable in the F1 frequencies. Since the data showed that the formant frequencies at the midpoint within the same vowel type were consistent across target places and

[5] For Voiced contrast type (/b/-/d/-/g/), the minimal triplets of low-back vowels were *cob-cod-cog* and *dub-dud-dug*, containing different vowels (/ɑ/ and /ʌ/) because of the availability of appropriate triplets. They were collapsed together for averaging because the formant transition patterns of the two triplets showed almost identical patterns.

the F2 and F3 transitions are known to cue the place information of stop consonants (e.g., Kent & Read, 2001; Raphael et al., 2011), further inspection focused only on the F2 and F3 frequencies at the offset points.

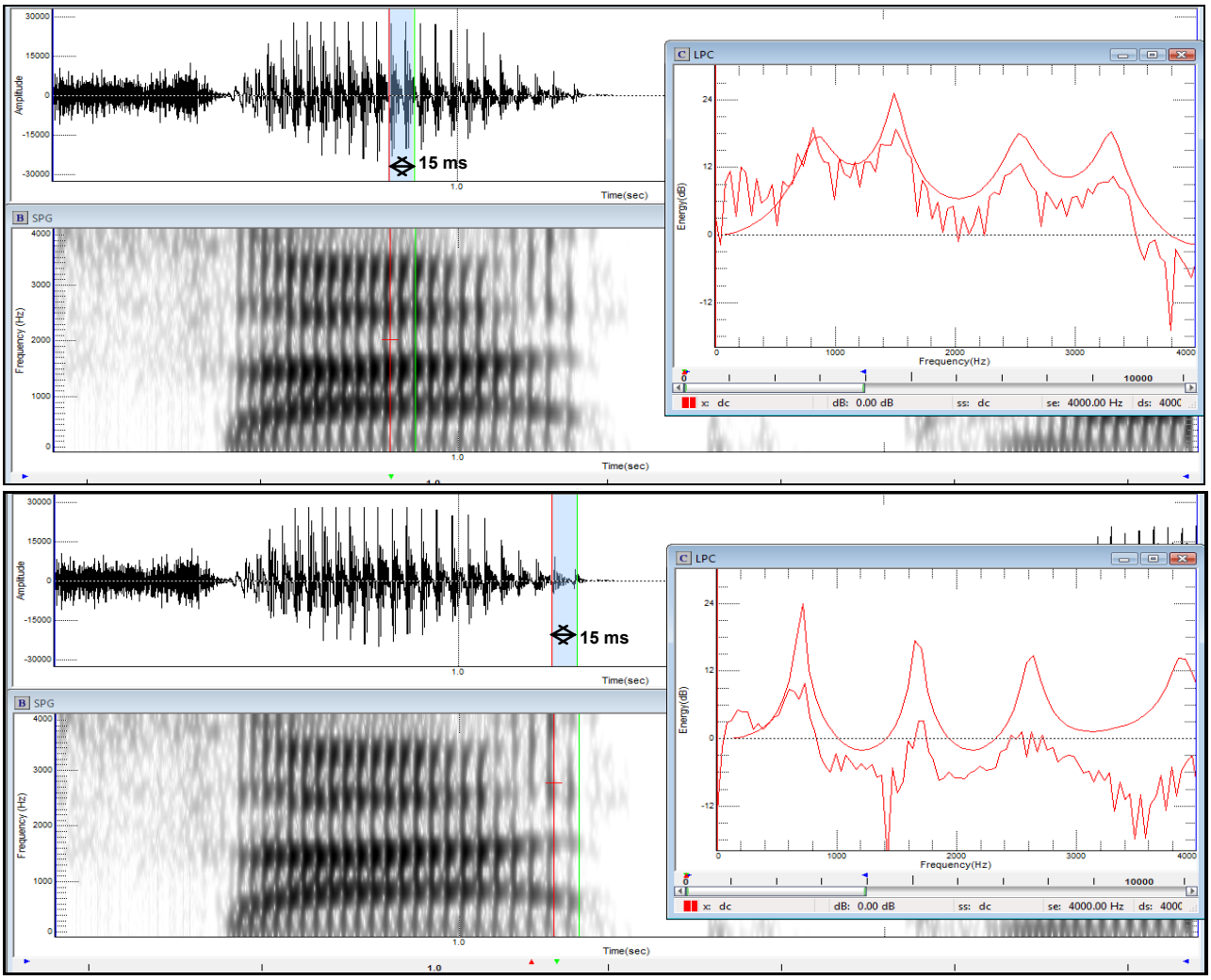


Figure 2. Measurements of formant frequencies at the midpoint (Top panel) and at the offset point (Bottom panel) of preceding vowel using MultiSpeech (*sat cautiously* in Clear speech)

The ranges of the F2 and F3 at the offset point were plotted in Bark for each vowel type sorted by target place and following place, and are presented in Appendix E. For each vowel type (/ɪ/, /æ/, or /ʌ/ & /ɑ/), much smaller variability among following places within the same target

place was seen than the variability across target places. This indicates that target consonant-to-following consonant coarticulatory effects were rather small and that the place cues in the formant transitions remained distinctive across all following contexts (i.e., there was little or no place assimilation). It is also observed that the structural patterns of F2 and F3 as a function of target place were strikingly uniform even across contrast types and across speech modes, with one exception indicating that the F2 values of /æk/ in Clear Speech were relatively low compared to the other velar stops preceded by /æ/. It is not clear what caused this inconsistency because the F2 values of the same /æk/ in Fast Speech were as high as the other velar stops having the preceding /æ/. Future studies should address this issue to clarify the reasons by analyzing larger numbers of tokens.

To further describe the distinctive place information available in F2-F3 transitions into stop closure, the tokens belonging to the same target place sharing the same vowel across the three contrast types were grouped together and were expressed in boxplots in Figure 3. [6] In spite of the fact that the three contrast types were combined together, each formant offset frequency was rather narrowly distributed (F2 differences ranging from 0.77 to 3.38 Barks for Clear Speech, 0.78 to 3.02 Barks for Fast Speech; F3 differences ranging from 0.75 to 1.73 Barks for Clear Speech, 0.44 to 1.56 Barks for Fast Speech), strongly indicating the existence of vowel-context-specific place cues in the F2-F3 structural characteristics.

The observed patterns were seen in both Clear and Fast Speech so clearly that the graphs looked almost identical, suggesting that the acoustic information of F2 and F3 formant transitions was present without diminution or truncation even in casually spoken fast speech. The

[6] The groups sharing low-back vowels /ʌ/ and /ɑ/ were combined together because they showed almost identical F2 and F3 patterns across target place (see Appendix E).

relatively wide variability of F2 frequencies seen in Velar preceded by /æ/ in Clear Speech was caused by the lower F2 of /æk/ mentioned above.

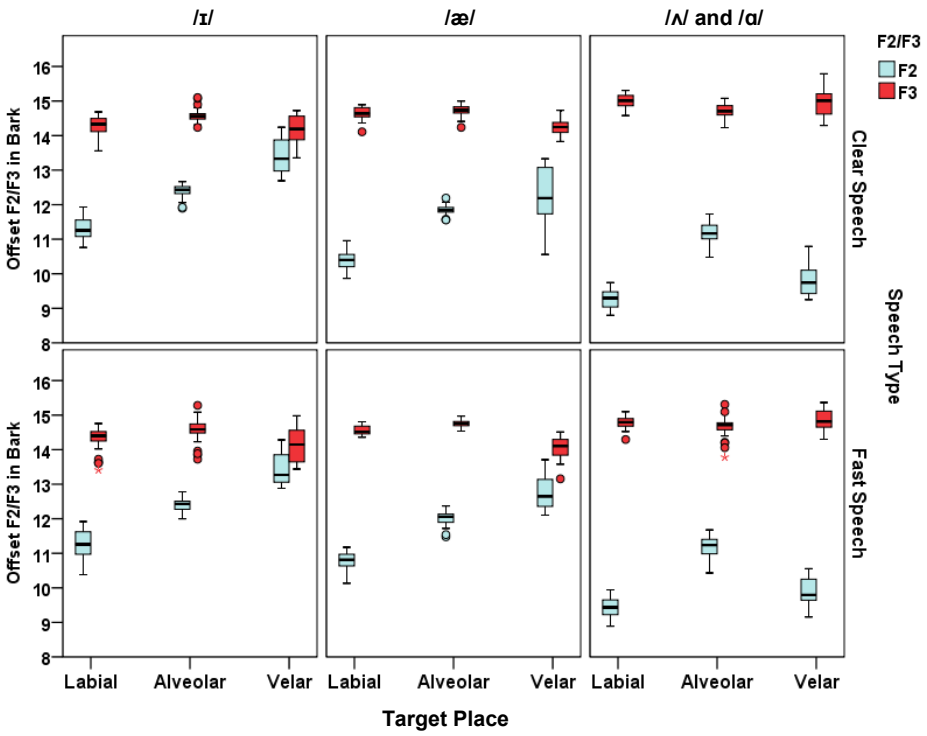


Figure 3: Offset F2 & F3 of Preceding Vowels in Bark Collapsed across Contrast Type

The observed F2-F3 structural patterns for each target place, differing by the vowel types, are described below.

The stops following front vowels /ɪ/ and /æ/ were characterized by the lowest F2 frequencies with wide F2-F3 spacings for Labials (word-final /p/, /b/, and /m/), the highest F2 frequencies with merging F2-F3 structures for Velars (word-final /k/, /g/, and /ŋ/), and intermediate F2 frequencies with intermediate F2-F3 spacings for Alveolars (word-final /t/, /d/, /n/). Little or no overlap in F2 offset values for Labials and Alveolars was observed. Velars were differentiated from Alveolars by sharply converging F2-F3 patterns, which was more

pronounced in the stops following /ɪ/ than /æ/, seemingly caused by the intrinsically higher F2 of /ɪ/ than /æ/.

For stops following back vowels /ʌ/ or /ɑ/, the F2 and F3 characteristics of Labials and Alveolars were similar to those following /ɪ/ and /æ/ with somewhat lower F2 offset frequencies, which also seemed to be caused by the intrinsically lower F2 midpoint frequencies of /ʌ/ and /ɑ/. The most noteworthy point of the /ʌ/ or /ɑ/ group was that the F2-F3 structures for Velars were distinctly different from those seen in the /ɪ/ and /æ/ groups. The F2 frequencies were notably lower to the extent that they were almost as low as those of Labial, showing wide F2-F3 spacings that resulted in a similar F2-F3 structural pattern to that seen in Labial.

In summary, the vowel-context-specific place cues in the F2-F3 formant transitions were found to be present across contrast types as well as across speech modes in the stimuli used in the current study. These cues are most likely to be utilized by native listeners, especially in the cases where other place cues, such as the acoustic information in stop releases, are not present.

2.3.3. Presence or absence of oral stop release.

The presence or absence of the oral stop release was examined by displaying the selected area on the screen in temporally aligned waveforms and wideband spectrograms using MultiSpeech. The settings of the spectrogram were the same as the formant transition measurements except that the frequency range was set from 0 to 8,000 Hz. The following criteria were adopted for the presence of the release[7]: 1) a trace of sound energy is visible in both

[7] There was one token, *rat cautiously*, in Fast Speech in which no stop release was audible but acoustic signal meeting the above criteria was present in both waveform and spectrogram with a relatively long duration (8 ms) and high amplitude (38 dB SPL) for a Fast Speech token. Although there was no audible release noise, it was counted as a release since it met the set criteria.

waveform and spectrogram; and 2) a vertical line indicating sound energy across at least 80% of the entire frequency range in spectrogram is observable.

Table 2 summarizes the results of the measurements. The data for the stimuli having the phoneme sequence of the same place of articulation at the word boundary (the stimuli in *Same contexts*, hereafter; e.g., *sip positively*, *cod tauntingly*, and *king cautiously*) are expressed in bold face. Clear patterns were seen in the data. In Clear Speech, the word-final stops of the target words (*target stops*, hereafter) were not released in most cases of the stimuli in Same contexts whereas the target stops were almost always released in the stimuli in which the word-initial stop following the target stop did not share the same place of articulation (the stimuli in *Different contexts*, hereafter). These patterns were seen in both Voiceless and Voiced contrasts, and were most clearly seen in Labial in which the stops were never released in Same contexts and were always released in Different contexts. In Fast Speech, the difference in release pattern between Same and Different contexts was less marked because of the considerable decrease in number of the stop release observed in Different contexts.

The place order effect on the presence or absence of the releases was also examined by sorting Voiceless and Voiced stimuli in Different contexts into front-to-back sequences (/p-t/, /p-k/, /t-k/, /b-t/, /b-k/, and /d-k/ sequences) and back-to-front sequences (/k-p/, /k-t/, /t-p/, /g-p/, /g-t/, and /d-p/ sequences). A remarkable pattern observed in the Clear Speech condition was that the release rate was 100 % (72/72 tokens) for the front-to-back sequences although the rate for the back-to-front sequences was quite high as well (90.3%; 65/72 tokens). In the Fast Speech condition, the place order effect was notably evident. Whereas only 18.1% (13/72 tokens) was released in the back-to-front sequences as high as 72.2% (52/72 tokens) of the tokens in the

front-to-back sequences was released, strongly indicating that presence or absence of oral stop releases was largely dependent on the front-to-back/back-to-front place order.

Table 2: Presence of Stop Release in Voiceless and Voiced Stimuli

Target Place	Following Place	Clear Speech		Fast Speech	
		Voiceless	Voiced	Voiceless	Voiced
Labial	/p/ (N=12)	**0%**	**0%**	**0%**	**0%**
	/t/ (N=12)	100%	100%	66.7%	66.7%
	/k/ (N=12)	100%	100%	66.7%	92.7%
	Overall (N=36)	66.7%	66.7%	44.4%	52.8%
Alveolar	/p/ (N=12)	83.3%	75.0%	0%	8.3%
	/t/ (N=12)	**0%**	**8.3%**	**16.7%**	**8.3%**
	/k/ (N=12)	100%	100%	75.0%	66.7%
	Overall (N=36)	61.1%	61.1%	30.6%	27.8%
Velar	/p/ (N=12)	100%	100%	41.7%	16.7%
	/t/ (N=12)	100%	83.3%	8.3%	33.3%
	/k/ (N=12)	**25.0%**	**16.7%**	**0%**	**8.3%**
	Overall (N=36)	75.0%	66.7%	16.7%	19.4%

2.3.4. Magnitudes of oral stop releases: duration and amplitude.

The duration and amplitude of release bursts were measured to examine the magnitude of the stop release, using MultiSpeech for the durational measurement and Sound Forge 10.0 Pro for the amplitude measurement. The display settings of MultiSpeech were the same as those used in 2.3.3. The amplitude was measured by highlighting the waveform of the release burst. The obtained RMS dB V values were converted to dB SPL based on the fact that the amplitude-normalized stimuli were presented at 70 dB SPL (see the description of stimulus presentation in Procedure of Main experiment) to listeners in the experiment. Figure 4 illustrates the

distributions of the stop release magnitudes calculated by multiplying the amplitude (RMS dB V converted into dB SPL) of each release by its duration (ms), sorted by following place in each target place, contrast type, and speech mode. The numerical data of the release measurements sorted by following context are presented in Appendix F along with other durational measurements of the critical area. The observed patterns here are in line with those seen in the presence or absence of the stop releases above. The magnitudes of the stop releases were close to zero in Same contexts for both Clear and Fast Speech, as well as in Different contexts for Fast Speech.

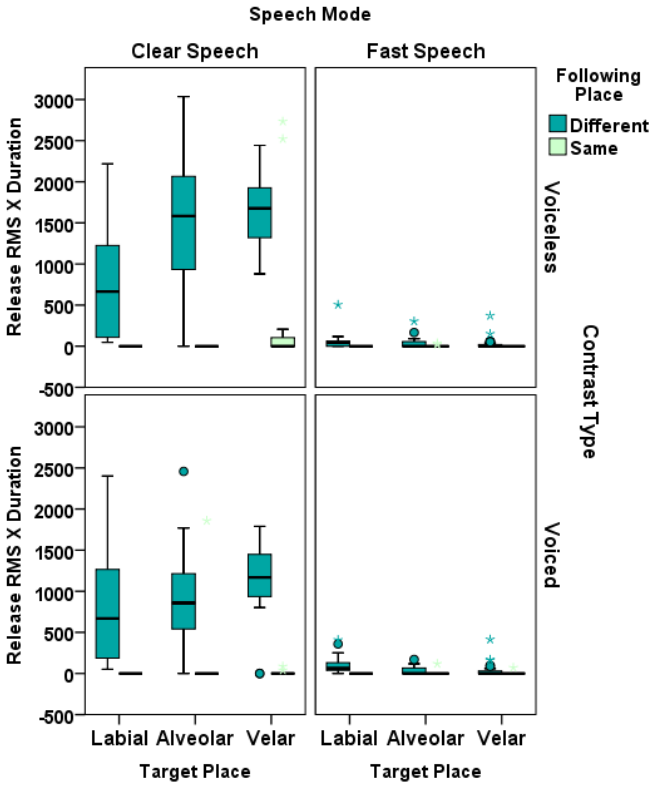


Figure 4: Magnitude of Oral Stop Release (RMS Amplitude in dB SPL × Duration in ms) Compared by Following Context

Combined with the results in the presence or absence of the stop releases, the data strongly indicate that the target stops of Voiceless and Voiced contrasts in Same contexts in both

Clear and Fast Speech were unreleased in most cases, and that even when released, the releases were extremely reduced in terms of both duration and amplitude. The results also indicate that the magnitudes of the stop releases in Different contexts were crucially reduced in Fast Speech.

2.3.5. Stop burst frequencies in oral stops.

It has been reported that the gross spectral shape of the stop burst provides invariant information for place of articulation of stop consonants (e.g., Stevens & Blumstein, 1978; Blumstein & Stevens, 1979). In the present study, the spectral shapes of the word-final oral stop bursts (/p/, /t/, /k/, /b/, /d/, and /g/) analyzed using Linear Predictive Coding (LPC) were examined by adopting the template matching criteria for each place of articulation used by Blumstein and Stevens (1979). The analysis was done only for the tokens in Clear Speech because the majority of the oral stops in Fast Speech did not contain bursts, and even if they did, the bursts had negligible energy in almost all cases. Approximating the settings used by Blumstein and Stevens (1979), the following settings for LPC analysis were adopted: 25-ms frame length; filter order of 24; pre-emphasis (0.9); with frequency range from 0 to 5,000 Hz.

Following the analysis by Blumstein and Stevens (1979), each token was judged in terms of whether it was correctly accepted by the template of the appropriate place and whether it was correctly rejected by the templates of the other two places. The labial template was called *Diffuse-falling* which was characterized by a diffuse spread of energy with either a predominance of lower-frequency spectral peaks over high-frequency peaks or an equal distribution of energy among various peaks. The alveolar template was called *Diffuse-rising* which was characterized by a diffuse spread of energy with high-frequency peaks having greater amplitude than the lower-frequency peaks. The velar template was called *Compact*, which was characterized by a prominent spectral peak in the mid-frequency range. Examples of typical spectral shapes of stop

bursts of the three places of articulation in LPC spectra taken from the actual tokens of the present study are presented in Figure 5. More detailed descriptions of the templates for the three places of articulation are presented in Appendix G.

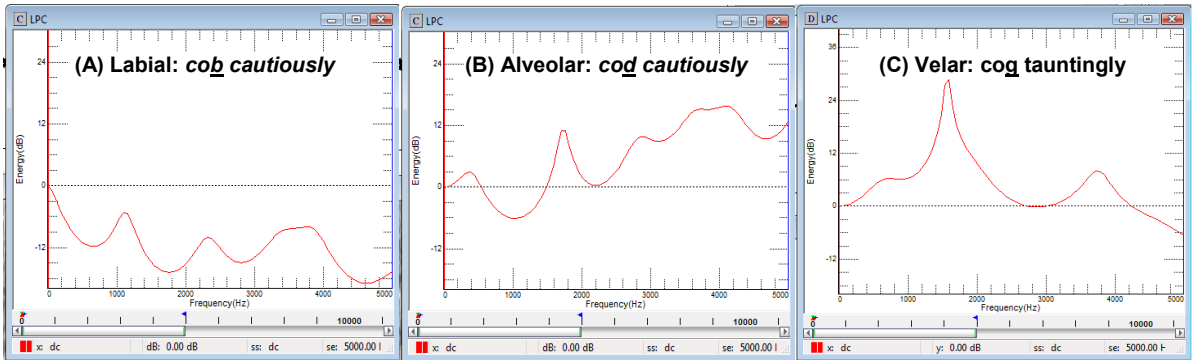


Figure 5: Examples of typical spectral shapes of stop bursts of the three places of articulation in LPC spectra taken from the actual tokens of the present study showing examples of (A) Diffuse-falling (Labial), (B) Diffuse-rising (Alveolar), and (C) Compact (Velar).

Out of 36 tokens in total for each contrast type (6 words × 3 adverbs × 2 repetitions), one third (12 tokens) are in Same context and most contained no stop release or negligible energy in stop bursts, and there were a few tokens with no stop release in Different context as well. Furthermore, the bursts shorter than 3 ms were not examined either. Only the tokens containing 3 ms or longer stop burst were examined, resulting in the analysis of 22 to 27 tokens for each contrast type. Table 3 shows the template matching results.

The results were overall in line with those of Blumstein and Stevens (1979) with a few exceptions. In the Blumstein and Stevens study, the correct acceptance rates of word-final /p/, /t/, k/ and /b/, /d/, /g/ ranged from 70% to 81% and the correct rejection rates from 78% to 91%, observing no conspicuous differences among the three places of articulation in either voiceless or voiced stops. In the present study, the correct acceptance rates for those stops were higher than

75%, replicating the Blumstein and Stevens study, except for /d/ and /g/. The correct acceptance rate was particularly low for /d/ (32%), in which many tokens did not show the rising spectral shape required for matching the alveolar template. The correct acceptance rate was also relatively low for /g/ (68%), which was attributable to the tokens with no prominent peak between 1200 and 3500 Hz and to those with an additional high peak at 3800 Hz or higher aside from the prominent peak. Regarding the correct rejection rates, however, the present study revealed quite high rates for all stops (range: 82% to 100%), replicating the results of the Blumstein and Stevens study (1979).

Table 3: Template Matching Results for Voiceless and Voiced Oral Stop Stimuli

Voiceless/ Voiced	Contrast Type	Correct Acceptance	Correct Rejection Labial	Correct Rejection Alveolar	Correct Rejection Velar
Voiceless	/p/ (N=23)	87.0%	---	95.7%	95.7%
	/t/ (N=22)	90.9%	100%	---	100%
	/k/ (N=27)	77.8%	100%	100%	---
Voiced	/b/ (N=23)	91.3%	---	100%	95.7%
	/d/ (N=22)	31.8%	86.4%	---	90.9%
	/g/ (N=22)	68.2%	81.8%	100%	---

In sum, the template matching analysis of the spectral shapes of stop bursts revealed that the invariant acoustic information signaling place of articulation of oral stops was indeed present in the word-final oral stop bursts in the Clear Speech stimuli of the present study in most cases, except for /d/.

2.3.6. formant frequencies of preceding /æ/ in Nasal contrast.

It has been reported that the vowel /æ/ may be raised or "tensed"before the velar nasal /ŋ/ in American English (e.g., Labov, 2007). The speaker of the present study also indicated a strong possibility of tongue raising and fronting of the /æ/ before /ŋ/ in his production. Since it may cue the velar quality of the following nasal, the F1 frequency, lowering of which would indicate the tongue raising, and the F2 frequency, heightening of which would indicate the tongue fronting, of the vowel /æ/ preceding nasals at the midpoint were examined. Mean F1 frequencies for the /æ/ preceding /m/, /n/, and /ŋ/ (12 tokens for each) with the standard deviations in parentheses were 700.3Hz (77.0), 737.5 Hz (55.3), 682.8 Hz (51.9) in Clear Speech and 724.8 Hz (72.2), 730.7 Hz (72.9), 628.8 Hz (38.9) in Fast Speech, respectively. Likewise, the mean F2 frequencies and their standard deviations for the /æ/ preceding /m/, /n/, and /ŋ/ were 1,648.2 Hz (38.2), 1,669.9 Hz (37.8), 1,737.9 Hz (61.2) in Clear Speech and 1,594.3 Hz (50.1), 1,610.8 Hz (67.1), and 1,721.8 Hz (96.3) in Fast Speech, respectively. The patterns showing the lowest F1 means and the highest F2 means for the /æ/ preceding /ŋ/ in both speech modes were in line with the assumption of the tongue raising and fronting of /æ/ before a velar nasal. These patterns were better observed in Fast Speech.

2.3.7. Acoustic differences in stop occlusion between voiceless, voiced and nasal stops.

It should be noted that acoustic properties in the occlusion of target stop characterizing each contrast type (i.e., Voiceless, Voiced, or Nasal) were observed in the acoustic analysis of each stimulus. The Voiceless stimuli typically showed a silence during the occlusion with a very short and weak voicing energy indicating the residue of the preceding vowel at the onset of the stop occlusion. The occlusion of the Voiced stimuli was characterized by the presence of a voice

bar throughout the occlusion. In the Nasal stimuli, a solid nasal murmur was observed during the occlusion, which had decreased energy from the preceding vowel but still had much greater energy than the occlusions of voiceless and voiced stops. These characteristics were very consistent in both speech modes, although the durations of the occlusions were shorter for the Fast Speech stimuli in accordance with the durational difference of the stimulus sentences between Clear Speech and Fast Speech. Since nasal murmurs in the occlusion portion contain place information (Kent & Read, 2001; Raphael et al., 2011), it is reasonable to consider that the nasal stimuli had place information during the closure whereas the oral stop stimuli (both voiced and voiceless) had minimal place information during the occlusion.

2.4. Procedures

Participants were tested individually in a sound-attenuated room in the Speech-Language-Hearing Sciences department at the Graduate Center CUNY, using computer software, Paradigm (written by Bruno Tagliaferri) in which participants proceeded by using a mouse. The experiment consisted of the main experiment and a word familiarity rating task, both of which were self-paced, followed by a standardized English proficiency test called Versant ™ English Test (for Japanese participants only) administered by telephone. The entire session was completed within two hours for Japanese participants and within one and half hours for AE participants.

2.4.1. Main experiment.

Two experimental conditions, Clear Speech and Fast Speech, each of which contained two physically different tokens of the stimulus sentences in the same speech mode, were constructed. The stimuli were presented binaurally through headphones (Telephonics TDH 39) in random order. The sound level was set in such a way that a 1000-Hz pure tone with the same

amplitude as the RMS value of the stimuli was presented at 70 dB SPL. For each trial, a stimulus sentence with one member of a triplet as the target word (e.g., "He said the word *sit* positively") was presented, and the three words of the triplet (e.g., *sip, sit* and *sick*) appeared on the computer screen at the offset of the auditory stimulus.

Participants were informed by written instructions on the screen that they would hear English two-word sequences embedded in the sentence "He said the word ____ ____," in which the first word is the target word followed by either *positively*, *tauntingly*, or *cautiously*. They were asked to identify the target word by clicking on one of the three written alternatives appearing on the computer screen. The participants were specifically instructed that their choice should be based only on what they heard, not on the familiarity of the words. Each condition consisted of 12 blocks, each of which contained 39 trials, and one repetition of each sentence was presented in the first 6 blocks and the other repetition in the remaining 6 blocks. In order to avoid fatigue effects, a 5-minute break was inserted after the third, the sixth, and the ninth block.

Before the actual test, participants were presented 12 sample sentences produced by the same speaker in familiarization trials, where participants got feedback (i.e., correct or incorrect) for each answer they made. Six of the sentences contained members of word-initial minimal triplets and the other six those of word-final minimal triplets. The Clear Speech condition had Clear Speech stimuli for familiarization and the Fast Speech condition had Fast Speech stimuli. None of the stimuli presented in the familiarization trials appeared in the actual trial blocks.

2.4.2. Word familiarity rating task.

The main experiment was followed by a word familiarity rating task that asked participants to rate their familiarity with the target words on a seven-piont Likert scale. Each written target word was presented on the screen without auditory presentation, and participants

were asked to rate their familiarity with the word by clicking on one of seven boxes corresponding to the seven levels of familiarity (1 = *I don't know the word at all*, 7 = *I know the word very well*).

2.4.3. Versant English test.

After the word familiarity test, the Versant ™ English Test was administered only to Japanese participants. It is a 20-minute computerized telephone test that assesses English spoken language skills of non-native speakers. In the test, six tasks (reading sentences, repeating sentences, sentence building, short answer questions, story retelling, and open questions) were administered and four diagnostic subscores (Pronunciation, Fluency, Sentence Mastery, and Vocabulary) in addition to an overall score were given as the results. The scores can range between 20 and 80. Native English speakers typically perform at ceiling on the test (scores of 80). Scores are interpreted in the following way: 79-80: Can understand with ease virtually everything heard or read. Can express him/herself spontaneously, very fluently and precisely even in complex situations; 69-78: Can understand a wide range of demanding, longer texts, and recognize implicit meaning. Can express him/herself fluently and spontaneously; 58-68: Can understand the main ideas of complex text on both concrete and abstract topics. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue; 47-57: Can understand the main points of clear standard input. Can produce simple connected text on familiar topics; 36-46: Can understand sentences and frequently used expressions related to areas of most immediate relevance. Can communicate in simple and routine tasks on familiar and routine matters; 26-35: Can understand and use familiar everyday expressions and very basic phrases. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help (excerpted from the Versant Test website: http://ordinate.com/). It has been

reported that the scores correlate greater than 90% with extended one-on-one interviews by two experts (Bernstein, 2009).

Chapter 3. Results

The results of the experiment are reported in terms of the following: 1) performance on the main experiment by AE and Japanese groups; 2) correlations between the magnitude of oral stop release and performance by Japanese and AE listeners; 3) error analysis for each language group; 4) correlations of Japanese listeners' performance with their English language experience; and 5) effects of word familiarity and word frequency on listeners' performance.

3.1. Performance in Main Experiment by AE and Japanese Groups

This section reports the performance by AE and Japanese listeners for each speech mode in terms of percent correct response accuracy. First, overall performance by the two language groups in the two speech modes was examined to see whether there was a main effect of language group (i.e., language effect), an effect of Clear Speech benefit, and an interaction between the two. Further analyses for language effect and Clear Speech benefit were calculated separately for each stimulus category: contrast type and its subcategories, target place. Comparisons of performance by each language group by stimulus type are discussed, comparing accuracy of contrast type (Voiceless *vs*. Voiced *vs*. Nasal) and target place (Labial *vs*. Alveolar *vs*. Velar), for each speech mode. Finally, the influence of the following-place contexts on AE and Japanese performance is reported.

3.1.1. Overall performance in Clear and Fast Speech conditions.

Comparisons of overall performance by AE and Japanese listeners in the Clear Speech condition and in the Fast Speech condition collapsed across the three contrast types are presented in Figure 6, illustrating the distributions of response accuracy as box plots. The numerical values of overall mean percent correct accuracies with standard deviations (SDs), medians, semi-interquartile ranges on target stimuli and fillers are presented in Appendix H (Overall

performance on final consonants is given in the first column, performance on initial consonants [fillers] in the last column). As expected, performance by both Japanese and AE listeners on the filler items was at ceiling, showing that neither group exhibited difficulty in perceiving stop place of articulation in word-initial position, which contrasts phonologically in both languages. Thus, no further examination of performance on filler items was carried out.

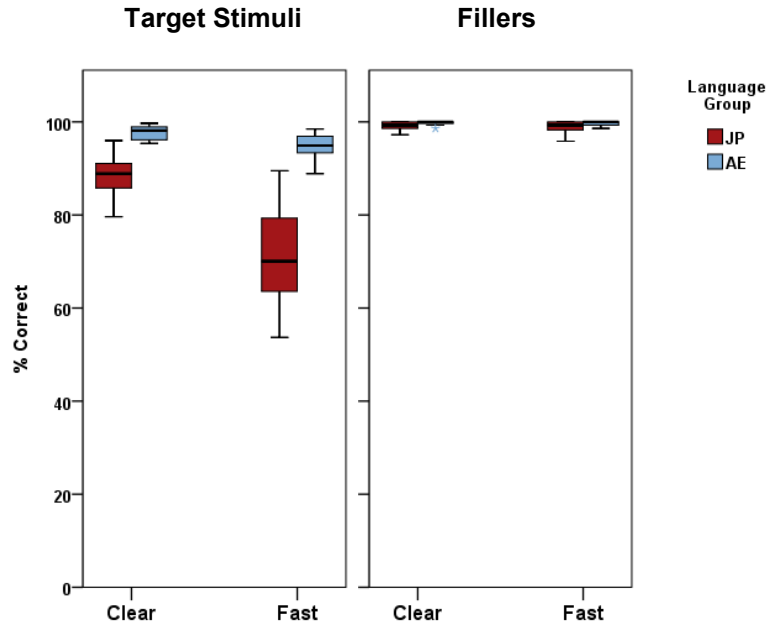


Figure 6: Overall Percent Correct Accuracy by Japanese and AE Listeners

In contrast, performance on the target words, for which the word-final stops contrasted in place of articulation, exhibited larger differences between the two language groups. The box plot in Figure 6 indicates that Japanese listeners' performance was less accurate, compared to AE listeners, even in the Clear Speech condition (median performance 88.9% *vs*. 98.5 % correct, respectively), and the difference between language groups was even greater, with greater variability, in the Fast Speech condition (median performance 70.1% *vs*. 94.9 % correct, respectively). The AE group's performance in both speech modes (98% overall in Clear Speech,

95% in Fast Speech) was slightly lower than those on the fillers (99%), but still above 95% on average with relatively small variability (SDs 2.6% or less) among listeners. The performance difference between the speech modes (i.e., Clear Speech benefit) was larger for the Japanese group than for the AE group (18.8% *vs*. 3.4% benefit), in part due to ceiling effects for the latter group even in the Fast Speech condition.

For statistical analysis of the effect of language group (i.e., AE *vs*. Japanese) on overall performance within each speech mode, non-parametric tests were chosen because of extreme heterogeneity of variance across language groups, ceiling effects for AE listeners, and unequal number of participants within the two language groups in Clear Speech (N = 12 for AE *vs*. N = 24 for Japanese). Two separate Mann Whitney *U* tests (assessed in terms of z scores) comparing the performance by the AE and Japanese groups for each speech mode revealed a highly significant effect of language group on overall performance accuracy ($U = 3$, $z = -4.74$, $p < 0.001$ for Clear Speech, $U = 1$, $z = -5.92$, $p < 0.001$ for Fast Speech). Effect sizes were assessed using the *r* statistic[8]. The language group effect size was very large in Clear Speech ($r = 0.79$), and even larger in Fast Speech ($r = 0.85$), confirming that Japanese had more difficulty in differentiating place of articulation in final oral and nasal stops than native listeners in both speech modes, but especially in Fast speech mode.

To examine the effect of speech mode on performance within language groups, two separate Mann Whitney *U* tests comparing overall performance in Clear *vs*. Fast Speech conditions for each language group were carried out. Results showed significantly better performance in the Clear Speech condition than in the Fast Speech condition (i.e., Clear Speech benefit) for both language groups ($U = 39.5$, $z = -3.51$, $p < 0.001$ for AE, $U = 34$, $z = -5.24$, $p <$

[8] The present study adopted Cohen (1988) criteria of the *r* effect size as follows: 0.1 = small effect, 0.3 = medium effect, 0.5 = large effect.

0.001 for Japanese). While the effect size for AE group was relatively large ($r = 0.59$), that of Japanese was even larger ($r = 0.76$). Combined with the fact that the language effect was larger in Fast Speech than in Clear Speech, the results indicate a larger Clear Speech benefit for the Japanese group than for the AE group.

3.1.2. Language effect (AE *vs.* Japanese) by consonant type within speech mode.

3.1.2.1. Performance by contrast type (Voiceless, Voiced, Nasal).

The performance on each contrast type by AE and Japanese listeners in each speech mode are presented in Figure 7, illustrating the distributions of response accuracy as box plots. A pronounced pattern regarding the difference between the two language groups is the Japanese listeners' considerably poorer performance on Nasal contrasts in both Clear and Fast Speech conditions, making a striking contrast with ceiling performance in both speech modes on Nasal contrasts by the AE group. (See Appendix H for numerical values.)

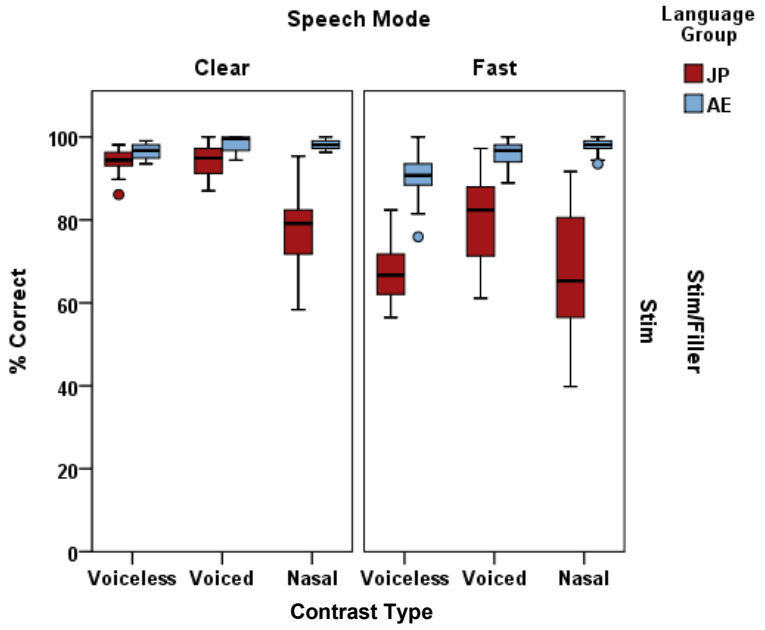


Figure 7: Percent Correct Accuracy on Contrast Type by AE and Japanese Listeners

A set of three Mann Whitney *U* tests for each speech mode was carried out to analyze the differences in performance between the two language groups within each contrast type (Voiceless, Voiced, Nasal), using Bonferroni adjusted alpha levels of .017 per test (.05/3), as presented in Appendix K. While significant language effects were evident for all contrasts in both Clear and Fast Speech conditions, different patterns with respect to the effect sizes were observed between the two speech modes. In the Clear Speech condition, the language effect was conspicuously large for Nasal ($r = 0.81$), compared to Voiceless ($r = 0.45$) and Voiced ($r = 0.53$) contrasts, underscoring Japanese listeners' particular difficulty in identifying the place of articulation of word-final nasals even in Clear Speech. In the Fast Speech condition, on the other hand, the effect sizes were larger for Voiceless ($r = 0.84$) and Voiced ($r = 0.71$) contrasts, but the effect size for Nasal contrasts did not change very much in Fast Speech ($r = 0.86$) relative to Clear Speech. These patterns indicate that correctly perceiving Nasal contrasts is much more challenging for Japanese listeners than for AE listeners regardless of the speech mode, whereas for the perception of Voiceless and Voiced contrasts, Japanese listeners were more disadvantaged than AE listeners by Fast Speech.

3.1.2.2. Performance by target place (Labial, Alveolar, Velar) within contrast type.

Figure 8 further breaks down the data to show performance on Labial, Alveolar, and Velar stops of each contrast type (Voiceless, Voiced, Nasal) for each language group in each speech mode. Another set of three separate Mann Whitney *U* tests for each contrast type for each speech mode was conducted to compare the performance between the two language groups on each target place, using Bonferroni adjusted alpha levels of .0056 per test (.05/9). The detailed statistical data for the language group comparison within each speech mode are presented in Appendix L.

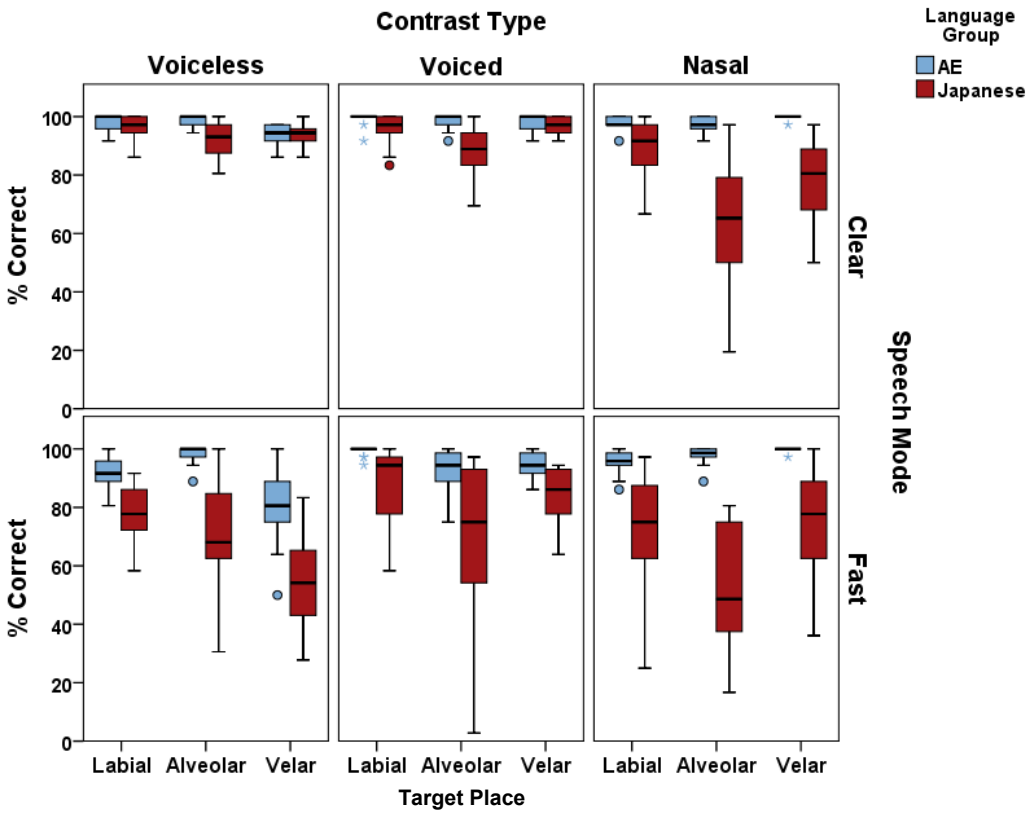


Figure 8: Percent Correct Accuracy on Target Place by AE and Japanese Listeners

In the Clear Speech condition, while the Japanese performance on Voiceless Alveolar /t/ and Voiced Alveolar /d/ was significantly poorer than that of AE listeners ($p < 0.001$ for /t/, $p < 0.0056$ for /d/) with large effect sizes ($r = 0.53$ for /t/, $r = 0.77$ for /d/), their performance on Labial /p/, /b/ and Velar /k/, /g/ did not differ significantly from that of the AE group. The results indicate that the overall language effect observed for Voiceless and Voiced contrasts in the Clear Speech conditions are attributable to the Japanese group's significantly worse performance only on alveolar oral stops. The language effects on all target places of Nasal contrasts were significant, and the effects were very large ($r = 0.76$ or larger) except for Labial /m/ in Clear Speech ($r = 0.49$, which is considered a medium effect size), underscoring Japanese listeners'

conspicuous perceptual difficulty in perceiving the place of articulation of word-final nasal stops followed by another consonant.

In Fast Speech, on the other hand, highly significant and large language effects for all target places were seen (z scores ranging from -4.08 to -6.07, p values all less than 0.001, r scores ranging from 0.59 to 0.88). The effects were especially large on Voiceless and Nasal Contrasts (r = 0.73 or higher for all three places), showing much more evident language effect than seen in the Clear Speech condition.

3.1.3. Clear speech benefit (Clear vs. Fast) by stimulus types within language group.

Since the analyses of this section were done for each language group separately, different statistical tests were carried out for the two language groups. Non-parametric tests (Mann Whitney U tests) were adopted for AE groups because of the heterogeneity of variance, ceiling effects, and unequal sample sizes between the two speech modes, while parametric tests (ANOVAs) were adopted for Japanese groups. For the numerical data, such as mean percent correct accuracies, SDs, medians, and semi-interquartile ranges, see Appendices H (for contrast type) and I (for target place).

3.1.3.1. Performance by AE group.

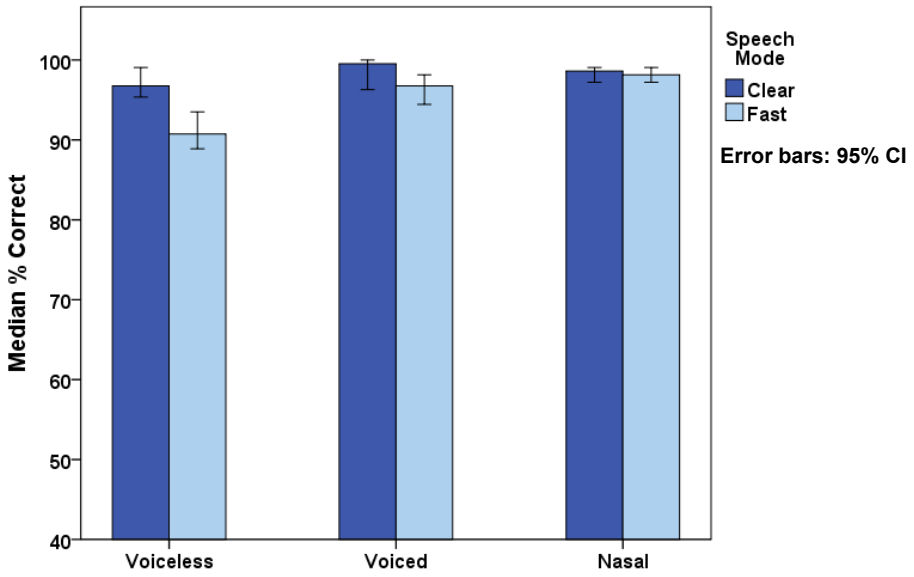
For the comparisons of the AE performance between Clear Speech and Fast Speech on each contrast type (Voiceless, Voiced, Nasal), three separate Mann Whitney U tests were conducted, using Bonferroni adjusted alpha levels of 0.0167 per test (0.05/3). The statistic data are presented in the "Contrast Type" row of Table 4. Figure 9 presents the median response accuracies by AE listeners on contrast types in both speech modes as a bar chart. Results showed that the effect of Clear Speech benefit was significant for Voiceless (U = 26.5, z = -3.95, $p < 0.001$, r = 0.66) and Voiced contrasts (U = 63.5, z = -2.72, $p < 0.0167$, r = 0.45) but not for Nasal contrasts (U = 129,

z = -0.52, p = 0.61, r = 0.09). That is, the Clear Speech benefit reported for the AE's overall performance was due to the better performance on oral consonants by AE listeners in the Clear Speech condition.

Table 4: Performance difference between Clear and Fast speech by AE Listeners Sorted by Contrast Type and Target Place (Mann-Whitney *U* Test)

		U	*z*	Sig.	*r*
Speech Mode	Overall	39.5	z = -3.51	$p < 0.001$**	0.59
Contrast Type	Voiceless	26.5	z = -3.95	$p < 0.001$**	0.66
	Voiced	63.5	z = -2.72	$p = 0.006$*	0.45
	Nasal	129	z = -0.52	$p = 0.61$	0.09
Target Place	Voiceless Labial /p/	47	z = -3.33	$p < 0.001$**	0.55
	Voiceless Alveolar /t/	135	z = -0.35	$p = 0.727$	0.06
	Voiceless Velar /k/	36	z = -3.65	$p < 0.001$**	0.61
	Voiced Labial /b/	137	z = -0.39	$p = 0.696$	0.07
	Voiced Alveolar /d/	70.5	z = -2.56	$p = 0.011$*	0.43
	Voiced Velar /g/	83.5	z = -2.10	$p = 0.036$	0.35

Bonferroni adjustment applied to contrast type comparisons:
$\beta = 0.0167$ for $\alpha = 0.05$; $\beta = 0.0033$ for $\alpha = 0.01$. ** indicates $p < 0.0033$, and * indicates $p < 0.0167$.
Bonferroni adjustment applied to target place comparisons:
$\beta = 0.0083$ for $\alpha = 0.05$; $\beta = 0.0016$ for $\alpha = 0.01$. ** indicates $p < 0.0016$, and * indicates $p < 0.0083$.

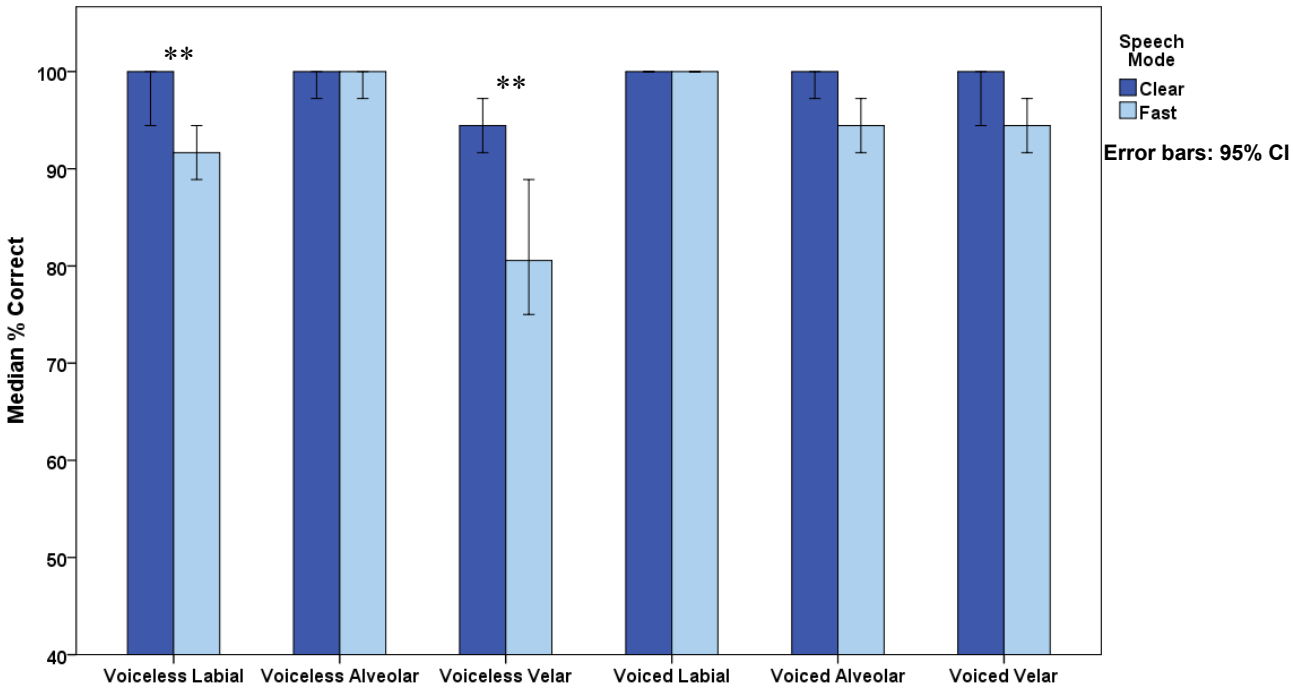


Bonferroni adjustment applied to contrast type comparisons: β = 0.0167 for α = 0.05; β = 0.0033 for α = 0.01. ** indicates $p < 0.0033$, and * indicates $p < 0.0167$.

Figure 9: Percent Accuracy by AE on Contrast Type

For the comparisons of the AE performance between the speech modes for each target place within voiceless and voiced oral stops, Figure 10 shows the AE group's median response accuracies and 95% confidence intervals with speech modes juxtaposed and broken down by target place within each contrast type of oral stops. It is observable from the graphs that the lower overall performance reported above in the Fast Speech condition for the AE listeners was due to lower accuracy levels on particular consonants even in Fast Speech, with others yielding ceiling effects. Since the effect of speech mode was not significant for Nasal contrasts as a whole, no further analysis of these data were conducted. Three separate Mann Whitney *U* tests were carried out for Voiceless and Voiced contrasts (i.e., a total of six tests), using Bonferroni adjusted alpha levels of .0083 per test (0.05/6). Detailed statistic data are presented in the "Target Place" row of Table 4. Results showed that significantly better performance on Clear Speech

than on Fast Speech was seen in Voiceless Labial and Voiceless Velar only. Effect sizes (*r* statistic) of these significant effects were 0.55 and 0.61, respectively.



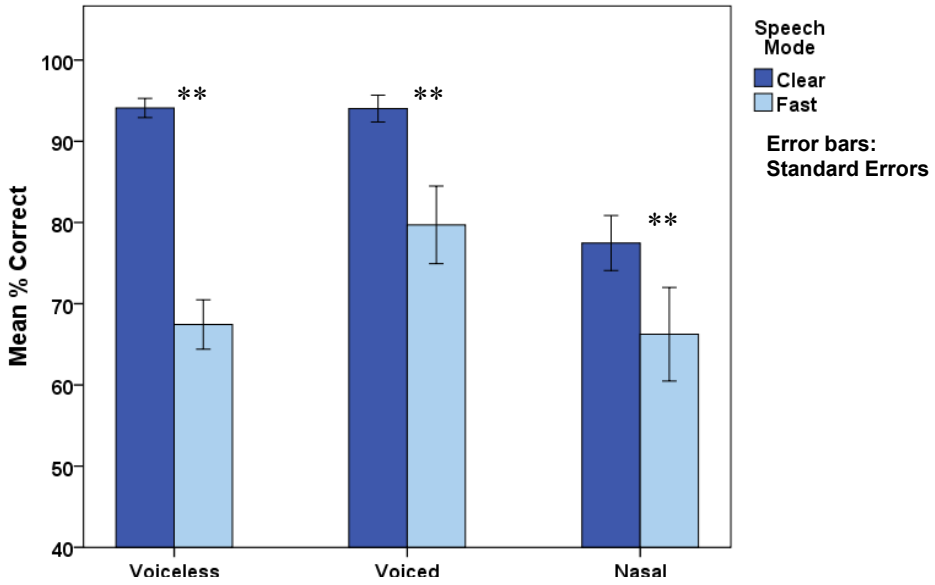
Bonferroni adjustment applied: β = 0.0083 for α = 0.05; β = 0.0016 for α = 0.01.
** indicates $p < 0.0016$, and * indicates $p < 0.0083$.

Figure 10: Percent Accuracy by AE on Target Place

3.1.3.2. Performance by Japanese group.

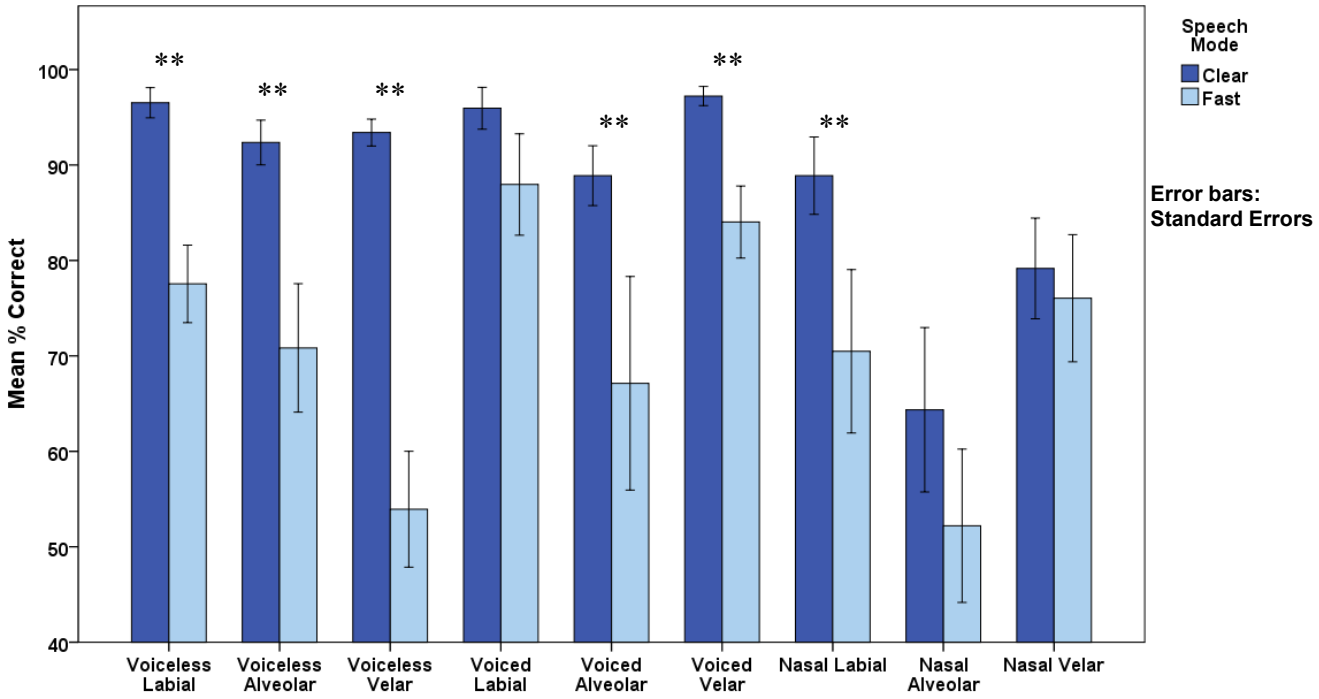
The Japanese group's performance on different contrast types and performance on target places in both speech modes are presented as bar charts in Figures 11 (contrast type) and 12 (target place within contrast type). For the distributions of the performance, see the box plots in Figures 7 (contrast type) and 8 (target place) as well. In order to examine the main effect of speech mode (Clear *vs*. Fast), contrast type (Voiceless *vs.* Voiced *vs.* Nasal) and target place (Labial *vs*. Alveolar *vs*. Velar), and their interactions, a mixed design ANOVA was conducted

with speech mode as a between groups factor, and contrast type and target place as repeated measures factors. The statistical data are presented in Table 5. In the cases where the assumption of sphericity was violated, the Greenhouse-Geisser correction was adopted. There was a significant main effect of speech mode [F (1, 46) = 59.56, $p < 0.001$, $\eta_p^2 = 0.56$], contrast type [F (1.7, 77.4) = 91.84, $p < 0.001$, $\eta_p^2 = 0.67$], and target place [F (1.5, 70.3) = 25.94, $p < 0.001$, $\eta_p^2 = 0.36$]. There was a significant interaction of contrast type × speech mode [F (2, 92) = 26.88, $p < 0.001$, $\eta_p^2 = 0.37$]. There was no interaction of target place × speech mode [F (2, 92) = 0.54, $p = 0.59$, $\eta_p^2 = 0.01$], but there was a significant three-way interaction of contrast type × target place × speech mode [F (4, 184) = 9.94, $p < 0.001$, $\eta_p^2 = 0.18$].



Bonferroni adjustment for contrast type comparisons: β = 0.0167 for α = 0.05; β = 0.0033 for α = 0.01. ** indicates $p < 0.0033$, and * indicates $p < 0.0167$.

Figure 11: Percent Accuracy by Japanese on Contrast Type



Bonferroni adjustment applied: $\beta = 0.0056$ for $\alpha = 0.05$; $\beta = 0.0011$ for $\alpha = 0.01$.
** indicates $p < 0.0011$, and * indicates $p < 0.0056$.

Figure 12: Percent Correct Accuracy by Japanese on Target Place

Table 5: Main Effect of Speech Mode, Contrast Type and Target Place on Japanese Listeners' Performance and Their Interaction (Mixed Design ANOVA)

	df	*F*	Sig.	η_p^2
Speech Mode Main Effect	1, 46	59.555	$p < 0.001$**	0.56
Contrast Type Main Effect	1.7, 77.4	91.84	$p < 0.001$**	0.67
Target Place Main Effect	1.5, 70.3	25.94	$p < 0.001$**	0.36
Contrast Type × Speech Mode Interaction	2, 92	26.88	$p < 0.001$**	0.37
Target Place × Speech Mode Interaction	2, 92	0.540	$p = 0.585$	0.01
Contrast Type × Target Place Interaction	4, 184	18.71	$p < 0.001$**	0.29
Contrast Type × Target Place × Speech Mode Interaction	4, 184	9.942	$p < 0.001$**	0.18

To examine the effect of speech mode on each contrast type, three separate one-way ANOVAs for pair-wise comparisons of contrast type using Bonferroni adjusted alpha levels of .0167 per test (0.05/3) was conducted. The statistic data are presented in Table 6. Results revealed significantly better performance on Clear Speech than on Fast Speech in all three contrasts, but their effect sizes varied considerably, showing a very large effect of Clear Speech benefit for Voiceless contrasts ($\eta_p^2 = 0.85$) and a much smaller effect for Nasal ($\eta_p^2 = 0.20$) with the effect for Voiced contrasts in the middle ($\eta_p^2 = 0.41$).

Table 6: Performance Difference between Clear and Fast Speech by Japanese Listeners Sorted by Contrast Type and Target Place (One-way ANOVAs)

		df	*F*	Sig.	η_p^2
Contrast Type	Voiceless	1, 46	267.13	$p < 0.001$**	0.85
	Voiced	1, 46	32.05	$p < 0.001$**	0.41
	Nasal	1, 46	11.30	$p = 0.002$*	0.20
Target Place	Voiceless Labial /p/	1, 46	75.74	$p < 0.001$**	0.62
	Voiceless Alveolar /t/	1, 46	36.50	$p < 0.001$**	0.44
	Voiceless Velar /k/	1, 46	160.23	$p < 0.001$**	0.78
	Voiced Labial /b/	1, 46	7.72	$p = 0.008$	0.14
	Voiced Alveolar /d/	1, 46	14.01	$p = 0.001$**	0.23
	Voiced Velar /g/	1, 46	45.71	$p < 0.001$**	0.50
	Nasal Labial /m/	1, 46	15.11	$p < 0.001$**	0.25
	Nasal Alveolar /n/	1, 46	4.26	$p = 0.045$	0.09
	Nasal Velar /ŋ/	1, 46	0.54	$p = 0.466$	0.01

Bonferroni adjustment applied to contrast type comparisons:
$\beta = 0.0167$ for $\alpha = 0.05$; $\beta = 0.0033$ for $\alpha = 0.01$. ** indicates $p < 0.0033$, and * indicates $p < 0.0167$.
Bonferroni adjustment applied to target place comparisons:
$\beta = 0.0056$ for $\alpha = 0.05$; $\beta = 0.0011$ for $\alpha = 0.01$. ** indicates $p < 0.0011$, and * indicates $p < 0.0056$.

For further analysis, a set of three separate one-way ANOVAs for pair-wise comparisons of target place for each contrast type (i.e., a total of nine tests) was carried out, again using Bonferroni adjusted alpha levels of .0056 per test (0.05/9). The statistic data are also presented in Table 6. The results showed significant effects of speech mode for all three target places of Voiceless contrasts and of Voiced contrasts except for Voiced Labial /b/. For Voiceless and Voiced contrasts, the effect sizes tended to be large for Velar ($\eta_p^2 = 0.78$ for Voiceless Velar, $\eta_p^2 = 0.50$ for Voiced Velar). For Nasal contrasts, however, the significantly better performance on Clear Speech was only seen in Labial /m/ and not seen in Alveolar /n/ and Velar /ŋ/, showing very different patterns of Clear Speech benefit.

3.1.4. Contrast type comparisons (Voiceless *vs*. Voiced *vs*. Nasal) within language group and speech mode.

This section reports the differences in performance between contrast types within each speech mode and each language group. Again, non-parametric tests were adopted for the AE group and parametric tests for the Japanese group.

3.1.4.1. Performance by AE group.

For the analysis of the AE performance, the following testing procedure was adopted for each speech mode: a Friedman's test was carried out first to see if there is a significant difference in performance across contrast types, then three separate Wilcoxon Signed-Rank tests using Bonferroni adjusted alpha levels of .0167 per test (0.05/3) were conducted when the Friedman's test showed a significant difference. The numerical statistic data are presented in Table 7. For the comparisons of performance on different contrast types, see the bar chart in Figure 9 and the box plots in Figure 7.

Table 7: Performance Difference between Contrast Types by AE Listeners
(Friedman's & Wilcoxon Signed Ranks Tests)

	Friedman's				Wilcoxon Signed Ranks			
		df	χ^2	Sig.		*z*	Sig	*r*
Clear Speech (N = 12)	Contrast Type	2	10.84	$p < 0.01$**	Voiceless *vs*. Voiced (*Md* 96.8%, 99.5%)	-2.73	$p = 0.006$*	0.56
					Voiceless *vs*. Nasal (*Md* 96.8%, 98.6%)	-2.77	$p = 0.006$*	0.57
					Voiced *vs*. Nasal (*Md* 99.5%, 98.6%)	-0.55	$p = 0.58$	0.11
Fast Speech (N = 24)	Contrast Type	2	31.49	$p < 0.001$**	Voiceless *vs*. Voiced (*Md* 90.7%, 96.8%)	-3.99	$p < 0.001$**	0.58
					Voiceless *vs*. Nasal (*Md* 90.7%, 98.1%)	-4.22	$p < 0.001$**	0.61
					Voiced *vs*. Nasal (*Md* 96.8%, 98.1%)	-3.09	$p = 0.002$**	0.45

Bonferroni adjustment applied to Wilcoxon Signed Ranks tests: $\beta = 0.0167$ for $\alpha = 0.05$; $\beta = 0.0033$ for $\alpha = 0.01$. ** indicates $p < 0.0033$, and * indicates $p < 0.0167$.

The results of the Friedman's test for the performance by AE groups in Clear Speech revealed a significant difference by contrast type [χ^2 (2, n = 12) = 10.84, $p < 0.01$]. The Wilcoxon tests revealed that their performance on Voiceless contrasts was significantly worse than that on Voiced contrasts and Nasal contrasts. The difference in performance between Voiced and Nasal contrasts was not significant ($p = 0.58$), showing the following relationship in terms of the goodness of the performance: Voiceless < Voiced ≈ Nasal. The results indicate that the difference seen in the AE group's performance in the Clear Speech condition across contrast types was attributable to the significantly worse performance on Voiceless stops, especially /k/ than those on the other two contrast types.

There was, however, one token out of 108 Voiceless tokens on which 10 out of 12 AE listeners answered incorrectly: one token from *hock cautiously*. After taking the data on this

token out, the difference between Voiceless contrasts and the other two contrasts became non-significant (z = -1.60, p = 0.11, r = 0.33 for Voiceless *vs*. Voiced; z = -1.07, p = 0.29, r = 0.22 for Voiceless *vs*. Nasal). However, it is not clear whether disregarding this particular token is valid because the other token of *hock cautiously* was the second most inaccurate token for AE listeners (6 out of 12 listeners misidentified it) and that the stimuli containing *hock* tended to be error-prone items for both language groups in both speech modes. Further inspection regarding the error-prone k-ending items is carried out and is reported in Error Analysis.

The Friedman's test for the AE group's performance in Fast Speech also revealed a significant difference by contrast type [χ^2 (2, n = 24) = 31.49, $p < 0.001$]. The results of the Wilcoxon tests further revealed that their performance on the three contrast types were significantly different from each other, Voiceless < Voiced < Nasal.

3.1.4.2. Performance by Japanese group.

For comparisons of the Japanese performance, the following testing procedure was adopted for each speech mode: a one-way repeated measures ANOVA was carried out first to assess the main effect of contrast type, then three separate repeated measures ANOVAs for pair-wise comparisons of contrast type using Bonferroni adjusted alpha levels of .0167 per test (0.05/3) were conducted when the main effect was seen. Again, the Greenhouse-Geisser correction was adopted when the sphericity was not assumed. The numerical statistical data are presented in Table 8. For the comparisons of performance between different contrast types as a bar chart, see Figure 11. For the distributions of performance, see the box plots in Figure 7.

Table 8: Performance Difference between Contrast Types by Japanese Listeners (Repeated Measures ANOVAs)

Speech Mode			df	*F*	Sig	η_p^2
Clear Speech	Contrast Type Main Effect		1.3, 30.2	113.25	$p < 0.001$**	0.83
	Pair-wise Comparison	Voiceless vs. Voiced	1, 23	0.01	$p = 0.910$	0.001
		Voiceless vs. Nasal	1, 23	131.58	$p < 0.001$**	0.85
		Nasal vs. Voiced	1, 23	118.80	$p < 0.001$**	0.84
Fast Speech	Contrast Type Main Effect		1.4, 32.6	33.24	$p < 0.001$**	0.59
	Pair-wise Comparison	Voiceless vs. Voiced	1, 23	40.78	$p < 0.001$**	0.64
		Voiceless vs. Nasal	1, 23	0.29	$p = 0.598$	0.01
		Nasal vs. Voiced	1, 23	135.42	$p < 0.001$**	0.86

Bonferroni adjustment applied to pair-wise comparisons: $\beta = 0.0167$ for $\alpha = 0.05$; $\beta = 0.0033$ for $\alpha = 0.01$. ** indicates $p < 0.0033$, and * indicates $p < 0.0167$.

3.1.4.2.1. Japanese performance in Clear Speech.

The results of the repeated measures ANOVA for the Japanese group's performance in Clear Speech showed that there was a significant effect of contrast type [F (1.3, 30.2) = 113.25, $p < 0.001$]. The effect was very large ($\eta_p^2 = 0.83$). The pair-wise comparisons revealed that their performance on Nasal contrasts was significantly poorer than that on both Voiced contrasts and on Voiceless contrasts; the effect sizes for both comparisons were very large ($\eta_p^2 > 0.8$ for both). The difference in their performance between Voiced contrasts and Voiceless contrasts, on the other hand, did not differ significantly ($p = 0.91$), showing the following relationship regarding the goodness of the performance: Nasal < Voiceless ≈ Voiced. The results strongly indicate that the Japanese group's difference in performance across contrast types was due to the significantly

worse performance on Nasal contrasts than those on the other two contrast groups, showing a very different pattern from that seen in the corresponding AE results.

3.1.4.2.2. Japanese performance in Fast Speech.

The results of the repeated measures ANOVA for the Japanese group's performance in Fast Speech revealed that there was a significant effect of contrast type [F (1.4, 32.6) = 33.24, p < 0.001]. The effect was smaller than that seen in Clear Speech but was still very large (η_p^2 = 0.59). The ANOVAs for pair-wise comparisons revealed that their performance on Voiced contrasts was significantly better than that on Voiceless and on Nasal contrasts with very large effect sizes (η_p^2 = 0.64 for Voiced *vs*. Voiceless, η_p^2 = 0.86 for Voiced *vs*. Nasal). On the other hand, their performance between Voiceless and Nasal contrasts did not differ significantly, showing the following relationship regarding the goodness of the performance: Nasal ≈ Voiceless < Voiced. The results indicate that the difference in performance by Japanese listeners across stimulus types in Fast Speech was due to the significantly better performance on Voiced contrasts than those on the other two contrast types, again showing a distinct pattern from that seen in the corresponding AE results.

3.1.5. Target place comparisons: performance on Labial vs. Alveolar vs. Velar within contrast type, language group, and speech mode.

The performance differences between target places within each contrast type within each speech mode for each language group are briefly reported here. Detailed description of statistical analysis for this section is presented in Appendix M. Table 9 (right side columns) summarizes comparisons of the identification accuracy for each place of articulation of target consonants.

Table 9: Summary of Performance Differences: Contrast Type and Target Place

Language Group	Speech Mode	Performance Differences between Contrast Types	Contrast Type	Performance Differences between Target Places
AE	Clear (N = 12)	Voiceless < Voiced ≈ Nasal	Voiceless	/p/ ≈ /k/ < /t/ ≈ /p/
			Voiced	N/A
			Nasal	m/ ≈ /n/ < /ŋ/ ≈ /m/
	Fast (N = 24)	Voiceless < Voiced < Nasal	Voiceless	/k/ < /p/ < /t/
			Voiced	/d/ ≈ /g/ < /b/
			Nasal	/m/ ≈ /n/ < /ŋ/
Japanese	Clear (N = 24)	Nasal < Voiceless ≈ Voiced	Voiceless	/t/ ≈ /k/ < /p/
			Voiced	/d/ < /b/ ≈ /g/
			Nasal	/n/ < /ŋ/ < /m/
	Fast (N = 24)	Nasal ≈ Voiceless < Voiced	Voiceless	/k/ < /p/ ≈ /t/
			Voiced	/d/ < /g/ ≈ /b/
			Nasal	/n/ < /m/ ≈ /ŋ/

The observed patterns here are: 1) in Clear Speech, performance differences between target places were non-significant for the AE group except for the better performance on /t/ than on /k/ and on /ŋ/ than on /n/; 2) identification of Labials was relatively easy for the Japanese group in all contrast types in both speech modes; 3) Alveolars tended to be most difficult for Japanese in both speech modes except for in Voiceless contrasts in Fast Speech; 4) Voiceless Velar /k/ was the hardest to identify for both language groups in Fast Speech.

3.1.6. Following place contexts comparisons (Different *vs.* Same contexts) within target place, contrast type, language group, and speech mode.

This section inspected whether the following context, that is, the place of articulation of the following word-initial stops: /p/ in *positively*, /t/ in *tauntingly*, or /k/ in *cautiously* (*following place*, hereafter), influenced listeners' performance. Depending on whether the following place coincides with the target place, the stimuli of each target place were categorized into Different or

Same context for each contrast type. The mean percent correct accuracies, SDs, medians, and semi-interquartile ranges for both language groups are presented in Appendix J.

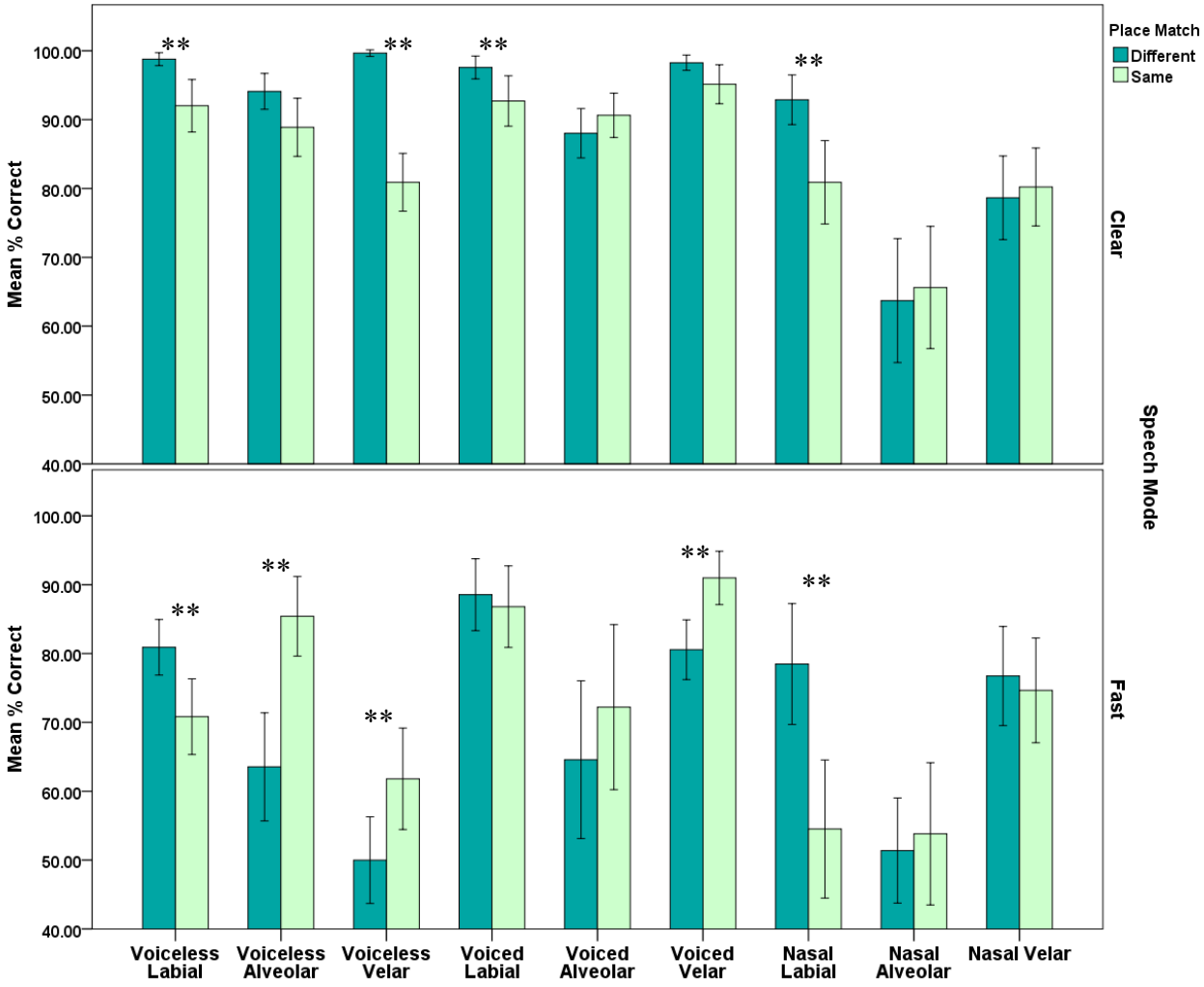
In the AE performance, very few differences with relatively small effect sizes were seen in the direction of better performance on Different than Same context only. The description of statistic analysis on the AE performance is presented in Appendix Q, followed by the figure (Appendix R) and table (Appendix S).

3.1.6.1. Context effects in performance by Japanese group.

The performance by Japanese listeners on each contrast type broken down by target place and following place is presented in Figure 13 as a bar chart with mean performance and standard errors indicated. Clear Speech (top) and Fast Speech (bottom) conditions are shown separately. Repeated measures ANOVAs were conducted for each contrast type (Voiceless, Voiced, Nasal) within each speech mode (i.e., a total of six analyses). In the cases where the assumption of the sphericity was violated, the Greenhouse-Geisser correction was adopted. The statistical data are presented in Appendix T.

In Clear Speech, the main effect of target place (Labial, Alveolar, Velar) was significant for all contrast types [F (2, 46) = 5.70, $p < 0.01$, $\eta_p^2 = 0.20$ for Voiceless; F (2, 46) = 13.53, $p < 0.001$, $\eta_p^2 = 0.37$ for Voiced; F (1.3, 29.5) = 11.82, $p < 0.001$, $\eta_p^2 = 0.34$ for Nasal], and the main effect of following place (Same *vs*. Different) was significant for Voiceless [F (1, 23) = 109.73, $p < 0.001$, $\eta_p^2 = 0.83$] and Voiced contrasts [F (1, 23) = 8.45, $p = 0.008$, $\eta_p^2 = 0.27$] but not for Nasal contrasts [F (1, 23) = 3.82, $p = 0.063$, $\eta_p^2 = 0.14$]. Significant interactions of target place × following place in all contrast types were also found. A conspicuous pattern seen in Clear Speech was that the main effect of following place in Voiceless contrasts was markedly

large ($\eta_p^2 = 0.83$), indicating the strong influence of following place on perception of preceding voiceless stops.



Bonferroni adjustment applied: $\beta = 0.0056$ for $\alpha = 0.05$; $\beta = 0.0011$ for $\alpha = 0.01$.
** indicates $p < 0.0011$, and * indicates $p < 0.0056$.

Figure 13: Percent Correct Accuracy by Japanese on Following Place

For the pair-wise comparisons of Different and Same contexts, a set of three separate repeated measures ANOVAs was carried out for each contrast type within each speech mode (i.e., a total of 18 tests), adopting Bonferroni adjusted alpha levels of 0.0056 per test (0.05/9). Table

10 presents the results of the statistical analysis, with the following places of significantly better performance indicated in bold face. In Clear Speech, significantly better performance on Different context was seen in Voiceless Labial /p/, Voiceless Velar /k/, Voiced Labial /b/, and Nasal Labial /m/, and none of the Same contexts was perceived better than the corresponding Different contexts in any of the contrast types (See top panel of Figure 13). The effect was particularly large for /k/ ($\eta_p^2 = 0.77$), indicating its greatest contribution to the main effect of following place seen in Voiceless contrasts in Clear Speech.

In the Fast Speech condition, the main effect of target place (Labial, Alveolar, Velar) was significant for all contrast types [F (2, 46) = 14.38, $p < 0.001$, $\eta_p^2 = 0.39$ for Voiceless; F (1.38, 31.70) = 8.97, $p < 0.01$, $\eta_p^2 = 0.28$ for Voiced; F (2, 46) = 13.04, $p < 0.001$, $\eta_p^2 = 0.36$ for Nasal]. The main effect of following place was also significant for all contrast types [F (1, 23) = 17.88, $p < 0.001$, $\eta_p^2 = 0.44$ for Voiceless; F (1, 23) = 16.81, $p < 0.01$, $\eta_p^2 = 0.42$ for Voiced; F (1, 23) = 15.78, $p < 0.001$, $\eta_p^2 = 0.41$ for Nasal]. Their interactions were all significant for all three contrast types as well. The interaction of target place × following place was especially large for Voiceless contrast ($\eta_p^2 = 0.68$). The effect of following place was seen in all three places in Voiceless contrasts, but not in the same direction: /p/ showed better performance in Different contexts whereas Same context was better for /t/ and /k/ (See bottom panel of Figute 13). While the directions were not uniform, the effects were relatively large, especially for /t/ ($\eta_p^2 = 0.73$), illustrating the pattern of the interaction between target place and following place seen in Voiceless contrasts, which was notably indicated in the preceding analysis. For Voiced contrasts, the effect was only seen in /g/ with better performance on Same context. For Nasal contrasts, only /m/ showed significantly better performance on Different context.

Table 10: Performance Difference between Following Place by Japanese Listeners in Clear Speech and in Fast Speech (Repeated Measures ANOVAs)

Speech Mode	Contrast Type	Target Place	Following Place	df	*F*	Sig	η_p^2
Clear Speech (N = 24)	Voiceless	Labial	**Different** *vs*. Same	1, 23	13.80	$p = 0.001$**	0.38
		Alveolar	Different *vs*. Same	1, 23	4.89	$p = 0.037$	0.18
		Velar	**Different** *vs*. Same	1, 23	77.63	$p < 0.001$**	0.77
	Voiced	Labial	**Different** *vs*. Same	1, 23	14.09	$p = 0.001$**	0.38
		Alveolar	Different *vs*. Same	1, 23	2.71	$p = 0.113$	0.11
		Velar	Different *vs*. Same	1, 23	3.51	$p = 0.074$	0.13
	Nasal	Labial	**Different** *vs*. Same	1, 23	28.76	$p < 0.001$**	0.56
		Alveolar	Different *vs*. Same	1, 23	0.53	$p = 0.475$	0.02
		Velar	Different *vs*. Same	1, 23	0.29	$p = 0.593$	0.01

Speech Mode	Contrast Type	Target Place	Following Place	df	*F*	Sig	η_p^2
Fast Speech (N = 24)	Voiceless	Labial	**Different** *vs*. Same	1, 23	20.43	$p < 0.001$**	0.47
		Alveolar	Different *vs*. **Same**	1, 23	60.74	$p < 0.001$**	0.73
		Velar	Different *vs*. **Same**	1, 23	16.45	$p < 0.001$**	0.42
	Voiced	Labial	Different *vs*. Same	1, 23	1.66	$p = 0.211$	0.07
		Alveolar	Different *vs*. Same	1, 23	5.22	$p = 0.032$	0.19
		Velar	Different *vs*. **Same**	1, 23	29.24	$p < 0.001$**	0.56
	Nasal	Labial	**Different** *vs*. Same	1, 23	44.03	$p < 0.001$**	0.66
		Alveolar	Different *vs*. Same	1, 23	0.53	$p = 0.474$	0.02
		Velar	Different *vs*. Same	1, 23	0.41	$p = 0.529$	0.02

Contexts with significantly better performance are expressed in bold face.
Bonferroni adjustment applied to pair-wise (Different *vs*. Same) comparisons: $\beta = 0.0056$ for $\alpha = 0.05$; $\beta = 0.0011$ for $\alpha = 0.01$. ** indicates $p < 0.0011$, and * indicates $p < 0.0056$.

To summarize, the patterns of the effect of following place observed in Japanese results are as follows: 1) in Clear Speech, Same context was never perceived better than Different context in either of the contrast types; 2) in both Clear and Fast speech, Labials tended to be identified better in Different context in most cases (5 out of 6 comparisons), and none of the

Labials were perceived better in Same context; 3) /m/ was better perceived in Different context; and 4) performance on /t/, /k/ and /g/ in Fast Speech was better in Same Context, which was the opposite direction to the pattern seen in Labials. Further considerations are given in Discussion.

The results showing the Japanese listeners' better performance on Different context than Same context for Voiceless and Voiced contrasts in Clear Speech suggest that Japanese listeners were utilizing stop release cues for place identification of word-final oral stops. The next analysis investigates the listeners' reliance on the release cues for the perception of word-final oral stops more directly by examining the correlations between the magnitude of stop releases and the listeners' perceptual accuracies.

3.2. Correlations between Magnitude of Oral Stop Release and Performance

As discussed in the introduction, studies have indicated that the presence or absence of release bursts in word-final oral stops is expected to affect their perception and that L2 perception is expected to be more negatively affected by the absence of robust word-final oral stop releases (e.g., Deelman & Connine, 2001; Flege & Wang, 1989). This section examines the correlations of the amplitude and duration of word-final oral stop releases obtained from the acoustic analysis with the performance on each token by Japanese and AE listeners. The following parameters of the oral stop release of each token were correlated with the corresponding percent correct accuracy by Japanese and by AE listeners: RMS amplitude of the release (dB SPL), the peak intensity of the release (dB SPL), and a combined measure (RMS × duration of the bursts [ms]). Spearman rank order correlations were carried out for the performance on Voiceless and Voiced contrasts. The *rho* values examined are presented in Table 11.

Table 11: Correlation between Oral Stop Release and Performance (Spearman's *rho*)

Language Group	Oral Stop Type	Clear Speech			Fast Speech		
		Release RMS	Peak Amplitude	RMS x Duration	Release RMS	Peak Amplitude	RMS x Duration
Japanese	Voiceless	**0.542****	**0.521****	**0.549****	-0.052	-0.042	-0.038
	Voiced	**0.306****	**0.301****	**0.340****	0.023	0.033	0.034
AE	Voiceless	**0.283****	**0.251****	**0.300****	0.073	0.083	0.087
	Voiced	0.113	0.113	0.127	0.112	0.111	0.132

As can be seen, positive correlations of the Japanese performance on both Voiceless and Voiced tokens in Clear Speech with all three variables were evident, indicating Japanese listeners' heavy reliance on the place information in the stop release. The correlations were especially strong for the performance on the Voiceless tokens for Japanese listeners. For AE listeners' performance in the Clear Speech condition, significant, but less strong correlations with the three variables were seen in the performance on Voiceless tokens only. None of the correlations with the three variables reached significance for the performance on Voiced tokens, showing similar, but much weaker patterns of correlations to those of the Japanese performance in Clear Speech. For both language groups, the multiplication of the RMS value by the duration among the three variables exhibited the highest correlations, indicating it is the best predictor of perceptual performance.

Performance in the Fast Speech condition, on the other hand, did not show significant correlations with any of the variables by either of the language groups, suggesting that the acoustic information of stop releases was too limited to utilize effectively for the place perception of word-final oral stops. The scatter plots showing the correlations between

performance (by Japanese and AE) and the combined measure, which had the highest correlations among the three parameters, are presented in Figure 14.

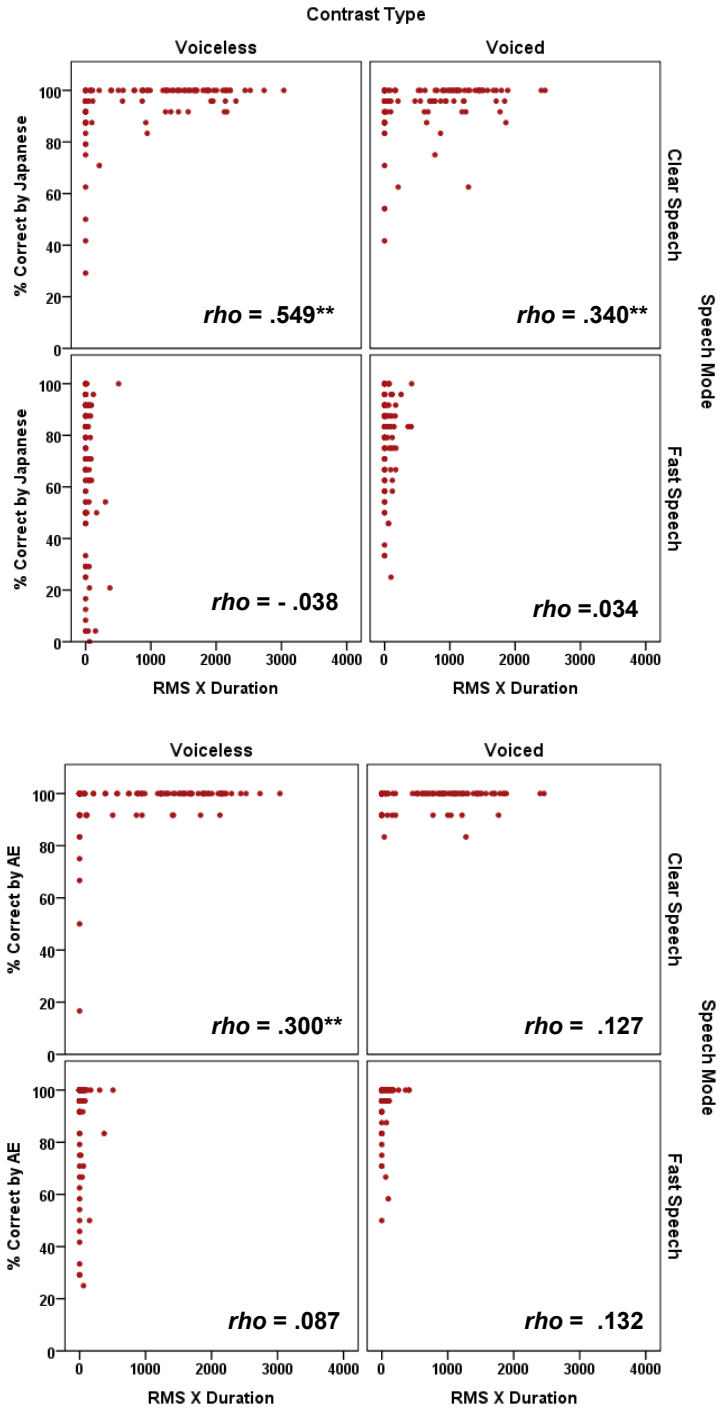


Figure 14: Correlations between Magnitude of Oral Stop Release and Performance (top 4 panels = Japanese Listeners; bottom 4 panels = AE Listeners)

3.3. Error Analysis

This section examines the errors made by AE and Japanese listeners to see if any patterns of difficulties are observed in each language group and/or in each speech mode. In the following subsections, the errors made for each contrast group are sorted by target place first, and are further divided by the following place. Individual error-prone items are also presented for each language group in each speech mode.

3.3.1. Errors by target place.

Confusion matrices summed across responses by the two language groups are presented in Table 12 (AE listeners) and Table 13 (Japanese listeners), each of which collapsed across the target words in the same categories in the Clear and Fast Speech conditions. Binominal Tests were carried out to examine whether or not frequencies of the two places of error responses for each target place (e.g., Alveolar and Velar responses when the correct answer is Labial) were significantly different from each other. The numerical data, including the *z* values of the Binominal Tests, are presented in Appendix U.

Table 12: Confusion Matrix by AE Listeners

		Clear Speech			Fast Speech		
Contrast Type	Contrast Place	Labial Response	Alveolar Response	Velar Response	Labial Response	Alveolar Response	Velar Response
Voiceless	Labial	**98%**	1%	1%	**92%**	8%	0%
	Alveolar	0%	**98%**	1%	1%	**98%**	1%
	Velar	1%	5%	**94%**	4%	15%	**81%**
Voiced	Labial	**99%**	0%	0%	**100%**	0%	0%
	Alveolar	1%	**98%**	0%	5%	**93%**	1%
	Velar	1%	1%	**98%**	1%	4%	**95%**
Nasal	Labial	**98%**	2%	0%	**96%**	4%	0%
	Alveolar	2%	**97%**	1%	1%	**98%**	1%
	Velar	0%	0%	**100%**	0%	0%	**100%**

Table 13: Confusion Matrix by Japanese Listeners

		Clear Speech			Fast Speech		
Contrast Type	Contrast Place	Labial Response	Alveolar Response	Velar Response	Labial Response	Alveolar Response	Velar Response
Voiceless	Labial	**96.5%**	3.0%	0.5%	**77.5%**	18.8%	3.7%
	Alveolar	3.1%	**92.4%**	4.5%	16.6%	**70.8%**	12.6%
	Velar	1.0%	5.6%	**93.4%**	15.3%	30.8%	**53.9%**
Voiced	Labial	**95.9%**	3.0%	1.0%	**88.0%**	7.8%	4.3%
	Alveolar	7.4%	**88.9%**	3.7%	14.1%	**67.1%**	18.8%
	Velar	1.0%	1.7%	**97.2%**	4.2%	11.8%	**84.0%**
Nasal	Labial	**88.9%**	6.1%	5.0%	**70.5%**	15.9%	13.7%
	Alveolar	10.0%	**64.4%**	25.7%	10.9%	**52.2%**	36.9%
	Velar	1.3%	19.6%	**79.2%**	4.6%	19.3%	**76.0%**

For the errors on Voiceless contrasts, it is clearly observed that for Japanese listeners, both Labial and Velar (i.e., /p/ and /k/) are more frequently misheard as Alveolar /t/ than the other place (i.e., /k/ for /p/, and /p/ for /k/) in both Clear and Fast Speech ($p < 0.01$ for all). This tendency is also seen in the AE listeners' error responses in Fast Speech and the errors on /k/ in Clear Speech ($p < 0.01$), suggesting that it is a general perceptual bias, rather than the misperception led by language-specific perceptual patterns. In contrast, the error responses on /t/ were more evenly distributed to /p/ and /k/ by both Japanese and AE listeners in both speech modes, indicating the confusability of /t/ by both language groups.

For the errors on Voiced contrasts, the Binominal Tests further revealed that the response bias toward Alveolar /d/ was seen in the Japanese errors on Labial /b/ and on Velar /g/ in Fast Speech and on /b/ in Clear Speech, as well as in the AE's errors on /g/ in Fast Speech ($p < 0.01$ for all). The error patterns on /d/ were not consistent. It was misperceived as /b/ more frequently than as /g/ in Clear Speech ($p < 0.01$) whereas it was misheard as /g/ more often than as /b/ in Fast Speech ($p < 0.05$) by Japanese listeners. It is interesting that the AE listeners' error

responses on /d/ in Fast Speech showed the opposite direction; it was misperceived as /b/ more often than as /g/ ($p < 0.01$), suggesting different perceptual patterns.

The Japanese listeners' error patterns for Nasal contrasts are very clear and consistent across speech modes. The errors on Labial /m/ are almost evenly distributed to Alveolar /n/ and Velar /ŋ/ whereas strong response biases toward /ŋ/ for the errors on /n/ and toward /n/ for the errors on /ŋ/ were evident, highlighting the confusion between word-final /n/ and /ŋ/ by Japanese listeners. For the AE listeners' errors, a significant response bias toward /n/ was seen in the errors on /m/ only, showing a totally different pattern of error responses from those seen in the Japanese errors.

3.3.2. Errors by following place in each target place.

The numbers of error responses by Japanese and AE listeners sorted by the three following places within each target place, along with the results of Binominal Tests, are presented in Appendix V.

The patterns seen in the errors on Voiceless contrasts are as follows: 1) in Fast Speech, a strong bias of error responses toward Alveolar for Labial and Velar targets was observed in both language groups regardless of the following place; 2) in Clear Speech, the bias toward Alveolar was seen in the Japanese error responses only in geminate contexts, such as the target /p/ followed by /p/ and the target /k/ followed by /k/. This pattern was seen in the AE error responses on the target /k/ followed by /k/ in Clear Speech as well; 3) The Japanese error responses for Alveolar targets showed significant biases toward the same places of articulation as the following places (i.e., the word-final /t/ was more frequently heard as /p/ when followed by /p/ and was heard as /k/ when followed by /k/) in Fast Speech, indicating the influence of the

following place on the perception of target place. This pattern was also seen in the Japanese error responses on Alveolar targets followed by /p/ in Clear Speech.

The error patterns seen in Voiced contrasts were less evident, but were similar to those seen in Voiceless contrasts, especially for the Japanese. The error pattern 1) above, that is, the response bias toward Alveolar regardless of following place, was seen in the Japanese error responses on Velar targets /g/ in Fast Speech. The error pattern 2), the bias toward Alveolar only in Same context in Clear Speech, was seen in the Japanese errors on Labial targets /b/ followed by labial /p/. The error pattern 3), the response bias toward the following place for Alveolar targets, was observed when followed by Labial /p/ (i.e., /d/ was misheard as /b/ when followed by /p/) in the Japanese error responses in both speech modes and in the AE's error responses in Fast Speech. It was also observed in the Japanese errors when followed by Velar /k/ (i.e., /d/ was misheard as /g/ when followed by /k/) in Fast Speech.

The errors made by both language groups on Nasal contrasts showed very different patterns from those seen in Voiceless and Voiced oral consonants. AE listeners' errors were very few in both speech modes and no obvious patterns were seen except for the tendency to be biased toward Alveolar for Labial targets /m/ in Fast Speech, regardless of the following place. The Japanese errors, on the other hand, exhibited conspicuous patterns as follows: 1) the error responses for Velar targets /ŋ/were strongly biased toward Alveolar in both Clear and Fast Speech, regardless of the following place; 2) the error responses for Labial targets /m/ followed by Labial /p/ were biased toward Alveolar in both Clear and Fast Speech; 3) the error responses for Alveolar targets /n/ in Fast speech were strongly biased toward Velar regardless of the following place. The same tendency was seen in Clear Speech except for the Alveolar targets followed by Labial /p/. These patterns strongly indicate the confusion of word-final Alveolar /n/

and Velar /ŋ/ by Japanese listeners. Word-final Labial /m/ seems to be less confusable for Japanese unless it is followed by a labial consonant.

3.3.3. Individual error-prone target words for Japanese and AE listeners.

The words with 20 highest error rates by Japanese listeners for each speech mode are presented along with the error rates on the same items by AE listeners in Table 14. Six physically different tokens (i.e., each item followed by three different adverbs appearing twice in an experiment) for each target word were presented to each subject. Thus, the number of the entire presentations of each item was 144 (6 tokens × 24 subjects) for Japanese listeners in both speech modes and for AE listeners in Fast Speech, and 72 (6 tokens × 12 subjects) for AE listeners in Clear Speech. The error rates were calculated based on these numbers. The entire word lists of error responses by Japanese and AE listeners are presented in Appendix W, and more detailed word lists sorted by following place are presented in Appendix X to Appendix AA.

In Clear Speech, the error-prone words for the two language groups exhibited very different patterns. Whereas the words with high error rates by Japanese listeners were dominated by words with final nasals, especially /n/ and /ŋ/ (e.g., *din*, *ban*, *kin*, *bang*), AE listeners' most error-prone item was the /k/-ending *hock* (23.6% error rate), and the error rates of other words were below 5%. The word *hock* also had a high error rate by Japanese listeners (21.5% error rate). Further examination revealed that the error rates of *hock* were higher for both Japanese and AE listeners when followed by Velar /k/, creating a geminate context (i.e., *ho<u>ck c</u>autiously*). The other words appearing in the table for both language groups, such as *kin* and *rid*, did not show any consistent pattern of context effects across the language groups. As for the Japanese error-prone nasal-ending words, the error rates of the /n/-ending and /ŋ/-ending words appeared in the table were found to be high regardless of following place whereas the error rates of the /m/-

ending words were higher when followed by Labial /p/, indicating the context effect. See Appendices X (Japanese) and Y (AE) for the context effect of each word in Clear Speech.

Table 14: Error-prone Items by Japanese and AE listeners in Clear and Fast Speech

Clear Speech			**Fast Speech**		
	Japanese	AE		Japanese	AE
Target Word	Error Rate (%)	Error Rate (%)	Target Word	Error Rate (%)	Error Rate (%)
din	50.7	4.2	hock	92.4	52.1
ban	45.8	1.4	shock	71.5	45.1
kin	43.1	8.3	kin	54.2	1.4
ran	40.3	1.4	ban	53.5	0
bang	34.7	0.0	lit	52.1	4.9
rid	33.3	5.6	bam	51.4	11.8
sung	31.3	0.0	sun	50.7	4.2
king	23.6	0.0	cog	48.6	22.2
hock	21.5	23.6	ran	47.9	3.5
run	20.8	1.4	din	47.2	2.8
bam	18.8	2.8	sack	46.5	3.5
lit	15.3	2.8	rid	43.1	9.0
ram	13.9	2.8	sit	42.4	0.7
ding	13.9	0.0	rat	38.9	0.0
rum	13.2	0.0	rap	36.8	23.6
sun	13.2	0.0	sap	36.8	14.6
rung	12.5	0.0	ram	35.4	0.0
rat	10.4	0.0	dud	34.0	2.1
tad	10.4	0.0	rum	34.0	3.5
sum	9.7	2.8	run	33.3	1.4

In Fast Speech, on the other hand, the patterns seen in the error-prone words for two language groups were much more similar. It is prominent that the words with the two highest error rates for both language groups were exactly the same; *hock* was the highest (error rates: 92.4% for Japanese; 52.1% for AE) and *shock* was the second (error rates: 71.5% for Japanese; 45.1% for AE). For both language groups, the error rates of these two items were markedly

higher than the rest of the items. The examination of the context effect revealed that the error rates of *hock* by Japanese were extremely high regardless of following place (95.8% to 89.6%) while those by AE showed a higher error rate for the /k/-following geminate context (70.8%), although the other two contexts were also relatively high (45.8% for the /p/-following context and 39.6% for the /t/-following context). The context effects of *shock* for both language groups were less obvious. Although the error rate of *shock* by Japanese listeners for the /k/-following geminate context was relatively low (47.9%), those for the other two contexts (i.e., /p/- and /t/-following context) were exactly the same (83.3%). The error rates by AE listeners also changed very little by following place, ranging from 47.9% to 41.7%. The other words appearing in the table for both language groups, such as *rap*, *cog*, *sap*, *bam*, and *rid*, did not show any consistent pattern across the language groups, regarding the following context effects. See Appendices Z (Japanese) and AA (AE) for the context effect of each word in Fast Speech.

3.4. Correlations of Japanese Listeners' Performance with Language Experience and Proficiency

This section reports the correlations of the Japanese listeners' performance with their language experience variables: LOR; AOA; chronological age; English proficiency as measured by the Versant test; and L2 use as measured by the proportion of L1 and L2 language use in their daily life. Spearman rank order correlations were carried out on each variable. The detailed *rho* values examined are presented in Appendix AE. Since the results showed no correlations with age or with L2 use, only the correlations with LOR, AOA, and language proficiency measured by the scores of the Versant ^{TM} English Test are discussed in the following subsection. It has to be kept in mind in these analyses, however, that the Japanese participants in Clear Speech experiment and those in Fast Speech experiment are two different groups. Thus, although the

distributions of age, LOR, AOA, language proficiency in the two groups are quite similar to each other (see Table 1 for biographical information of the two groups), observed discrepancies in correlational patterns between the two groups may not be totally attributable to the speech mode.

3.4.1. Correlation with LOR.

Figure 15 shows scatter plots of the correlations of the Japanese listeners' overall performance and their performance on each contrast type with their LOR in Clear Speech and in Fast Speech, with *rho* values included. As can be seen, significant positive correlations were evident except for the performance on Voiceless contrasts in Fast Speech, which was one of the most challenging contrast types for both Japanese and AE groups. Another observable point is the relatively weak correlation seen in the performance on Voiced contrasts in Clear Speech, which was the easiest contrasts for both language groups, accounting for 24.1% of variance (*rho* = 0.491, $p < 0.05$). Overall, however, quite strong correlations were seen between Japanese listeners' performance and LOR across all three contrast groups.

The notable effect was the variability in the correlation of LOR with the performance in the Fast Speech condition for oral stops and for the nasals in both Clear and Fast Speech conditions. The scatter plots of all contrast types in Fast Speech and that of Nasal contrasts in Clear Speech show that whereas most Japanese listeners with less than 4 years of LOR did very poorly, performance of more experienced listeners varied from less than 60% accuracy to more than 90% accuracy. These scatter plots show again, that voiceless stops in Fast Speech were particularly difficult even for Japanese who had been in the US for over 10 years.

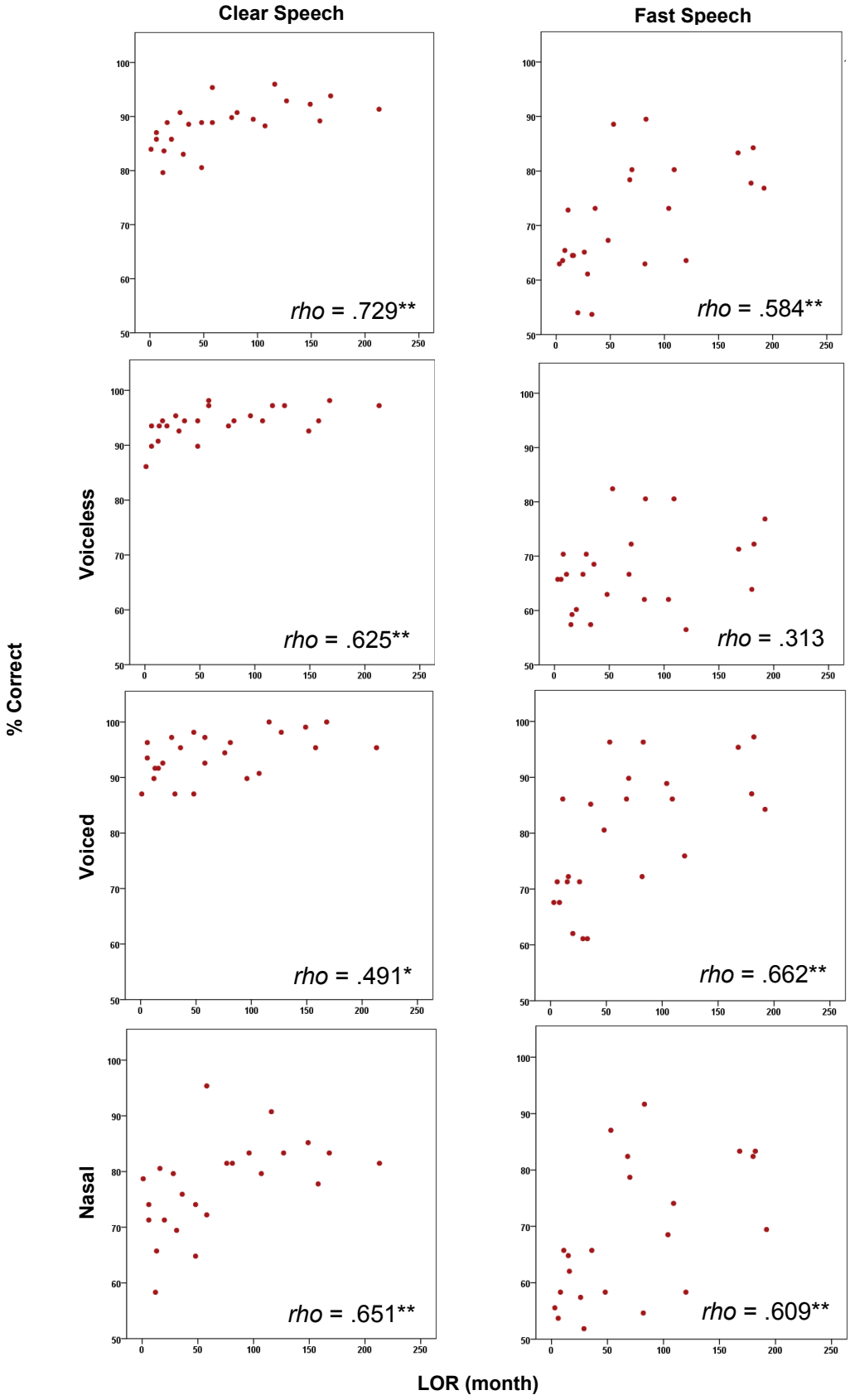


Figure 15: Correlation between Japanese Listeners' Performance and LOR

3.4.2. Correlation with AOA.

The correlations between the Japanese listeners' performance and their AOA are presented in Figure 16. In addition to the fact that the correlations were not as strong as those with LOR, a noticeable pattern seen in the correlation with AOA is that the expected negative correlations were lower in Clear Speech than in Fast Speech in all contrast types. In fact, the significant correlation was only seen with the performance on Voiceless contrasts in Clear Speech whereas the correlations were significant with their performance on all contrast types in Fast Speech. Another interesting result is that the correlation of their performance on Voiceless contrasts in Fast Speech, which did not reach significance with LOR, was significantly correlated with AOA, accounting for 18.1% of variance ($rho = 0.425$, $p < 0.05$).

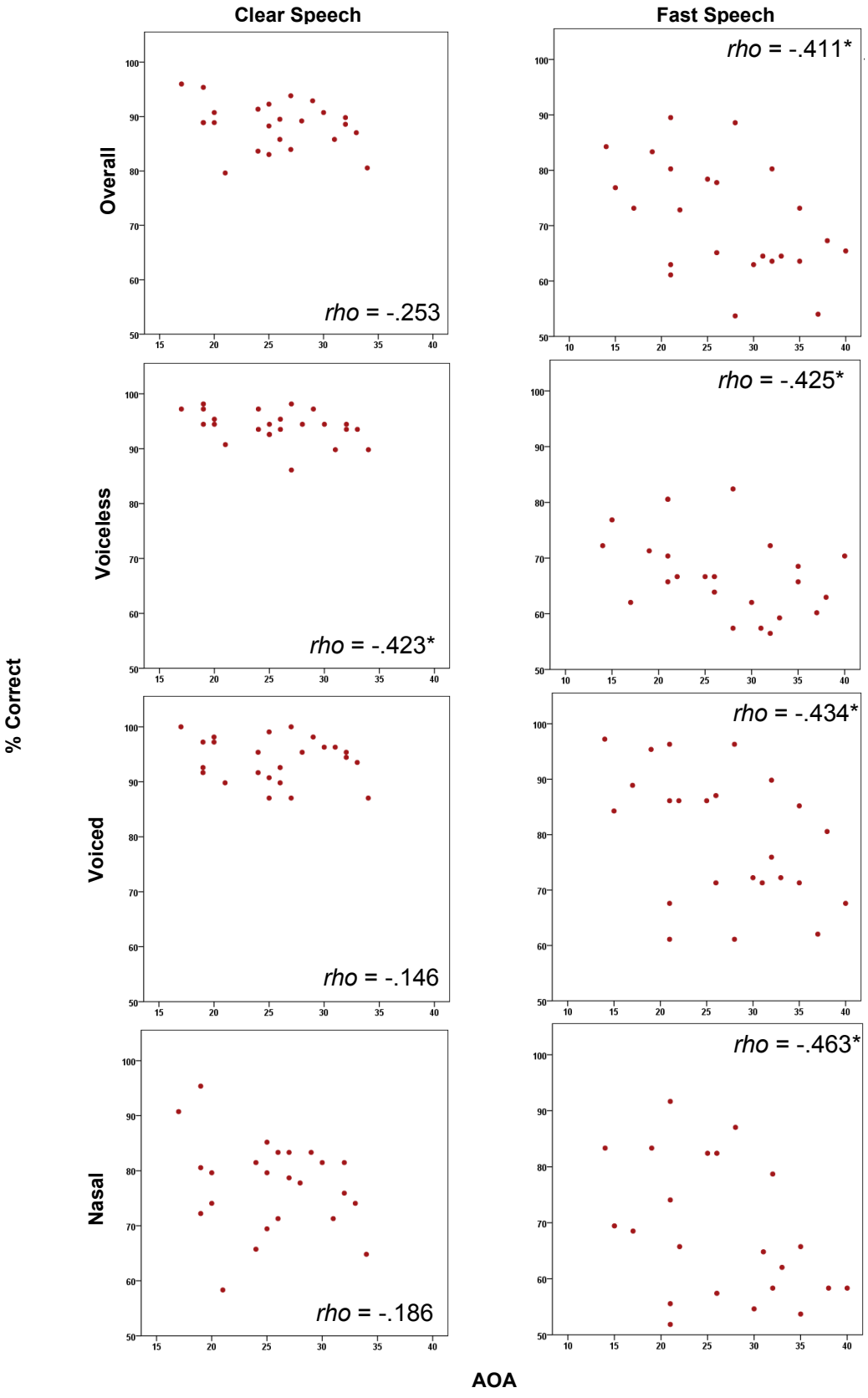


Figure 16: Correlation between Japanese Listeners' Performances and AOA

3.4.3. Correlation with language proficiency (Versant Test scores).

The correlations of the Japanese listeners' performance with the overall scores of Versant ™ English Test (*Overall Versant*, hereafter) are illustrated in Figure 17. The correlations of Overall Versant in Clear Speech were almost as strong as those of LOR, and in Fast Speech, they were even stronger than those of LOR, marking the highest correlation for their performance on Nasal contrasts, accounting for 66.7% of variance (*rho* = 0.817, $p < 0.01$). The correlation of Overall Versant with the Japanese performance on Voiceless contrasts in Fast Speech was also much higher than those of LOR, accounting for 22.1% of variance (*rho* = 0.470, $p < 0.05$). The patterns here seem to indicate that Overall Versant is a better predictor than LOR for the Japanese listeners' performance on relatively difficult word-final stop consonant place contrasts.

For further analyses, the correlations with the subscores of Versant Test were also examined. The scatter plots for each contrast type are available in Appendix AB to Appendix AD. An interesting finding was that the vocabulary score and the pronunciation score showed opposite directions in terms of the patterns of correlation with the perception of oral stops. Moreover, the sentence mastery score showed similar tendencies to those of the vocabulary score, and the fluency score those of the pronunciation score. The vocabulary score was most highly correlated with the performance on Voiceless contrasts in Fast Speech among all variables, accounting for 27.5% of variance (*rho* = 0.524, $p < 0.01$), while its correlations with all performance in Clear Speech were relatively low, compared to the other subscores. The correlation of the vocabulary score with Voiced Contrast in Clear Speech was especially poor, accounting for only 1.7% of variance (*rho* = 0.132, $p > 0.05$). More moderate but similar patterns were seen in the sentence mastery score as well.

In contrast, the highest correlation with the performance on Voiced contrasts in Clear Speech was seen in the pronunciation score that accounts for 28.7% of variance (*rho* = 0.536, $p < 0.01$), but its correlation with Voiceless contrasts in Fast Speech was very low, accounting for only 7.1% of variance (*rho* = 0.266, $p > 0.05$). The fluency score had similar tendencies to the pronunciation score.

It appears that the vocabulary scores and the sentence mastery scores predict Japanese performance well on relatively challenging oral stop contrasts, whereas the pronunciation fluency scores are good for predicting Japanese performance on relatively easy oral stop contrasts. For more detailed correlations of performance by Japanese listeners with their language backgrounds, see Appendix AE.

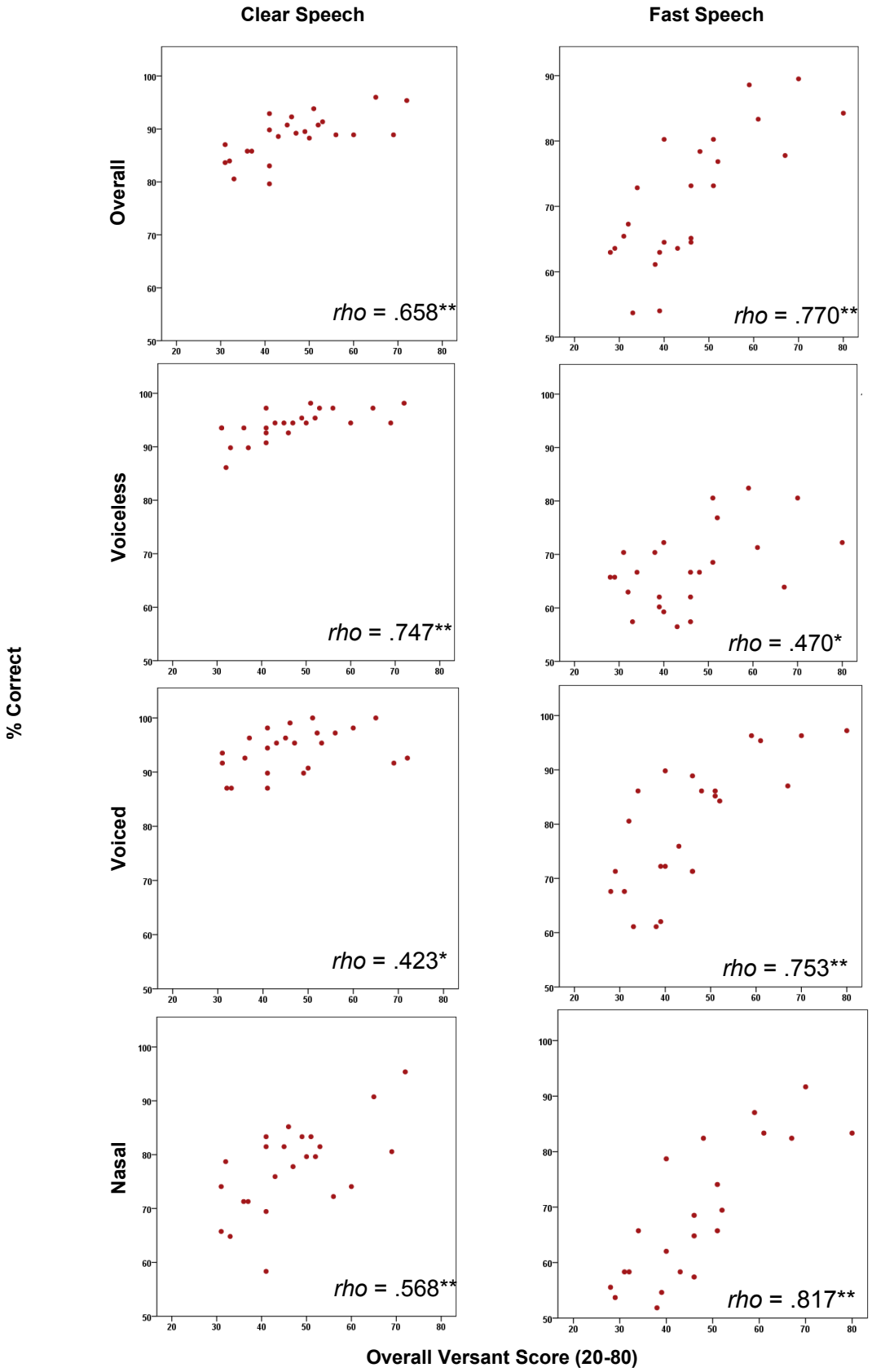


Figure 17: Correlation between Japanese Listeners' Performances and Language Proficiency

3.4.4. Summary of correlations of Japanese performance with subject variables.

To summarize, LOR was positively correlated with the Japanese performance quite strongly in most cases although it was not correlated with the performance on Voiceless contrasts in Fast Speech. English language proficiency measured by the scores of the Versant ™ English Test showed higher correlations than LOR with performance for the Fast Speech group, suggesting that language proficiency is a better predictor than LOR for the Japanese listeners' performance on relatively difficult contrasts in conversational speech environments. Among the subscores of the Versant Test, the vocabulary scores and the pronunciation scores exhibited contrasting patterns of correlations: the vocabulary scores appeared to be a good predictor of performance on challenging oral stop contrasts whereas the pronunciation scores seemed to be good for predicting the performance on easy oral stop contrasts. Negative correlations were seen between perceptual performance and AOA although they were not as high as the correlations seen in LOR and language proficiency. AOA tended to have higher correlations with performance for the Fast Speech group than that for the Clear Speech group. Participants' ages and proportion of L2 use in daily life did not show evident correlations with performance on any of the contrast types.

3.5. Lexical Effects on Listeners' Performance: Word Frequency and Word Familiarity

In order to inspect possible lexical effects on performance, correlations of performance on each target word by each language group with word frequency as measured by the $SUBTL_{WF}$ scores, and word familiarity as measured by the mean scores of the 7-scale familiarity ratings by each language group are reported here.[9] Spearman rank-order correlations revealed that the

[9] For the correlation measurement of the familiarity ratings, mean scores, instead of median scores were adopted because of a lack of variability of ratings by AE listeners, having the maximum 7 points for all target words, except for 6.5 points for the word *din* by the participants

$SUBTL_{WF}$ scores of the target words (*word frequency scores*, hereafter) were significantly correlated with the mean scores of the 7-scale familiarity ratings (*word familiarity scores*, hereafter) by Japanese, accounting for 54.5% of variance for Clear Speech (*rho* = 0.738, $p < 0.001$) and 48.3% of variance for Fast Speech (*rho* = 0.695, $p < 0.001$) and by AE listeners, accounting for 44.2% of variance for Clear Speech (*rho* = 0.665, $p < 0.001$) and 51.7% of variance for Fast Speech (*rho* = 0.719, $p < 0.001$).

The assumption here is that if lexical information of words biased phonetic perception of word-final consonants, lower performance in the experiment would be seen in less familiar words or in less frequent words. Thus, positive correlation of the performance accuracy with the word familiarity score or/and with the word frequency score would indicate the lexical effect on the outcomes. For this reason, one-tailed correlations were adopted for the examination of the lexical effects, disregarding negative correlations. The results of Spearman rank-order correlations showed that the percent correct accuracy on target words and the corresponding word familiarity scores by either of the language groups were not correlated in either of the speech modes, accounting for 0.2% of variance for Japanese in Clear Speech (*rho* = 0.04, $p > 0.05$), 0.8% for Japanese in Fast Speech (*rho* = 0.09, $p > 0.05$), 0.4% for AE in Clear Speech (*rho* = -0.06, $p > 0.05$) and 0.5% for AE in Fast Speech (*rho* = 0.07, $p > 0.05$). None of the correlations between accuracy and word frequency scores were significant either, accounting for 0.01% of variance for Japanese in Clear Speech (*rho* = 0.01, $p > 0.05$), 0.09% for Japanese in Fast Speech (*rho* = 0.03, $p > 0.05$), 2.2% for AE in Clear Speech (*rho* = -0.15, $p > 0.05$) and 0.5% for AE in Fast Speech (*rho* = -0.07, $p > 0.05$), indicating no lexical effects on performance.

---

in the Clear Speech condition. The mean familiarity rating scores by both AE and Japanese groups showed higher correlations with word frequency scores and seemed to reflect the familiarity of each subject group more sensitively.

Chapter 4. Discussion

This section first reviews the results of the study to recapitulate the findings; in the process, it also discusses whether the hypotheses of the present study proposed in the introduction were born out or not. It is followed by the examination of error patterns made by the AE and Japanese groups, and finally, conclusions and future directions are considered. Theoretical implications are deliberated throughout the discussion wherever a relevant topic arises.

4.1. Review of Results and Testing Hypotheses

4.1.1. Language effect: less accurate perception by Japanese than AE listeners.

Not surprisingly, the present study hypothesized that Japanese listeners' place perception would be less accurate than AE listeners for English word-final oral stops and for word-final nasal stops. This language effect was evident in the results, showing significantly less accurate performance by Japanese listeners on all contrast types (Voiceless, Voiced, and Nasal) as well as the overall performance in both Clear and Fast Speech.

In Clear Speech, the effect was especially large for Nasal contrasts compared to Voiceless and Voiced oral stop contrasts. Whereas the AE group's performance was near or at ceiling on all contrast types (median > 96%), the Japanese group's performance on Nasal contrasts was noticeably lower (median = 79%), highlighting Japanese listeners' perceptual difficulty in identifying the place of articulation of word-final nasal stops even in Clear Speech. For Voiceless and Voiced contrasts, the Japanese group's performance was quite high (median > 94%), resulting in smaller language effects. This suggests that place identification in word-final oral stops in Clear Speech is relatively manageable for Japanese listeners.

In Fast Speech, on the other hand, the language effect was very large, not only for Nasal contrasts but also for Voiceless and Voiced contrasts, showing that Japanese listeners' increased difficulty in identifying word-final oral stops in Fast Speech. The AE group's performance in Fast Speech was slightly lower than that in Clear Speech but was still quite accurate for all contrast types (median > 90%), whereas the Japanese group's performance was markedly lower (median < 67% for Voiceless and Nasal contrasts, 82% for Voiced contrasts). The larger language effect size in Fast Speech than in Clear Speech indicates that Japanese listeners were more disadvantaged by Fast Speech than native-speaking AE listeners.

Further inspection of the data by target place for each contrast type revealed that the overall language effect observed for Voiceless and Voiced contrasts in Clear Speech were seen only for Alveolar /t/ and /d/, indicating that alveolar stops are hardest for Japanese listeners to identify among word-final oral stops in Clear Speech. In contrast, the language effects were seen in all Nasal contrasts in Clear Speech and all contrast types in Fast Speech. The uneven pattern of language effects for oral stop contrasts in Clear Speech was not expected. Considering the fact that the effects were relatively small for these contrast types and that the number of AE participants was smaller (N = 12) than for the Japanese group (N = 24), there seems to be a possibility that the inconsistency becomes insignificant after obtaining more data from AE listeners. Further examination will be necessary before determining the finding conclusive.

4.1.1.1. Inconsistency between Aoyama (2003) and current study.

It also should be noted that Aoyama (2003) did not find a significant difference between AE and Japanese listeners in their identification on word-final /m/-/n/ comparisons, which is inconsistent with the results of the current study, although she did find better performance by AE listeners on the /m/-/ŋ/ and /n/-/ŋ/ comparisons. A crucial difference between the Aoyama study

and the current study was that the stimuli used in the Aoyama study were isolated words whereas the current study adopted the words produced in connected speech where the target word-final stops were followed by another stop with varied place of articulation. This difference is critical considering the Japanese phonological rule of obligatory place assimilation for final nasals. It is reasonable that Japanese listeners were more confused by word-final nasals in context than those in isolated words for place identification because of the context-dependent nature of final nasals in their L1.

The more complex stimuli and higher task demands of the current study than the Aoyama study also may have played a role in the inconsistency between the two studies. In the Aoyama study, each stimulus was repeated twice while participants were looking at the written alternatives, whereas each stimulus was played only once and the written alternatives appeared only at the end of the auditory presentation in the current study. In the former case, listeners were aware of which segment to listen for whereas in the latter case, listeners were not certain to which segment to attend until the auditory presentation was completed. Furthermore, the Aoyama study adopted a two-alternative forced-choice identification task in which the chance level was 50% whereas the present study adopted a three-alternative forced-choice identification task, the chance level of which was 33%.

The ASP model of speech perception (Strange, 2011) stresses the important role that the complexity of the stimulus materials and task demands play in L2 perception. According to the ASP model, when the stimulus materials and task demands are simple, listeners are able to utilize a phonetic mode of perception where they are able to attend to detailed acoustic-phonetic differences, leading to good discrimination even of quite difficult L2 contrasts. However, as the materials and task demands become more complex, the phonetic mode becomes less available to

L2 listeners and they increasingly turn to their L1 phonological mode, which may not be effective for L2 perception. As a result, the same L2 contrasts that listeners were able to discriminate in a simple task may become very challenging in a high-demand task with complex stimuli. There may be a possibility that the more demanding task with more complex stimuli used in the current study, which is closer to real-life listening conditions, contributed to the different results between the two studies.

4.1.2. Clear Speech benefit: better perception in Clear Speech than in Fast Speech.

4.1.2.1. Overall performance.

Another major hypothesis of the present study concerns the Clear Speech benefit, which was based on the assumption that the acoustic properties cueing the place of articulation of word-final stops would be less robust in Fast Speech than in Clear Speech. The hypothesis regarding the Clear Speech benefit predicted that both AE and Japanese listeners' place perception of word-final stops would be better in Clear Speech than in Fast Speech. The hypothesis was supported by the results revealing significantly more accurate overall performance in Clear Speech than in Fast Speech by both language groups.

An interaction between Clear Speech benefit and language effect, that is, a larger Clear Speech benefit for Japanese than for AE listeners, was also hypothesized. AE listeners were expected to be less affected by Fast Speech than Japanese listeners because of their better capability to use the remaining acoustic information available in Fast Speech. Although the study could not directly measure the significance of the interaction statistically because parametric tests were not applicable to this comparison, there was a strong indication of an interaction. A very large effect of speech mode by the Japanese group ($r = 0.76$) in comparison with the effect seen in the AE group ($r = 0.59$) was observed. Along with the larger language effect in Fast

Speech than in Clear Speech discussed above, the results support the hypothesis of the interaction between speech mode and language effect. The finding of Japanese listeners being more negatively affected by Fast Speech than AE listeners is in accordance with the claim of the ASP model (Strange, 2011) that a high-demand task has a more negative impact on L2 perception than on L1 perception, as discussed earlier. The durational measurements in the acoustic analysis confirmed that the Fast Speech stimuli were constantly shorter than the Clear Speech stimuli. This means that, within the same amount of time, a listener had to process more speech segments in Fast Speech than in Clear Speech, facing a more taxing listening condition that may affect L2 listeners more negatively. Further consideration of the Clear Speech benefit for different stimulus types within each language group is discussed in the next section.

4.1.2.2. Clear Speech benefit for each contrast type.

It was hypothesized that Clear Speech benefit seen in the overall performance would be more evidently seen for oral stop contrasts, in which absence or crucial reduction of release bursts in Fast Speech was expected, than for nasal stop contrasts that would better retain the acoustic information in the nasal murmur even in Fast Speech. The acoustic analysis confirmed that invariant acoustic information of place of articulation was actually present in most final oral stop release bursts, showing a consistent pattern of spectral shapes for most Clear Speech tokens (68-81%) except for /d/ (32%). The acoustic analysis also confirmed the assumption that the stop release would be absent or crucially reduced in Fast Speech. Aside from the release burst cues, very consistent place cues were found in the F2 and F3 transitions of the preceding vowels, which were uniformly seen across speech modes as well as across contrast types sharing the same preceding vowel. Thus, the F2 and F3 formant transitions are assumed to be the primary cue utilized for place distinctions among unreleased oral stop contrasts. Compared to the oral

stops, the reduction of acoustic cues in the nasal stops seemed to be much less evident in Fast Speech because the nasal murmurs were shorter in Fast Speech but were still present in all tokens. In addition, the aforementioned F2 and F3 formant transition cues showing the same patterns as those seen in oral stop contrasts were present. Thus, there were redundant acoustic cues for place of articulation in final nasal stops even in Fast Speech.

The larger Clear Speech benefit for oral stop contrasts than for nasal stop contrasts was expected to be seen in both language groups because the hypothesis was based on the assumption that the degrees of curtailment of the acoustic information between oral and nasal stop contrasts in Fast Speech would be different, and results of the experiments were as expected. The AE group's performance was significantly more accurate in Clear Speech than in Fast Speech for Voiceless and Voiced contrasts but not for Nasal contrasts. The effect was larger for Voiceless (r = 0.66) than for Voiced contrasts (r = 0.45), which is compatible with the past finding that the native perception of English final voiced stops was less dependent on the stop release information than that of voiceless stops (Deelman & Connine, 2001). The AE group's identification of Nasal contrasts was highly accurate in both speech modes (median > 98%), indicating that sufficient information about place of articulation of nasal stops for native listeners remained in Fast Speech. For the Japanese performance, a Clear Speech benefit was seen for all contrast types, but the effect was smallest for Nasal contrasts. The decrease in accuracy in Fast Speech was especially evident for Voiceless contrasts (median = 94.4% in Clear, 66.7% in Fast Speech), exhibiting the largest effect size among the three contrast types (η_p^2 = 0.85). The effect size for Voiced contrasts was smaller than for Voiceless contrasts (η_p^2 = 0.41) but was still larger than Nasal contrasts (η_p^2 = 0.20). The interaction between speech mode and contrast type was

significant, supporting the hypothesis of a much smaller Clear Speech benefit for word-final nasal stops for the Japanese group.

The Japanese listeners' perceptual accuracy on Voiceless and Voiced contrasts was assumed to be primarily dependent on the availability of information in stop releases, which was heavily affected by the speech mode. In contrast, their perception of Nasal contrasts was hypothesized to be negatively influenced by their L1 phonological rule, so the availability of acoustic information may have little influence on their perception, which seems to have caused the small Clear Speech benefit. Thus, although the results showing a small Clear Speech benefit for Nasal contrasts may look similar to those of AE listeners showing no Clear Speech benefit, the reasons for each group were different. No Clear Speech benefit was seen for Nasal contrasts in the AE perception because their performance was highly accurate in both speech modes whereas the Clear Speech benefit was small in the Japanese perception because they had perceptual difficulty not only in Fast speech but also in Clear Speech.

4.1.3. Contrast type comparisons: Voiceless *vs.* Voiced *vs.* Nasal.

4.1.3.1. Perception by AE listeners.

Since AE listeners were expected to have no difficulty in identifying English word-final oral and nasal stops in clearly articulated speech, the AE group's significantly less accurate performance on Voiceless contrasts than on Voiced and on Nasal contrasts in Clear Speech was an unexpected finding. In fact, the overall differences were very small and became nonsignificant when one token of the stimuli *hock cautiously*, on which 10 out of 12 AE subjects answered incorrectly, was excluded from the analysis. However, considering the fact that the other token of *hock cautiously* was the second least accurate token and that the stimuli containing *hock* tended to be error-prone items for both language groups in both speech modes (see the error

analysis section in Results), it does not seem to be appropriate to simply ignore the token to consider that AE listeners' perception was equally good on all contrast types in Clear Speech. This issue will be discussed more in a later section (4.2.1. Errors on oral stop contrasts).

The AE listeners' performance pattern in Fast Speech showed that participants were most accurate on Nasal contrasts and least accurate on Voiceless contrasts with Voiced contrasts in the middle. The significantly more accurate performance on Voiced contrasts than on Voiceless contrasts supported the prediction based on the findings of Deelman and Connine (2001). The most accurate performance on Nasal contrasts also had been assumed because the nasal stops were likely to retain more acoustic information than the oral stops in Fast Speech because of the presence of the nasal murmurs during the occlusion portions. Verifying the predictions, results of the AE performance on each contrast type in Fast Speech showed the patterns in line with the premise that AE listeners' perception would be affected by the relative availability of redundant acoustic cues.

4.1.3.2. Perception by Japanese listeners.

As already discussed above, the results point to Japanese listeners' conspicuous perceptual difficulty in identifying word-final nasal stops in both speech modes. Thus, as expected, Japanese performance on Nasal contrasts was significantly less accurate than that on Voiceless and on Voiced oral contrasts in Clear Speech. A significant difference in performance was found between Nasal and Voiceless contrasts and between Nasal and Voiced contrasts, but not between Voiceless and Voiced contrasts in Clear Speech. The results were also consistent with the small Clear Speech benefit for Nasal contrasts for Japanese listeners because of their perceptual difficulty even in Clear Speech. Anecdotally, almost all Japanese participants in the experiments, regardless of their English proficiency, expressed surprise over their inability to

identify the word-final nasal stops. This indicates a lack of awareness of this perceptual problem by Japanese learners of English.

In Fast Speech, the Japanese listeners' performance on Nasal contrasts was even lower, but the difference between Nasal and Voiceless contrasts was not significant because their performance on Voiceless contrasts was very low in Fast Speech as well. This decrease in accuracy for Voiceless contrasts was reflected in the large Clear Speech benefit in Voiceless contrasts discussed earlier. Results showed significantly more accurate performance on Voiced contrasts than both Nasal and Voiceless contrasts in Fast Speech.

The patterns of performance seen in the Japanese group comparing contrast types were all as predicted except for the unpredicted significant difference between Voiceless and Voiced contrasts in Fast Speech. As with the case for the AE group, Japanese listeners' better performance on Voiced than on Voiceless contrasts seems to reflect the different degree of availability of acoustic information between Voiceless and Voiced stops in Fast Speech. This perceptual pattern had been predicted for AE listeners but not for Japanese listeners because Japanese listeners were expected to rely heavily on the release cues for oral stops and to be unable to use other cues effectively. Finding better performance on Voiced than Voiceless contrasts by Japanese listeners suggests that Japanese listeners were actually utilizing acoustic information other than stop release cues, most likely the F2 and F3 transitions, to some extent when identifying word-final oral stops. In order to examine whether the information of F2 and F3 transitions were actually better present for voiced than for voiceless in the Fast Speech tokens, more detailed measures of F2-F3 transitions, such as examination of the slope differences in offsets, would be necessary. Follow-up studies should clarify this issue.

4.1.4. Influence of following place contexts: Different *vs.* Same.

4.1.4.1. Influence of stop release: oral stops in Same context in Clear Speech.

The results of the current study have indicated that Clear Speech benefit was more evident in Voiceless and Voiced contrasts than in Nasal contrasts by both language groups. The study discussed the possible involvement of the availability of the target stop release cues in the performance, especially for the Japanese group. The acoustic analysis showed that, for the speaker of this study with a New York dialect, the release of a word-final oral stop preceding another oral stop with the same place of articulation (i.e., in Same context) was deleted or critically reduced even in Clear Speech. Thus, it was hypothesized that Japanese listeners' performance would be less accurate in Same context even in the Clear Speech condition, due to their heavy reliance on the release cues. For AE listeners, who are accustomed to unreleased final stops followed by a consonant sharing the same place of articulation, identifying the place of articulation of final oral stops in Same context without release in Clear Speech was expected to be no worse than perceiving place of those in Different context.

Results showed that the Japanese listeners' performance in Same contexts in Clear Speech was significantly less accurate than that in Different contexts for /p/ and /k/ in Voiceless contrasts and /b/ in Voiced contrasts. Although not all target places in Voiceless and Voiced contrasts reached significance, no significant trend in the opposite direction was found. Thus, the hypothesis that Japanese listeners depend on burst release cues even in Clear Speech was further supported. An unexpected finding was that less accurate performance in Same context in Clear Speech was seen in the AE performance on Voiceless Velar /k/ as well. The significance was maintained even after disregarding the previously mentioned error-prone token *hock cautiously*, which may suggest that English word-final /k/ in geminate context without release is potentially

confusable at a general auditory level even when it was clearly articulated, but unreleased. More consideration regarding this issue will be given in section 4.2.1.

4.1.4.2. Other observations.

Another finding was that significantly less accurate performance in Same context was also observed in the Japanese performance on Nasal Labial /m/ in both Clear and Fast Speech. This perceptual pattern by Japanese listeners is noteworthy because the disadvantage of Same context was seen even in Nasal contrasts where stop releases did not occur. The error analysis section will deliberate this issue further.

The influence of the following context in Fast Speech did not seem to have very systematic patterns for either language group. For AE listeners, /p/ and /k/ were better perceived in Different context, but /t/ and /m/ were better perceived in Same context. For Japanese listeners, /p/ and /m/ were better perceived in Different context, but /t/, /k/, and /g/ were better perceived in Same context. In general, labials tend to be perceived better in Different context by Japanese listeners and /k/ was better perceived by AE listeners in Different context. These patterns could be attributable to merely random factors or to idiosyncrasies of this speaker's productions. Further research is needed to determine whether they are general patterns or not.

4.1.5. Influence of oral stop release on performance.

The hypotheses predicting Japanese listeners' difficulty in correctly perceiving Voiceless and Voiced contrasts in Fast Speech and Voiceless and Voiced contrasts in Same context in Clear Speech were based on the assumption that Japanese listeners would heavily rely on the acoustic information in the stop releases for the place identification of word-final oral stops. The correlations between the magnitudes of the oral stop releases based on the acoustic analysis and the performance on the corresponding tokens by each language group confirmed this assumption.

The Japanese performance on both Voiceless and Voiced contrasts in Clear Speech showed rather strong positive correlations, indicating Japanese listeners' reliance on the acoustic information in the stop release, especially for Voiceless contrasts, although there were some tokens with little or no release that were identified accurately. A follow-up study will look into these tokens to see if transitional information was better present in them.

The AE group's performance on Voiceless contrasts also indicated a positive correlation in Clear Speech, but the correlation was weaker than that of Japanese, and their performance on Voiced contrasts did not show a correlation, due to ceiling effects truncating the range in performance. The results are also consistent with past findings that listeners whose L1 has no or limited word-final stops rely on L2 stop releases (e.g., Flege, 1989; Flege and Wang, 1989) and that AE listeners are able to tap into anticipatory acoustic information available in the preceding vowel and transitional segments of word-final stops (Warren & Marslen-Wilson, 1987; 1988).

Neither of the language groups showed significant correlations with any of the acoustic measures of release bursts in Fast Speech, due largely to the truncation of range of variation in those variables.

4.1.6. Correlations of Japanese performance with language experience and proficiency.

The last hypothesis concerns the predicted positive relationship between Japanese listeners' performance and their language experience and proficiency. The Japanese listeners' overall performance was significantly correlated with their LOR as was performance on all contrast types in both speech modes except for Voiceless contrasts in Fast Speech. The Spearman's *rho* values for different contrast types ranged from .31 (Voiceless in Fast Speech) to

.66 (Voiced in Fast Speech), indicating that improvement with immersion experience differs for different kinds of final stop contrasts.

What was more notable in the scatter plots was the large variability of the correlations of LOR with performance on Voiceless and Voiced contrasts in Fast Speech and with performance on Nasal contrasts in both Clear and Fast Speech conditions. Japanese listeners with five or fewer years of LOR showed consistently poor performance on these contrasts, but Japanese listeners with longer LORs showed much greater variability in their performance. This pattern was especially evident in the correlations with the performance on Voiceless and on Nasal contrasts in Fast Speech, suggesting the considerable difficulty in correctly perceiving those contrasts even for those with over 10 years of immersion experience. These observations indicate that a long LOR alone may not guarantee the acquisition of difficult L2 contrasts, which has often been pointed out by Flege and colleagues (e.g., Flege & Liu, 2001).

As briefly mentioned in Introduction, the SLM (Flege, 1995) and the PAM-L2 (Best & Tyler, 2007) define "experienced" L2 learners using different criteria, and make different predictions about how much immersion experience produces asymptotic performance on phonetic perception performance. Whereas Flege and colleagues often adopted three years or much longer LOR as the criterion for "experienced" L2 learners, PAM-L2 suggests that most perceptual learning occurs in the first 6-12 months of L2 immersion. In the present study, the scatter plot showing Japanese performance on Nasal contrasts in the Clear Speech condition indicates that the improvement of the performance as a function of LOR is not evident after four to five years of LOR, suggesting a possible cut-off point of the effects of L2 immersion of perceptual performance. The plots of all contrast types in Fast Speech also indicate greater variability in individual performance after four to five years of LOR, suggesting that LOR may

not be the best predictor of perceptual mastery (see Figure 15). In general, however, the observed patterns are more in line with the SLM's claim that L2 perception may continue to improve beyond the first year of immersion.

Another noteworthy pattern of correlation was that the AOA showed significant negative correlations with Japanese listeners' performance on all contrast types only in Fast Speech. This may indicate that language experience before adulthood, even in the mid to late teens (>14 years), is advantageous for perceiving less clearly articulated speech.

The Japanese performance on all contrast types in both speech modes was also positively correlated with their language proficiency as measured by Versant ^{TM} English Test. In fact, correlation coefficients were in general higher for language proficiency than for LOR. This tendency was more apparent in Fast Speech, suggesting that for the perceptual mastery of challenging L2 phonetic contrasts, long immersion experience alone may not be good enough and improving overall L2 proficiency may be required before these difficult contrasts are mastered.

Among the subscores of the Versant Test, the vocabulary score and sentence mastery score appeared to be better predictors of performance on challenging oral stop contrasts, such as Voiceless contrasts in Fast Speech, whereas the pronunciation score and fluency score seemed to be stronger for predicting performance on easy oral stop contrasts, such as Voiced contrasts in Clear Speech. The observation of the high correlation between L2 vocabulary and performance on difficult L2 contrasts is compatible with the argument by Best and Tyler (2007) regarding the relationship between the establishment of the L2 lexicon and the perceptual development of L2 phonological contrasts. The authors claim that the "lexical pressure" forcing L2 listeners to "re-phonologize" L2 contrasts to fill a need to differentiate minimally contrasting L2 words

promotes establishing L2 phonological categories requiring the discernment of L2 phonetic details. Thus, based on their argument, L2 listeners with bigger vocabularies would have a stronger necessity to be able to discriminate hard L2 contrasts. These patterns should be compared with different language groups of L2 listeners in future studies to examine whether or not their argument can be born out.

4.2. Patterns of Errors Observed in Main Experiment

4.2.1. Errors on oral stop contrasts (Voiceless and Voiced).

The error patterns in Clear Speech discussed here are mainly the observations of the Japanese listeners' perceptual error patterns because only small numbers of errors were made by AE listeners. First, there was a tendency for Labials and Velars in Same contexts to be misperceived as Alveolars by Japanese listeners, which was seen in both Voiceless and Voiced oral consonants but was more evident in Voiceless contrasts. This response bias toward Alveolars was also observed in the errors made by AE listeners on Voiceless Velar /k/ followed by /k/. Since the target stops of Voiceless and Voiced contrasts were either unreleased or crucially reduced in Same context in Clear Speech, this error pattern can be interpreted as showing that Japanese listeners tended to misperceive the non-alveolar unreleased stops as alveolar. The fact that this error pattern was also seen in the AE perception of /k/ followed by /k/ seems to indicate that the place information of word-final /k/ in geminate context is limited compared to non-geminate contexts and other stops, and that the perceptual bias toward alveolar when listening to unreleased final stops is a general tendency rather than a language-specific L2 perceptual pattern.

The error responses by Japanese listeners on Alveolars were not as consistent as their errors on non-Alveolars, except for the pattern showing a bias toward labial when followed by

the labial /p/, which was seen for both /t/ and /d/. This perceptual pattern showing the influence of the following place was not seen in the AE perception in Clear Speech, but was seen more prevalently in Fast Speech for Japanese as well as for AE in certain contexts, as discussed below.

The error patterns observed in Fast Speech were essentially the same as those seen in Clear Speech, except that error rates were greater. The error bias toward Alveolar for Labials and Velars in Same contexts by Japanese listeners was seen in both Voiceless and Voiced contrasts. This bias toward Alveolar was also observed in the errors by AE listeners on /p/ followed by /p/ and /k/ followed by /k/, and on /g/ followed by /k/, suggesting that this error pattern may generally emerge when listeners do not have sufficient place information from the speech signal, regardless of L1 or L2 perception.

The error pattern on Alveolars showing a bias toward the following place, seen in the Japanese errors in Clear Speech, was more prevalent in Fast Speech: /t/ and /d/ followed by /p/ tended to be heard as /p/ and /b/; /t/ and /d/ followed by /k/ as /k/ and /g/ by Japanese listeners. This perceptual influence of the following place was also seen in the AE errors on /d/ followed by /p/. Although the perceptual patterns of biasing toward the following place seem to suggest the possibility of regressive place assimilation, the acoustic analysis of the F2 and F3 transition data showed no evidence of place assimilation. In fact, the data rather clearly indicated that there was no place assimilation, showing typical place characteristics of F2 and F3 transitions that were distinct from those of the other target places, regardless of the following context (see Appendix E).

To summarize, there seem to be two recognizable tendencies in the errors on Voiced and Voiceless contrasts: the bias toward Alveolars in the errors on Labials and Velars in Same context; and the bias toward the following place in the errors on Alveolars in Different context.

Both of these patterns seem to have a tendency to be more apparent in Voiceless than in Voiced, more apparent in the Japanese error patterns than in the AE error patterns, and more apparent in the Fast Speech condition than in the Clear Speech condition. Thus, it is reasonable to conclude that the emergence of these observed error patterns was caused by limited availability of acoustic information and/or listeners' difficulty in tapping into the availability of acoustic information for the place identification of word-final oral stops. Before discussing the errors on Nasal contrasts, a particular deliberation on the error pattern of unreleased word-final /k/ is given.

4.2.1.1. Confusability of unreleased word-final /k/.

As observed and briefly considered so far, there seemed to be an apparent tendency for both language groups to misperceive unreleased word-final /k/ and to be biased toward the alveolar /t/ in their incorrect responses. Specifically, high error rates for /k/ followed by /k/ in Clear Speech and for /k/ in Fast Speech regardless of the following context were frequently observed. Although these perceptual errors seem to be language-independent, the perceptual patterns and the corresponding acoustic similarities do not necessarily seem to coincide with each other. For example, the words *hock* and *shock* were the most typical error-prone items in Clear Speech in geminate context and in Fast Speech by both language groups, but the acoustic analysis revealed that F2 and F3 formant transitions of /k/ preceded by the vowel /ɑ/ are very close to those of /p/ but not those of /t/ (see Appendix E). Since F2 and F3 formant transitions are considered to be the major place cues for unreleased word-final oral stops, if the observed error response bias was caused by the acoustic similarity, /p/ would have been chosen as the answers instead of /t/. A possible explanation for this inconsistency seems to be the influence of lexical effect. According to the $SUBTL_{WF}$ scores indicating the word frequency per million words, the word frequency for *hot* (189.84 points) is 88 times higher than *hock* (2.16 points) and that of *shot*

(227.43 points) is 8 times higher than *shock* (28.78 points). The frequency for *hop* (19.16 points) and that for *shop* (53.55 points) are intermediate for both cases. The word familiarity results for the Japanese group also showed that the mean familiarity rates of *hot* (6.75 points for Clear, 7 points for Fast) was much higher than those of *hock* (4.79 points for Clear, 4.45 points for Fast). This interpretation does not contradict the observation of Ito and Strange (2009) and of Warren and Marslen-Wilson (1987; 1988) stating that word familiarity/word frequency effects were found only when listeners could not utilize the acoustic information effectively or when the acoustic information in the stimuli was ambiguous. Although the statistical analysis did not find an overall lexical effect on the performance in the present study, there might have been a certain degree of lexical influence when the acoustic information was limited in such a case as unreleased final /k/.

4.2.2. Errors on Nasal contrasts by Japanese listeners

The Japanese listeners' error rates for Nasal contrasts were conspicuously high, which was not seen in the AE performance. Since the difficulty had been predicted by the assumption that the Japanese phonological rule of final nasals would interfere with Japanese listeners' perception, the patterns of Japanese errors on Nasals was expected to be different from those found in Voiceless and Voiced contrasts that seem to be based on more general perceptual system. The observed error patterns for Nasals were in fact different. First, the error patterns by Japanese on Nasals in the two speech modes were very similar to each other. This observation is in line with the notion that the Japanese errors on Nasals stem from the negative influence of an L1 phonological rule and that, therefore, Clear Speech would not improve perception effectively. The following are the specific error patterns seen in both speech modes: 1) /ŋ/ tended to be misheard as /n/ regardless of the following place; 2) /n/ tended to be misheard as /ŋ/ regardless of

the following place, except for /n/ followed by /p/ in Clear Speech; 3) /m/ did not have a bias toward /n/ or /ŋ/, except for the bias toward /n/ in Same context; 4) /m/ was less confusable for Japanese listeners except for the one in Same context.

The overall patterns observed in the Japanese errors on Nasal contrasts had a number of similarities to the findings of Aoyama (2003), in spite of the fact that her study used isolated words whereas the present study adopted words in connected speech. The first two patterns of the present study highlighted the Japanese listeners' confusion between word-final /n/ and /ŋ/, which was one of the major findings of the Aoyama study. The bias toward /n/ in the errors on /ŋ/ was also common to the two studies although the bias toward /ŋ/ in the errors on /n/ found in the present study was not found in the Aoyama study. This word-final /n/-/ŋ/ confusion by Japanese listeners was seen even in the stimuli containing the preceding vowel /æ/, which is remarkable considering the fact that the more tensed vowel quality for the /æ/ when followed by /ŋ/ (see Acoustic Analysis 2.3.6.) would have signaled additional place information of the following nasal stop. Aoyama (2003) also pointed out the relatively easy identification of word-final /m/, which is compatible with the fourth observation of the present study. A unique finding of the present study was that /m/ became more confusable in Same context, showing a bias toward /n/, which seems to indicate a negative influence of the following context. In sum, the errors observed in Japanese listeners' place identification of word-final nasals in this study revealed to have consistent patterns, quite a few of which were in line with the previous findings of the study examining word-final nasals presented in isolation.

Since the Japanese results in the present study on the word-final nasals were in common in several points with those of Aoyama (2003), discussing the results of the present study in the framework of the PAM (Best, 1995) by adopting the perceptual assimilation data from the

Aoyama study seems to be worthwhile. According to Aoyama (2003), Japanese listeners assimilated the word-final /m/ to the Japanese /mɯ/ almost all the time, but they never assimilated the word-final /n/ and /ŋ/ to the Japanese /mɯ/. On the other hand, Japanese listeners chose the same Japanese orthographic representations transcribed as /N/ and /Ngɯ/ for both /n/ and /ŋ/, showing a considerable overlapped pattern for perceptual assimilation of the word-final /n/ and /ŋ/. Based on these data, the final /m/-/ŋ/ and /m/-/n/ contrasts were identified as Uncategorized-Categorized (UC), the perceptual distinction of which was rated as very good, and /n/-/ŋ/ as Uncategorized-Uncategorized (UU) assimilation, the perception of which was considered as "relatively poor" in this particular case because of the overlapped responses (Aoyama, 2003, p. 263). The author concluded that these perceptual assimilation patterns could explain the relatively easy perceptual distinction of the /m/-/ŋ/ and /m/-/n/ by Japanese listeners and the difficulty that they have in the /n/-/ŋ/ distinction. This explanation works well for the present study as well. Fewer mistakes were made for the identification of the word-final /m/ by Japanese listeners because their perceptual assimilation pattern of /m/ was different from those seen in /n/ and /ŋ/. The bias toward /ŋ/ for the errors on /n/ and vice versa can also be explained by the UU pattern showing the overlapping perceptual assimilation, confirming Aoyama's interpretations. The perceptual assimilation task using the stimuli of the present study for future research would consolidate this reasoning.

4.3. Conclusions and Directions for Future Research

The major objective of the present study was to shed light on the problems that listeners have in perceiving L2 word-final consonants in connected speech. The study examined Japanese listeners' perception of place-of-articulation contrasts of English word-final stops followed by word-initial stops. While studies have pointed out the perceptual difficulty in identifying L2 final

consonants produced in isolation, a possible negative impact of dynamic coarticulatory effects occurring in running speech on the L2 perception of final consonants had not been well taken into consideration. The present study addressed this issue by adopting word-final minimal triplets embedded in meaningful carrier sentences as stimuli, making the study closer to real life situations in listening to connected speech while controlling the phonetic environments of the stimuli. Detailed acoustic analyses were carried out and found that the place cues reported in the past studies of word-final stops examining isolated speech were present in the stimuli of the current study produced in context. The acoustic data of the present study also contribute details of how acoustic-phonetic information changes with changes in speech modes, at least for a single speaker of New York dialect.

Results of the experiment supported the hypotheses of the present study in most cases, while adding further information of L1 and L2 perception of English word-final stops by unexpected findings. It was found that Japanese listeners benefitted from clearly articulated speech for the perception of word-final oral stops, indicating that their perceptual problems are based on the availability of the acoustic information, as well as listeners' ability to take advantage of that information. The prediction of much heavier reliance on stop releases by Japanese than AE listeners was also supported by the results showing the Japanese listeners' marked disadvantage in identifying word-final oral stops with no or extremely reduced releases, as well as the positive correlations between the magnitudes of the oral stop releases and the Japanese performance. Unexpected findings were that in Clear Speech, final voiceless stops were relatively hard to perceive compared to voiced and nasal counterparts even for AE listeners, and that Japanese listeners perceived the voiced stops better than the voiceless stops in Fast Speech, suggesting the possibility of utilizing acoustic information remaining outside the release stop,

such as formant transitions of the preceding vowel. Both patterns were seen in both language groups, further supporting the notion that relative ease or difficulty of word-final oral stops is based on language-universal general perceptual patterns that depend on the availability of the acoustic information in the stimuli.

On the other hand, the benefit that Japanese listeners enjoyed from Clear Speech for the perception of final nasal stops in this study was relatively small, showing notably lower performance that contrasted with the AE listeners' ceiling performance in both speech modes. Moreover, quite a few instances of the observed errors were in line with the error patterns reported in a study examining Japanese listeners' perception of English word-final nasals produced in isolation (Aoyama, 2003), in which the patterns were explained by its perceptual assimilation data in the framework of the PAM. The fact that the PAM framework could offer the explanations of Japanese listeners' perceptual difficulty of final nasals also strongly suggests the involvement of L1 phonology in the Japanese perception of word-final nasal stops.

The present study also investigated correlations between the Japanese listeners' performance and the factors indicating their L2 language background, such as English immersion experience and proficiency. While the correlations of the performance with LOR were overall quite high, the highest correlations were seen with English proficiency, suggesting that effective improvement of English proficiency could lead to more successful perceptual mastery of challenging L2 phonetic contrasts than immersion experience in English-speaking countries alone. The higher correlations with AOA for the performance on the Fast Speech stimuli than for the Clear Speech stimuli implied that earlier L2 exposure is beneficial for effectively tapping into limited acoustic cues for L2 phonetic perception, even after the age of 14 years.

The current study investigated the perception of L2 phonetic contrasts while taking various factors that may affect listeners' perception into consideration, such as listeners' L1 (AE and Japanese), speech mode (Clear and Fast Speech), stimulus type (contrast type, target place, following place) and subject variables (age, LOR, AOA, English proficiency, language use). Although the present study introduced a number of findings and information that showed systematic patterns explicable as a function of these multiple variables, not all factors worth-exploring were incorporated in it. For example, the study adopted the stimuli produced by only one male speaker of one dialect, and the Japanese participants were not recruited in such a way that the study could identify the L2 perceptual cutoff point of LOR. The study also did not include a perceptual assimilation task for Japanese listeners that may have offered more detailed explanations of the perceptual difficulty of L2 phonetic contrasts by the Japanese, especially for the perception of the nasal stops. These issues should be addressed by replicating the study.

For an extension of the study, administering a training study to Japanese listeners would provide us with information about the effectiveness of direct training on the perception of challenging L2 place contrasts, such as unreleased oral stops and final nasal stops for Japanese. Training Japanese listeners for the final nasal contrasts would be especially interesting, considering their apparent lack of awareness of the perceptual problems. Manipulation of the stimuli, such as editing out the oral stop release or the following context and cross-splicing a release or nasal murmur onto a place-mismatched stimulus may also add the information regarding which acoustic cues are crucially utilized by L1 and L2 listeners.

Finally, the study needs to be extended to other language groups of English L2 learners to give a more comprehensive outlook of L2 perceptual difficulty of English word-final oral and nasal place contrasts. The present study assumes the role of the starting point of this line of

research. It should be followed by replicated and extended studies mentioned above to establish a series of studies, the accumulated findings of which would make a considerable contribution to the area of cross-language speech perception.

Appendix A

Stimulus Sentence List

Target Words (word-final minimal triplets: 54 words)			**Adverb**
contrast type	**vowel**	**minimal triplet stimuli**	
/p/-/t/-/k/	/ɪ/	**sip-sit-sick** **lip-lit-lick**	
	/æ/	**sap-sat-sack** **rap-rat-rack**	
	/ɑ/ or /ʌ/	**shop-shot-shock** **hop-hot-hock**	
/b/-/d/-/g/	/ɪ/	**rib-rid-rig** **bib-bid-big**	**positively**
	/æ/	**tab-tad-tag** **lab-lad-lag**	
	/ɑ/ or /ʌ/	**cob-cod-cog** **dub-dud-dug**	
/m/-/n/-/ŋ/	/ɪ/	**dim-din-ding** **Kim-kin-king**	
	/æ/	**ram-ran-rang** **bam-ban-bang**	**tauntingly**
	/ɑ/ or /ʌ/	**sum-sun-sung** **rum-run-rung**	
Fillers (word-initial minimal triplets: 24 words)			
contrast type		**minimal triplet stimuli**	**cautiously**
/p/-/t/-/k/		pick-tick-kick puff-tough-cuff pan-tan-can	
/b/-/d/-/g/		bet-debt-get bun-done-gun bait-date-gate	
/m/-/n/		mock-knock-(dock) mitt-knit-(bit) map-nap-(gap)	

Appendix B

Language Background Questionnaire for Japanese Subjects-1

Subject #
被験者番号 .....

Language Background Questionnaire (for Japanese subjects)
言語背景に関するアンケート(日本人被験者用)

Please answer the following questions:
以下の質問にお答え下さい：

Date Year Month Date
1. 日付： 年 月 日

Name
2. 氏名 ：

Address
3. 住所 ：

Telephone numbers:(Home) (Cellphone)
4. 電話番号：(自宅) - - (携帯) - -

email address
5. Eメールアドレス ：

Date of Birth Year Month Date
6. 生年月日：西暦 年 月 日

Gender Male Female
7. 性別： 男 / 女

Birthplace Prefecture
8. 生誕地 都・道・府・県

Native language
9. 母国語

Language(s) you speak and understand other than Japanese
10. 日本語以外に話し理解できる言語 1. 2. 3 4.

Foreign countries in which you have lived for more than 1 months
11. 1ヶ月以上住んだことのある外国

Country	City	From age	To age	year(s)	month(s)
国名	都市名	歳から	歳まで (	年	ヶ月)
国名	都市名	歳から	歳まで (	年	ヶ月)
国名	都市名	歳から	歳まで (	年	ヶ月)
国名	都市名	歳から	歳まで (	年	ヶ月)
国名	都市名	歳から	歳まで (	年	ヶ月)
国名	都市名	歳から	歳まで (	年	ヶ月)
国名	都市名	歳から	歳まで (	年	ヶ月)

Appendix B

Language Background Questionnaire for Japanese Subjects-2

Language use (Please skip the questions that are not applicable.)

12. 言語使用状況 (当てはまらない場合は記入しないで下さい。):

At home: Percentage of listening and speaking in Japanese
自宅：日本語で聞き話す割合：( ) %

Percentage of listening and speaking in English
英語で聞き話す割合：( ) %

Percentage of reading and writing Japanese
日本語で読み書く割合：( ) %

Percentage of reading and writing in English
英語で読み書く割合：( ) %

Percentage of watching TV in Japanese
日本語でテレビを観る割合：( ) %

Percentage of watching TV in English
英語でテレビを観る割合：( ) %

At school/work: Percentage of listening and speaking in Japanese
学校 / 職場：日本語で聞き話す割合：( ) %

Percentage of listening and speaking in English
英語で聞き話す割合：( ) %

Percentage of reading and writing Japanese
日本語で読み書く割合：( ) %

Percentage of reading and writing in English
英語で読み書く割合：( ) %

Other places: Percentage of listening and speaking in Japanese
その他の場所：日本語で聞き話す割合：( ) %

Percentage of listening and speaking in English
英語で聞き話す割合：( ) %

Percentage of reading and writing Japanese
日本語で読み書く割合：( ) %

Percentage of reading and writing in English
英語で読み書く割合：( ) %

If you use other languages than Japanese and English, please explain the circumstances in which you use those languages specifically.
日本語、英語以外の言語を使用する機会がある場合は、その言語と使用状況について具体的にお答え下さい。

Language
言語（ ）

Circumstances in which you use that language
使用状況 .....

.....

Language
言語（ ）

Circumstances in which you use that language
使用状況 .....

.....

Language
言語（ ）

Circumstances in which you use that language
使用状況 .....

.....

Experience in learning languages other than Japanese:

13. 日本語以外の言語を勉強した経験

Language
13-1. 言語.....

From age
.....歳から

To age
.....歳まで(

year(s)
年

month(s)
ヶ月)

Learning settings
学習方法:

Formal education in Japan
☐日本での学校教育

years
年、

conversation schools in Japan
☐日本の語学学校

years
年、

Teaching yourself
☐独学

years
年

Formal education overseas
☐海外での学校教育

Country
(国名

Type of school
：教育機関

Duration
：期間

years
年

months
ヶ月)

(国名 ：教育機関 ：期間 年 ヶ月)

(国名 ：教育機関 ：期間 年 ヶ月)

(国名 ：教育機関 ：期間 年 ヶ月)

Others
☐その他

Years
年

Please explain specifically:
(具体的にご説明下さい:)

Appendix B

Language Background Questionnaire for Japanese Subjects-3

Language　From age　To age　year(s)　month(s)

13-2. 言語＿＿＿＿＿＿＿＿＿ ＿＿＿歳から＿＿＿歳まで (　年　ヶ月)

Learning settings　Formal education in Japan　years　conversation schools in Japan　years　Teaching yourself　years

学習方法: ☐日本での学校教育　年、☐日本の語学学校　年、☐独学　年

Formal education overseas　Country　Type of school　Duration　years　months

☐海外での学校教育 (国名　：教育機関　：期間　年　ヶ月)

(国名　：教育機関　：期間　年　ヶ月)

(国名　：教育機関　：期間　年　ヶ月)

(国名　：教育機関　：期間　年　ヶ月)

Others　Years

☐その他　年

Please explain specifically:

(具体的にご説明下さい:　)

Language　From age　To age　year(s)　month(s)

13-3. 言語＿＿＿＿＿＿＿＿＿ ＿＿＿歳から＿＿＿歳まで (　年　ヶ月)

Learning settings　Formal education in Japan　years　conversation schools in Japan　years　Teaching yourself　years

学習方法: ☐日本での学校教育　年、☐日本の語学学校　年、☐独学　年

Formal education overseas　Country　Type of school　Duration　years　months

☐海外での学校教育 (国名　：教育機関　：期間　年　ヶ月)

(国名　：教育機関　：期間　年　ヶ月)

(国名　：教育機関　：期間　年　ヶ月)

(国名　：教育機関　：期間　年　ヶ月)

Others　Years

☐その他　年

Please explain specifically:

(具体的にご説明下さい:　)

Language　From age　To age　year(s)　month(s)

13-4. 言語＿＿＿＿＿＿＿＿＿ ＿＿＿歳から＿＿＿歳まで (　年　ヶ月)

Learning settings　Formal education in Japan　years　conversation schools in Japan　years　Teaching yourself　years

学習方法: ☐日本での学校教育　年、☐日本の語学学校　年、☐独学　年

Formal education overseas　Country　Type of school　Duration　years　months

☐海外での学校教育 (国名　：教育機関　：期間　年　ヶ月)

(国名　：教育機関　：期間　年　ヶ月)

(国名　：教育機関　：期間　年　ヶ月)

(国名　：教育機関　：期間　年　ヶ月)

Others　Years

☐その他　年

Please explain specifically:

(具体的にご説明下さい:　)

Thank you very much!!

ご協力ありがとうございました。

Appendix C

Language Background Questionnaire for American Subjects

Language Background Questionnaire (for English subjects)

Please complete the following questions:

1. Date: month / date / year

2. Name: _____

3. Home Address: _____

4. Telephone numbers: (Home) _____ - _____ - _____ (Cellphone) _____ - _____ - _____

5. email address: _____

6. Date of Birth month / date / year

7. Gender: Male / Female

8. Birthplace (City, State, Country): _____

9. Your first language is American English: Yes / No

10. Places in which you have lived for more than 1 year:

City/State/Country

_____ from age _____ to age _____ [_____ year(s) and _____ month(s)]

_____ from age _____ to age _____ [_____ year(s) and _____ month(s)]

_____ from age _____ to age _____ [_____ year(s) and _____ month(s)]

_____ from age _____ to age _____ [_____ year(s) and _____ month(s)]

_____ from age _____ to age _____ [_____ year(s) and _____ month(s)]

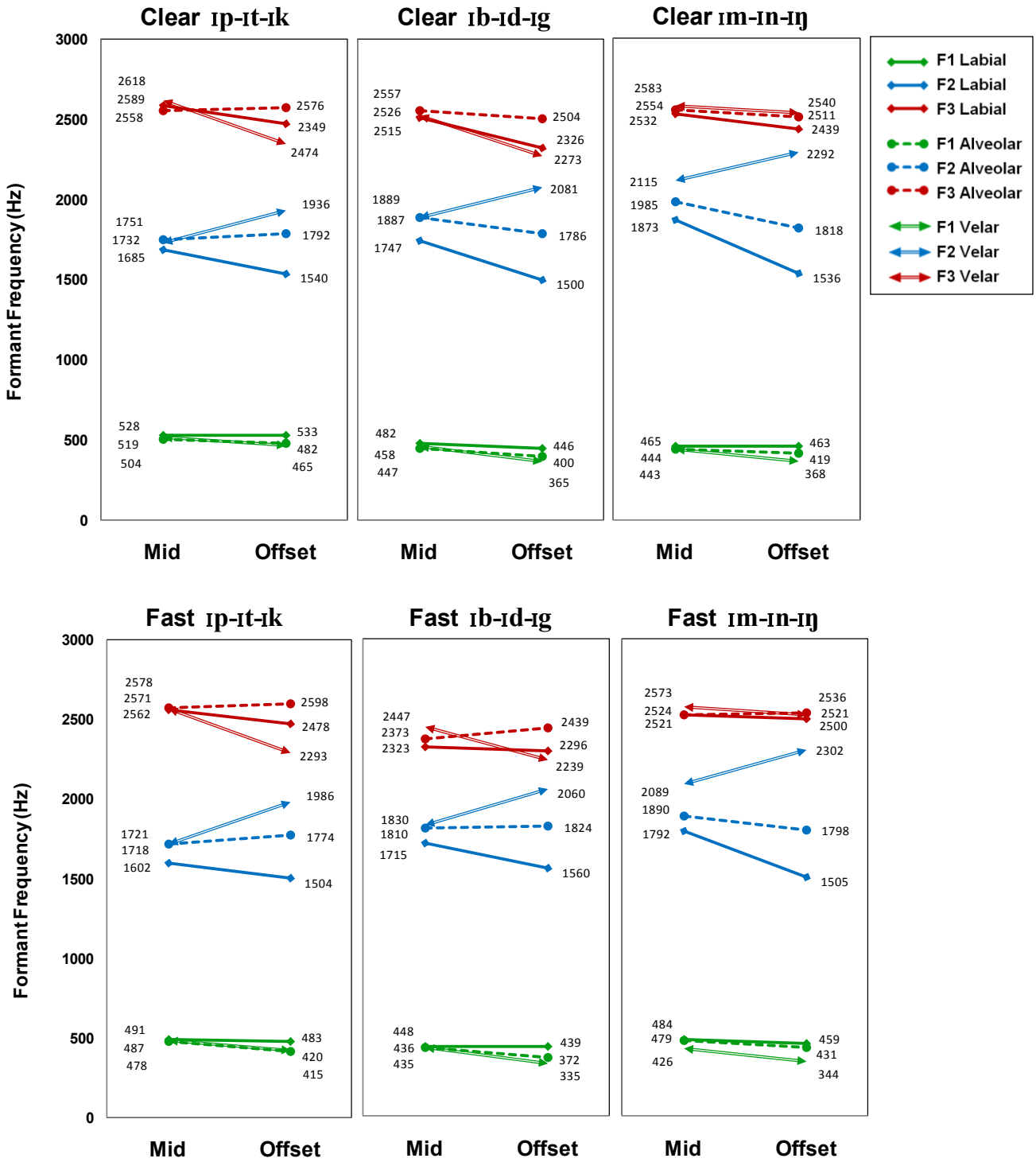
11. Language(s) you speak other than English: _____

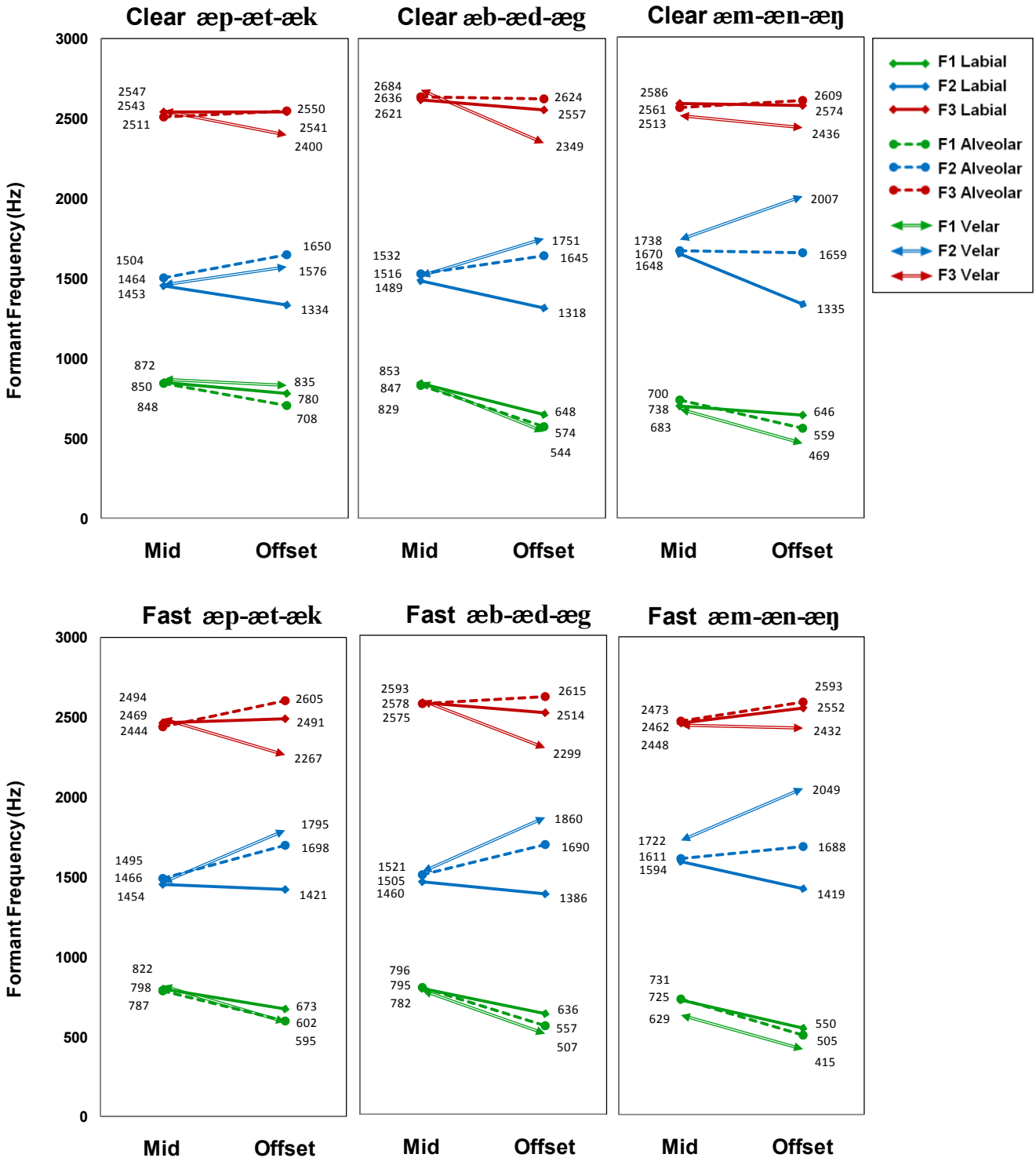
12. Experience in learning languages other than English:

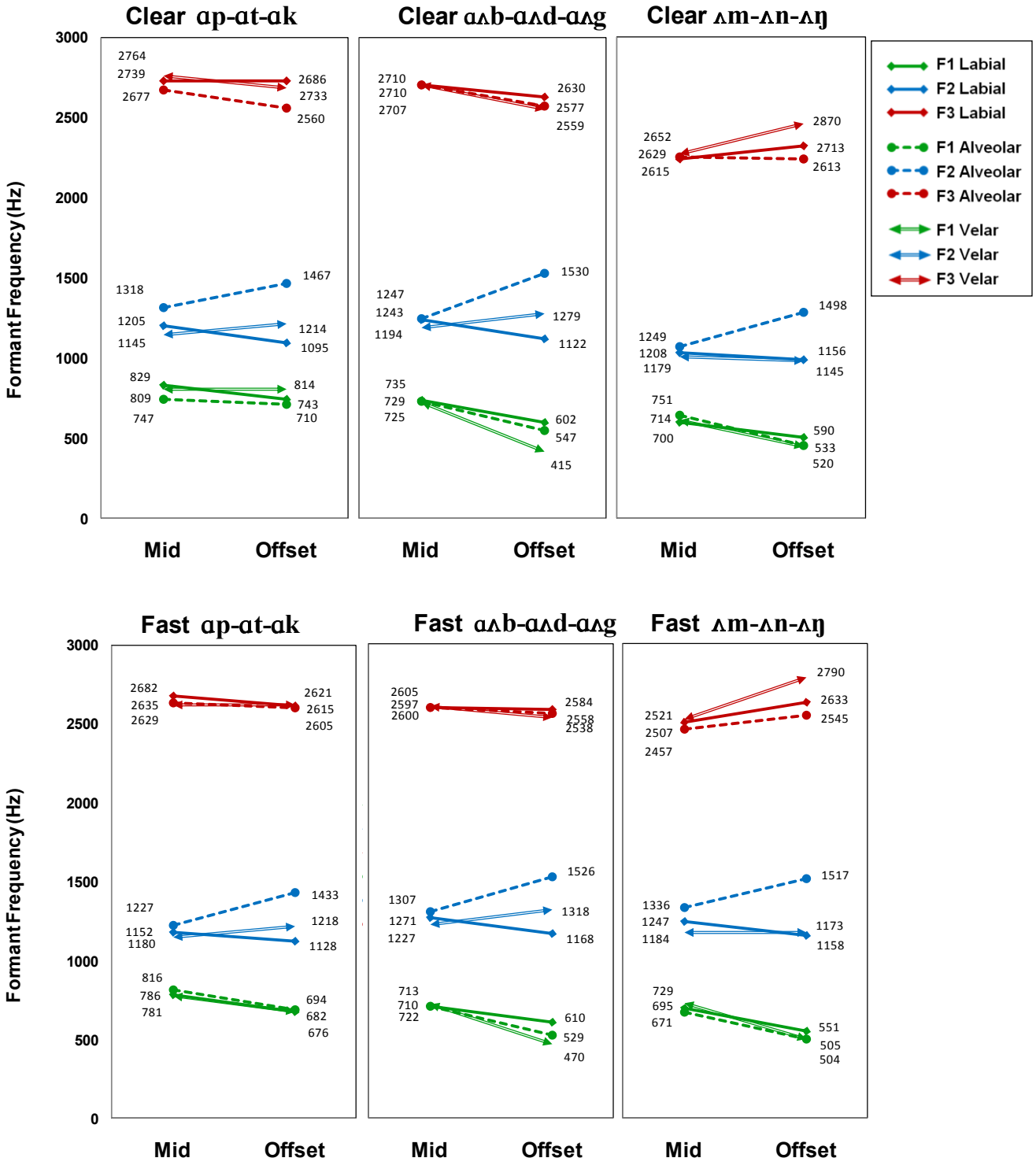
Language	age	learning settings (ex: at highschool, at home, etc.)
_____	age _____ to age _____	_____
_____	age _____ to age _____	_____
_____	age _____ to age _____	_____
_____	age _____ to age _____	_____

Appendix D

Formants Line Graphs Averaged Across Tokens Sharing Same Target Stops & Preceding Vowels

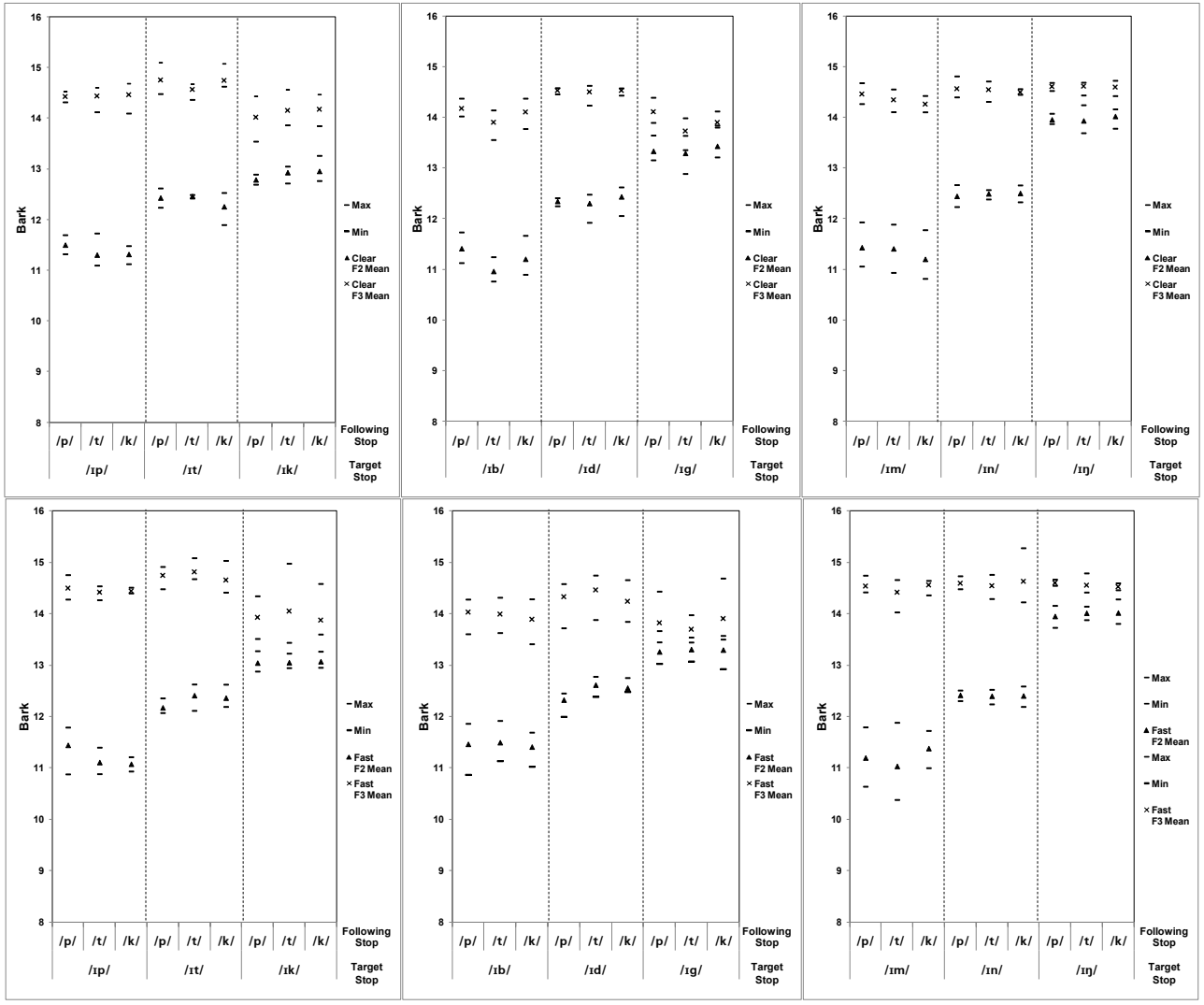



Clear æp-æt-æk
Clear æb-æd-æg
Clear æm-æn-æŋ
Fast æp-æt-æk
Fast æb-æd-æg
Fast æm-æn-æŋ
Formant Frequency (Hz)
3000
2500
2000
1500
1000
500
0
Mid
Offset
2547
2543
2511
2550
2541
2400
1504
1464
1453
1650
1576
1334
872
850
848
835
780
708
2684
2636
2621
2624
2557
2349
1532
1516
1489
1751
1645
1318
853
847
829
648
574
544
2586
2561
2513
2609
2574
2436
1738
1670
1648
2007
1659
1335
700
738
683
646
559
469
2494
2469
2444
2605
2491
2267
1495
1466
1454
1795
1698
1421
822
798
787
673
602
595
2593
2578
2575
2615
2514
2299
1521
1505
1460
1860
1690
1386
796
795
782
636
557
507
2473
2462
2448
2593
2552
2432
1722
1611
1594
2049
1688
1419
731
725
629
550
505
415
F1 Labial
F2 Labial
F3 Labial
F1 Alveolar
F2 Alveolar
F3 Alveolar
F1 Velar
F2 Velar
F3 Velar


Clear ɑp-ɑt-ɑk
Clear ɑʌb-ɑʌd-ɑʌg
Clear ʌm-ʌn-ʌŋ
Fast ɑp-ɑt-ɑk
Fast ɑʌb-ɑʌd-ɑʌg
Fast ʌm-ʌn-ʌŋ
Formant Frequency (Hz)
3000
2500
2000
1500
1000
500
0
Mid
Offset
F1 Labial
F2 Labial
F3 Labial
F1 Alveolar
F2 Alveolar
F3 Alveolar
F1 Velar
F2 Velar
F3 Velar
2764
2739
2677
2686
2733
2560
1318
1205
1145
1467
1214
1095
829
809
747
814
743
710
2710
2710
2707
2630
2577
2559
1247
1243
1194
1530
1279
1122
735
729
725
602
547
415
2652
2629
2615
2870
2713
2613
1249
1208
1179
1498
1156
1145
751
714
700
590
533
520
2682
2635
2629
2621
2615
2605
1227
1152
1180
1433
1218
1128
816
786
781
694
682
676
2605
2597
2600
2584
2558
2538
1307
1271
1227
1526
1318
1168
713
710
722
610
529
470
2521
2507
2457
2790
2633
2545
1336
1247
1184
1517
1173
1158
729
695
671
551
505
504

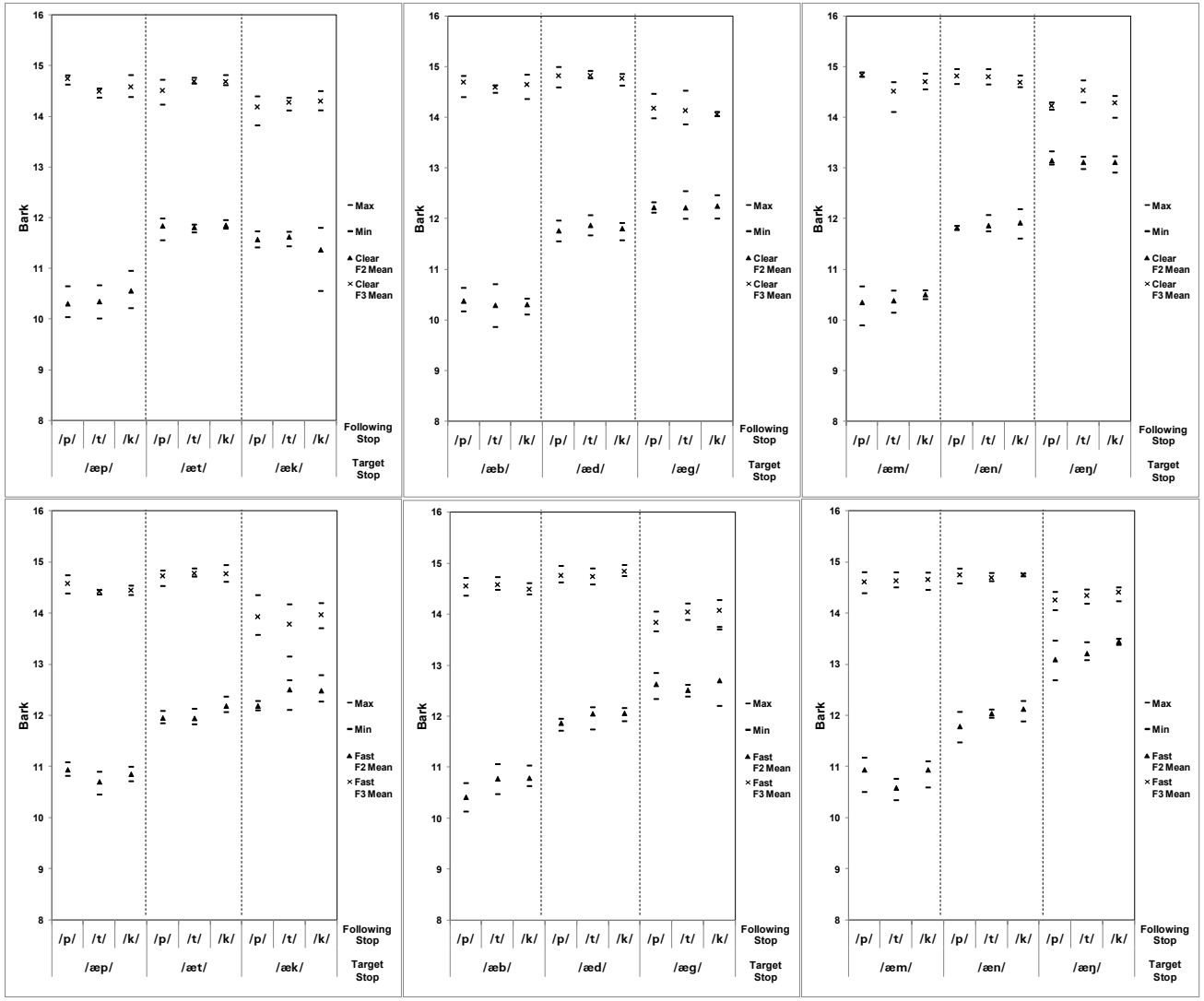
Appendix E1

F2 & F3 Ranges of Preceding Vowel /ɪ/ in Bark at Offset Point Sorted by Target Place & Following Place



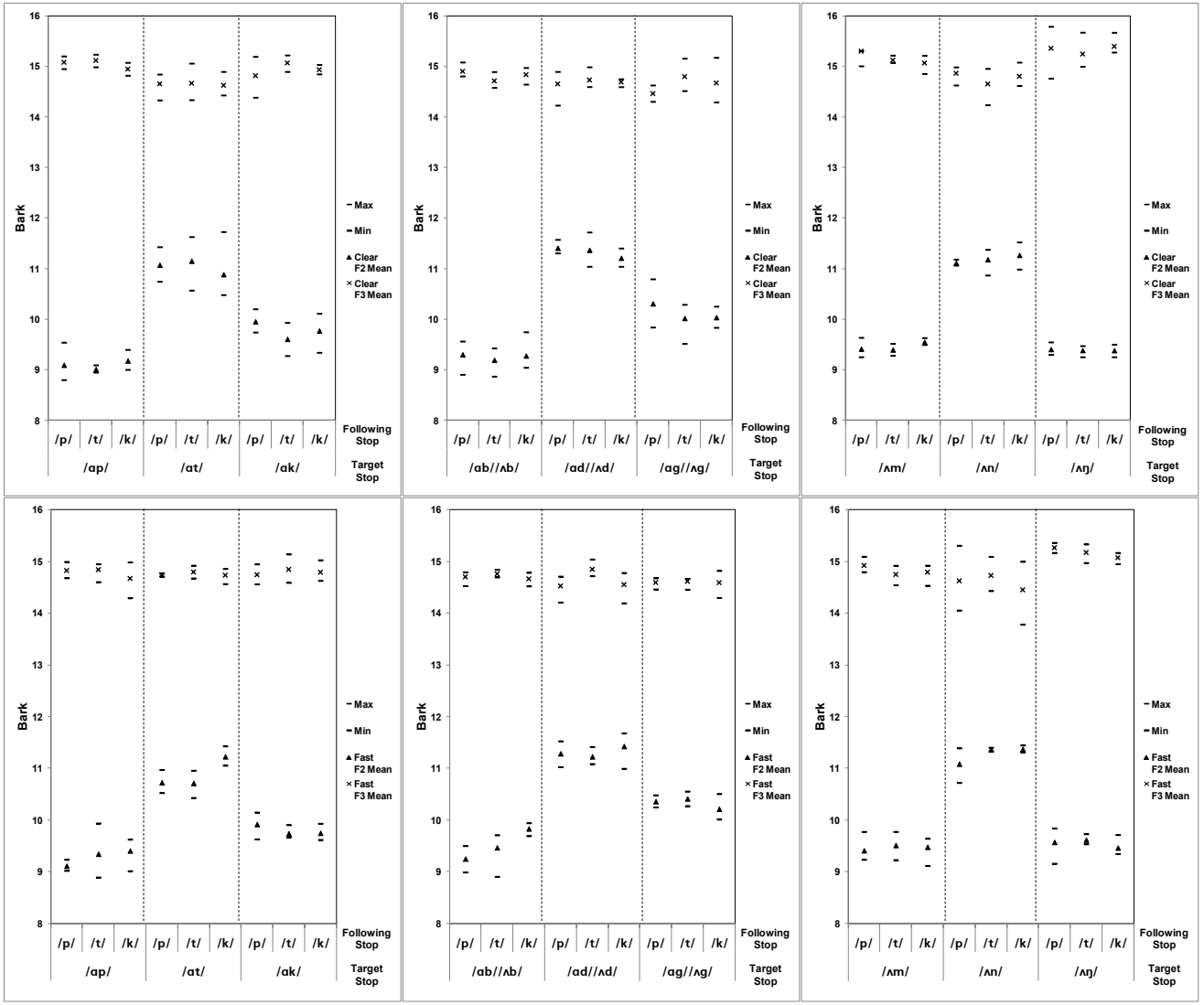
Appendix E2

F2 & F3 Ranges of Preceding Vowel /æ/ in Bark at Offset Point Sorted by Target Place & Following Place



Appendix E3

F2 & F3 Ranges of Preceding Vowel /ɑ/ & /ʌ/ in Bark at Offset Point Sorted by Target Place & Following Place



Appendix F

Durational and Amplitude Information of Critical Area Sorted by Following Context (Duration in ms)

				Clear Speech						Fast Speech					
Contrast Type	Target Place	Following Place		Word Final Closure	Release Duration	Word Initial Closure	Release dBSPL	Release RMS X duration	Release Peak dBSPL	Word Final Closure	Release Duration	Word Initial Closure	Release dBSPL	Release RMS X duration	Release Peak dBSPL
Voiceless	Labial	Different (N=24)	Mean	86.8	17.4	69.1	39.2	755.5	61.8	70.7	1.9	49.5	17.0	56.6	35.4
			Median	84.5	15.5	68.0	42.7	663.9	62.3	65.0	2.0	50.0	22.5	44.0	49.5
			SD	19.8	12.7	12.4	7.9	613.6	5.3	22.3	2.5	10.7	13.0	101.3	25.8
			Min	50.0	2.0	54.0	22.4	48.0	49.5	36.0	0.0	31.0	0.0	0.0	0.0
			Max	137.0	45.0	93.0	49.3	2218.5	73.0	134.0	12.0	67.0	42.1	505.2	62.8
		Same (N=12)	Mean	94.4	0.0	94.4	0.0	0.0	0.0	70.3	0.0	70.3	0.0	0.0	0.0
			Median	92.0	0.0	92.0	0.0	0.0	0.0	65.0	0.0	65.0	0.0	0.0	0.0
			SD	19.7	0.0	19.7	0.0	0.0	0.0	14.6	0.0	14.6	0.0	0.0	0.0
			Min	66.0	0.0	66.0	0.0	0.0	0.0	43.0	0.0	43.0	0.0	0.0	0.0
			Max	139.0	0.0	139.0	0.0	0.0	0.0	95.0	0.0	95.0	0.0	0.0	0.0
	Alveolar	Different (N=24)	Mean	70.2	32.7	81.5	41.7	1499.8	60.5	58.0	1.1	61.4	11.0	38.5	21.1
			Median	66.0	34.5	79.0	46.0	1584.0	65.8	55.0	0.0	59.5	0.0	0.0	0.0
			SD	17.7	16.5	21.5	13.2	765.1	18.7	13.2	1.8	13.6	14.8	71.0	27.8
			Min	47.0	0.0	55.0	0.0	0.0	0.0	28.0	0.0	35.0	0.0	0.0	0.0
			Max	102.0	69.0	152.0	49.9	3036.0	69.5	87.0	8.0	90.0	38.0	304.0	61.1
		Same (N=12)	Mean	78.2	0.0	78.2	0.0	0.0	0.0	54.5	0.3	54.6	3.1	4.5	7.6
			Median	74.0	0.0	74.0	0.0	0.0	0.0	55.5	0.0	55.5	0.0	0.0	0.0
			SD	18.6	0.0	18.6	0.0	0.0	0.0	14.5	0.6	14.2	7.1	11.2	17.9
			Min	53.0	0.0	53.0	0.0	0.0	0.0	29.0	0.0	29.0	0.0	0.0	0.0
			Max	110.0	0.0	110.0	0.0	0.0	0.0	76.0	2.0	76.0	18.9	35.6	46.8
	Velar	Different (N=24)	Mean	68.2	35.6	86.8	46.6	1657.2	67.9	64.6	0.9	67.5	7.7	30.6	13.9
			Median	69.0	34.5	90.0	46.6	1675.9	69.1	64.0	0.0	65.0	0.0	0.0	0.0
			SD	11.7	8.7	12.6	3.7	404.1	3.7	15.2	2.0	16.1	14.1	80.8	24.7
			Min	45.0	22.0	59.0	40.0	880.0	60.8	37.0	0.0	40.0	0.0	0.0	0.0
			Max	92.0	53.0	107.0	53.9	2442.9	74.3	103.0	9.0	103.0	41.3	371.7	62.9
		Same (N=12)	Mean	83.1	10.3	88.9	10.4	455.3	15.8	66.3	0.0	66.3	0.0	0.0	0.0
			Median	82.5	0.0	87.5	0.0	0.0	0.0	70.0	0.0	70.0	0.0	0.0	0.0
			SD	13.8	22.7	9.6	19.0	1017.6	28.7	11.4	0.0	11.4	0.0	0.0	0.0
			Min	65.0	0.0	72.0	0.0	0.0	0.0	49.0	0.0	49.0	0.0	0.0	0.0
			Max	105.0	63.0	105.0	46.7	2734.2	69.0	81.0	0.0	81.0	0.0	0.0	0.0
Voiced	Labial	Different (N=24)	Mean	72.7	18.0	62.2	40.2	807.6	62.6	72.3	3.0	45.4	23.0	101.0	43.5
			Median	71.5	16.0	66.0	42.2	668.4	63.5	72.0	3.0	47.0	28.5	63.5	51.5
			SD	14.0	12.6	11.2	8.8	654.4	5.6	16.0	2.9	11.3	13.4	107.4	23.4
			Min	47.0	2.0	31.0	17.9	52.2	48.2	47.0	0.0	22.0	0.0	0.0	0.0
			Max	104.0	46.0	77.0	52.2	2401.2	70.0	116.0	13.0	64.0	39.9	408.2	67.6
		Same (N=12)	Mean	99.7	0.0	65.4	0.0	0.0	0.0	85.3	0.0	57.0	0.0	0.0	0.0
			Median	97.0	0.0	67.0	0.0	0.0	0.0	90.0	0.0	60.5	0.0	0.0	0.0
			SD	23.3	0.0	13.6	0.0	0.0	0.0	17.9	0.0	9.9	0.0	0.0	0.0
			Min	54.0	0.0	31.0	0.0	0.0	0.0	63.0	0.0	41.0	0.0	0.0	0.0
			Max	136.0	0.0	81.0	0.0	0.0	0.0	118.0	0.0	73.0	0.0	0.0	0.0
	Alveolar	Different (N=24)	Mean	65.8	18.1	67.3	41.0	876.0	59.0	64.8	1.1	50.7	11.6	34.3	21.5
			Median	57.0	18.0	66.0	47.1	858.0	66.5	64.5	0.0	50.0	0.0	0.0	0.0
			SD	22.2	11.6	12.7	16.6	602.6	23.2	14.0	1.5	14.2	15.9	50.7	28.6
			Min	34.0	0.0	43.0	0.0	0.0	0.0	38.0	0.0	26.0	0.0	0.0	0.0
			Max	136.0	48.0	99.0	53.7	2457.6	74.9	105.0	4.0	95.0	42.4	169.6	68.0
		Same (N=12)	Mean	76.6	3.2	53.3	4.1	154.9	5.7	71.7	0.3	44.8	2.5	9.8	4.6
			Median	75.0	0.0	52.0	0.0	0.0	0.0	69.0	0.0	44.0	0.0	0.0	0.0
			SD	15.0	11.0	11.3	14.1	536.4	19.6	29.1	1.2	10.8	8.5	34.1	16.1
			Min	57.0	0.0	36.0	0.0	0.0	0.0	30.0	0.0	30.0	0.0	0.0	0.0
			Max	111.0	38.0	72.0	48.9	1858.2	67.8	142.0	4.0	62.0	29.5	118.0	55.6
	Velar	Different (N=24)	Mean	57.1	24.2	78.3	43.2	1146.6	61.8	59.5	1.0	54.2	9.0	41.2	15.0
			Median	54.0	26.0	79.0	45.7	1167.8	66.7	60.0	0.0	56.0	0.0	0.0	0.0
			SD	10.4	8.8	10.4	13.8	446.6	19.4	18.0	2.1	13.7	16.3	94.6	26.7
			Min	43.0	0.0	56.0	0.0	0.0	0.0	30.0	0.0	30.0	0.0	0.0	0.0
			Max	79.0	38.0	97.0	53.4	1789.8	75.7	99.0	9.0	77.0	46.0	414.0	66.3
		Same (N=12)	Mean	84.0	0.4	71.0	4.0	10.3	8.5	73.9	0.2	57.1	3.0	5.9	5.3
			Median	88.5	0.0	61.0	0.0	0.0	0.0	77.0	0.0	57.0	0.0	0.0	0.0
			SD	22.0	1.0	16.9	9.5	26.1	19.9	21.2	0.6	17.1	10.2	20.4	18.4
			Min	33.0	0.0	53.0	0.0	0.0	0.0	37.0	0.0	35.0	0.0	0.0	0.0
			Max	107.0	3.0	105.0	28.5	85.5	52.3	110.0	2.0	95.0	35.4	70.8	63.7

Appendix G

Criteria of Template Matching Analysis of Stop Bursts

1. Criteria for the labial (*Diffuse-falling*) template:
    1) a peak between 1200 and 3500 Hz is fitted to the top reference line and all other peaks above this frequency have to lie below the line.
    2) at least two peaks must fall within the reference lines that are about 10 dB apart, one peak falling below 2400 Hz and the other peak falling in the range of 2400 and 3600 Hz.
    3) the spectrum shape must be either falling or relatively flat, and spectral energy must occur at low to mid frequencies (1200-3500 Hz).
2. Criteria for the alveolar (*Diffuse-rising*) template:
    1) a peak above 2200 Hz touches an upward sloping reference line and all other peaks above this frequency have to lie below the line.
    2) at least two peaks must fall between the two reference lines that are about 10 dB apart.
    3) a peak falling above 2200 Hz must have higher amplitude than the other lower frequency peak.
    4) the spectrum shape has a tilting slope upwards with no one peak dominating the entire spectrum.
3. Criteria for the velar (*Compact*) template:
    1) there is a peak between 1200 and 3500 Hz projects through the reference line.
    2) there is no other peak of the same or greater amplitude occurring either below 1200 Hz or above 3500 Hz.

Appendix H

Percentage of Correct Responses by Japanese and AE Listeners

(Overall and Contrast Type)

Language Group	Speech Mode		Target Stimuli				Filler
			Overall	Voiceless	Voiced	Nasal	
AE	Clear Speech (N = 12)	Mean	97.8%	96.7%	98.5%	98.3%	99.8%
		SD	1.6%	1.9%	2.1%	1.3%	0.5%
		Median	98.5%	96.8%	99.5%	98.6%	100.0%
		Semi-interquartile	1.4%	1.5%	1.5%	0.9%	0.1%
	Fast Speech (N = 24)	Mean	94.8%	90.4%	95.9%	97.9%	99.7%
		SD	2.6%	5.0%	3.1%	1.8%	0.5%
		Median	94.9%	90.7%	96.8%	98.1%	100.0%
		Semi-interquartile	1.7%	2.4%	2.0%	0.9%	0.3%
Japanese	Clear Speech (N = 24)	Mean	88.5%	94.1%	94.0%	77.5%	99.2%
		SD	4.3%	2.9%	4.0%	8.3%	1.0%
		Median	88.9%	94.4%	94.9%	79.2%	99.3%
		Semi-interquartile	2.5%	1.3%	2.9%	5.0%	0.7%
	Fast Speech (N =24)	Mean	71.1%	67.4%	79.7%	66.2%	98.8%
		SD	10.2%	7.4%	11.7%	14.1%	1.4%
		Median	70.1%	66.7%	82.4%	65.3%	99.3%
		Semi-interquartile	7.6%	4.7%	8.1%	11.3%	0.8%

Appendix I

Percentage of Correct Responses by Japanese and AE Listeners (Target Place)

Language Group	Speech Mode		Voiceless			Voiced			Nasal		
			Labial /p/	Alveolar /t/	Velar /k/	Labial /b/	Alveolar /d/	Velar /g/	Labial /m/	Alveolar /n/	Velar /ŋ/
AE	Clear	Mean	97.9%	98.4%	93.8%	99.1%	98.4%	97.9%	97.9%	97.2%	99.8%
		SD	3.4%	2.2%	3.8%	2.5%	2.8%	2.9%	2.4%	2.6%	0.8%
		Median	100.0%	100.0%	94.4%	100.0%	100.0%	100.0%	97.2%	97.2%	100.0%
		SemiInter Quartile	1.7%	1.4%	2.8%	0.0%	1.4%	1.7%	1.4%	1.7%	0.0%
	Fast	Mean	92.2%	98.5%	80.6%	99.5%	93.4%	94.9%	95.9%	97.8%	99.9%
		SD	4.8%	2.6%	11.2%	1.3%	6.6%	4.3%	3.5%	2.8%	0.6%
		Median	91.7%	100.0%	80.6%	100.0%	94.4%	94.4%	95.8%	98.6%	100.0%
		SemiInter Quartile	3.1%	1.4%	6.9%	0.0%	3.8%	3.1%	1.7%	1.4%	0.0%
Japanese	Clear	Mean	96.5%	92.4%	93.4%	95.9%	88.9%	97.2%	88.9%	64.4%	79.2%
		SD	3.9%	5.7%	3.5%	5.4%	7.7%	2.5%	9.9%	21.1%	12.9%
		Median	97.2%	93.1%	94.4%	97.2%	88.9%	97.2%	91.7%	65.3%	80.6%
		SemiInter Quartile	2.8%	4.5%	1.7%	2.8%	5.6%	2.8%	6.9%	12.8%	10.1%
	Fast	Mean	77.5%	70.8%	53.9%	88.0%	67.1%	84.0%	70.5%	52.2%	76.0%
		SD	10.0%	16.5%	14.9%	13.0%	27.4%	9.2%	21.0%	19.7%	16.3%
		Median	77.8%	68.1%	54.2%	94.4%	75.0%	86.1%	75.0%	48.6%	77.8%
		SemiInter Quartile	6.3%	9.7%	10.4%	9.7%	18.8%	7.3%	11.1%	18.4%	12.8%

Appendix J

Percentage of Correct Responses by Japanese and AE Listeners (Following Place)

			Voiceless						Voiced						Nasal					
			Labial		Alveolar		Velar		Labial		Alveolar		Velar		Labial		Alveolar		Velar	
			Different	Same	Different	Same	Different	Same	Different	Same	Different	Same	Different	Same	Different	Same	Different	Same	Different	Same
AE	Clear	mean	98.3%	97.2%	98.6%	97.9%	99.3%	82.6%	99.3%	98.6%	98.3%	98.6%	98.6%	96.5%	97.9%	97.9%	96.9%	97.9%	100.0%	99.3%
		SD	3.3%	4.1%	2.7%	3.8%	1.6%	10.9%	1.6%	4.8%	3.3%	3.2%	2.1%	6.6%	2.2%	5.2%	3.1%	3.8%	0.0%	2.4%
		median	100.0%	100.0%	100.0%	100.0%	100.0%	83.3%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	97.9%	100.0%	95.8%	100.0%	100.0%	100.0%
		Semi-interquartile	0.5%	4.2%	0.5%	1.0%	0.0%	8.3%	0.0%	0.0%	0.5%	0.0%	2.1%	1.0%	2.1%	0.0%	2.1%	1.0%	0.0%	0.0%
	Fast	mean	97.9%	80.9%	97.9%	99.7%	84.2%	73.3%	99.5%	99.7%	92.5%	95.1%	94.4%	95.8%	95.3%	97.2%	97.0%	99.3%	100.0%	99.7%
		SD	3.0%	12.9%	3.7%	1.7%	11.8%	16.3%	1.9%	1.7%	7.3%	7.3%	5.0%	4.9%	4.3%	4.0%	3.8%	2.4%	0.0%	1.7%
		median	100.0%	79.2%	100.0%	100.0%	85.4%	75.0%	100.0%	100.0%	95.8%	100.0%	95.8%	100.0%	95.8%	100.0%	100.0%	100.0%	100.0%	100.0%
		Semi-interquartile	2.1%	8.3%	2.1%	0.0%	6.8%	10.4%	0.0%	0.0%	4.7%	4.2%	4.2%	4.2%	4.2%	4.2%	2.1%	0.0%	0.0%	0.0%
Japanese	Clear	mean	98.8%	92.0%	94.1%	88.9%	99.7%	80.9%	97.6%	92.7%	88.0%	90.6%	98.3%	95.1%	92.9%	80.9%	63.7%	65.6%	78.6%	80.2%
		SD	2.3%	9.4%	6.4%	10.3%	1.2%	10.3%	4.1%	9.0%	8.8%	7.9%	2.7%	6.9%	8.8%	14.8%	22.1%	21.7%	14.9%	13.9%
		median	100.0%	91.7%	95.8%	91.7%	100.0%	83.3%	100.0%	95.8%	87.5%	91.7%	100.0%	100.0%	95.8%	83.3%	62.5%	70.8%	79.2%	83.3%
		Semi-interquartile	0.5%	5.2%	4.2%	8.3%	0.0%	5.2%	2.1%	5.2%	4.7%	8.3%	2.1%	4.2%	6.3%	9.4%	17.2%	17.7%	8.3%	9.4%
	Fast	mean	80.9%	70.8%	63.5%	85.4%	50.0%	61.8%	88.5%	86.8%	64.6%	72.2%	80.6%	91.0%	78.5%	54.5%	51.4%	53.8%	76.7%	74.7%
		SD	9.9%	13.5%	19.2%	14.2%	15.4%	18.0%	12.8%	14.5%	28.0%	29.4%	10.6%	9.5%	21.5%	24.6%	18.7%	25.3%	17.6%	18.6%
		median	83.3%	66.7%	62.5%	83.3%	50.0%	58.3%	93.8%	91.7%	70.8%	79.2%	83.3%	91.7%	87.5%	58.3%	52.1%	50.0%	83.3%	83.3%
		Semi-interquartile	6.8%	9.4%	15.1%	12.5%	10.4%	12.5%	10.4%	12.5%	20.8%	25.0%	8.9%	4.2%	10.4%	14.6%	17.2%	20.8%	13.5%	16.7%

Appendix K

Performance Differences between Language Groups (AE *vs*. Japanese)

(Mann-Whitney *U* Test)

Speech Mode	Contrast Type	*U*	*z*	Sig.	*r*
Clear Speech (Japanese: N = 24 AE: N = 12)	Overall	3	-4.74	p < 0.001**	0.79
	Voiceless	63	-2.74	p = 0.006*	0.46
	Voiced	48	-3.25	p = 0.001**	0.54
	Nasal	0	-4.84	p < 0.001**	0.81
Fast Speech (Japanese: N = 24 AE: N = 24)	Overall	1	-5.92	p < 0.001**	0.85
	Voiceless	5	-5.84	p < 0.001**	0.84
	Voiced	49	-4.94	p < 0.001**	0.71
	Nasal	0	-5.97	p < 0.001**	0.86

Bonferroni adjustment for Contrast type comparisons: $\beta = 0.0167$ for $\alpha = 0.05$; $\beta = 0.0033$ for $\alpha = 0.01$. ** indicates $p < 0.0033$, and * indicates $p < 0.0167$.

Appendix L

Performance difference between Japanese and AE by Target Place in Clear and Fast Speech

(Mann-Whitney U Test)

Speech Mode	Contrast Type	Target Place	*U* value	*z*	Sig.	*r*
Clear (Japanese: N = 24 AE: N = 12)	Voiceless	Labial /p/	108.5	-1.27	$p = 0.20$	0.21
		Alveolar /t/	52	-3.16	$p < 0.001$**	0.53
		Velar /k/	132	-0.42	$p = 0.68$	0.07
	Voiced	Labial /b/	88	-2.11	$p = 0.04$	0.35
		Alveolar /d/	32.5	-3.80	$p = 0.003$*	0.77
		Velar /g/	116.5	-0.98	$p = 0.33$	0.16
	Nasal	Labial /m/	57.5	-2.96	$p = 0.003$*	0.49
		Alveolar /n/	7.5	-4.60	$p < 0.001$**	0.77
		Velar /ŋ/	2	-4.85	$p < 0.001$**	0.81
Fast (Japanese: N = 24 AE: N = 24)	Voiceless	Labial /p/	46	-5.03	$p < 0.001$**	0.73
		Alveolar /t/	20.5	-5.64	$p < 0.001$**	0.81
		Velar /k/	43.5	-5.05	$p < 0.001$**	0.73
	Voiced	Labial /b/	80.5	-4.68	$p < 0.001$**	0.68
		Alveolar /d/	91.5	-4.08	$p < 0.001$**	0.59
		Velar /g/	81.5	-4.30	$p < 0.001$**	0.62
	Nasal	Labial /m/	34	-5.30	$p < 0.001$**	0.76
		Alveolar /n/	0	-6.00	$p < 0.001$**	0.87
		Velar /ŋ/	13	-6.07	$p < 0.001$**	0.88

Bonferroni adjustment for target place comparisons for each contrast type: $\beta = 0.0056$ for $\alpha = 0.05$; $\beta = 0.0011$ for $\alpha = 0.01$. ** indicates $p < 0.0011$, and * indicates $p < 0.0056$.

Appendix M

Description of Target Place Comparisons by AE and Japanese Listeners (Results 3.1.5.)

The statistical analyses adopted in this section followed those adopted in the Contrast type comparison section and were carried out for each target place comparison within each contrast type: a Friedman's test and three separate Wilcoxon Signed-Rank tests for pair-wise comparisons when the Friedman's result was significant, for AE group; and a one-way repeated measures ANOVA to see the main effect of target place and three separate repeated measures ANOVAs for pair-wise comparisons when the main effect was seen, for Japanese group. For the pair-wise comparisons, Bonferroni adjusted alpha levels of .0167 per test (0.05/3) was adopted for both AE and Japanese group. For the values of mean percent correct accuracies with SDs, medians, and semi-interquartile ranges for each language group, see Appendix X5.

a. Performance by AE group

The distributions of response accuracy on each target place by AE listeners in both speech modes are presented as box plots in Figure X4 and as a bar chart in Figure X6. The detailed statistic data are presented in Appendix N.

For the analysis within Clear Speech, the Friedman's tests revealed that the performance on Voiceless (word-final /p/-/t/-/k/) and Nasal Contrasts (word-final /m/-/n/-/ŋ/) by AE listeners differed significantly by target place. The Wilcoxon tests found a significant difference in performance only between Alveolar and Velar for both Voiceless and Nasal Contrasts (but opposite directions), showing the following relationship of performance: /p/ ≈ /k/ < /t/ ≈ /p/; and /m/ ≈ /n/ < /ŋ/ ≈ /m/. Again, the significant difference between /k/ and /t/ disappeared after disregarding the performance on the most error-prone token *hock cautiously* (z = -1.73, p = 0.08, r = 0.35). The Friedman's test on Voiced Contrasts in Clear Speech by AE did not show a significant difference across target places, and no further tests were carried out.

For the comparisons within Fast Speech, the Friedman's test showed that the AE performance on Voiceless, Voiced and Nasal Contrasts in Fast Speech differed significantly by target place. The Wilcoxon tests further revealed that the relationship of performance as follows: /k/ < /p/ < /t/; /d/ ≈ /g/ < /b/; and /m/ ≈ /n/ < /ŋ/.

b. Performance by Japanese group

For the comparisons of the Japanese performance on different target places in each contrast type, see the bar chart in Figure X6 and also the box plots in Figure X4.

Japanese performance in Clear speech

The detailed statistic data examining the effect of target place on Japanese performance in Clear Speech are presented in Appendix O. The repeated measures ANOVA found that the effect of target place was significant for Voiceless Contrasts (word-final /p/-/t/-/k/). The further analyses revealed that their performance on Labial /p/ was significantly better than that on Alveolar /t/ and also better than that on Velar /k/, but their performance on /t/ and on /k/ did not differ significantly, showing the following relationship of performance: /t/ ≈ /k/ < /p/. The results indicate that the Japanese group's difference in performance across target places seen in Voiceless Contrasts in Clear Speech was due to the better performance on Labial than those on the other two (Alveolar and Velar) target places.

The ANOVAs revealed a significant effect of target place for Voiced Contrasts (word-final /b/-/d/-/g/) as well. The Japanese performance on Alveolar /d/ was significantly poorer than

that on Labial /b/ and also poorer than that on Velar /g/. Their performance on /b/ and on /g/, however, did not differ significantly, showing the following relationship of performance: /d/ < /b/ ≈ /g/. The results indicate that the Japanese group's difference in performance by target place observed in Voiced Contrasts in Clear Speech was attributable to the poorer performance on Alveolar than those on the other two (Labial and Velar) target places.

A significant effect of target place on Japanese performance was also seen in Nasal Contrasts (word-final /m/-/n/-/ŋ/) in Clear Speech, from the results of the ANOVAs. The Japanese performance on Alveolar /n/ was significantly worse than that on Labial /m/ and also worse than that on Velar /ŋ/. Furthermore, their performance on /m/ was significantly better than that on /ŋ/, showing the following relationship of performance: /n/ < /ŋ/ < /m/.

Performance in Fast speech

Appendix P presents the statistic data examining the effect of target place on Japanese performance in Fast Speech. The repeated measures ANOVA revealed that Japanese performance on Voiceless Contrast in Fast Speech differed significantly by target place. Their performance on Velar was significantly poorer than that on Labial /p/ and also poorer than that on Alveolar /t/, but their performance on /p/ and on /t/ did not differ significantly, showing the following relationship of performance: /k/ < /p/ ≈ /t/. The results indicate that the Japanese group's difference in performance across target places observed in Voiceless Contrasts in Fast Speech was due to the poorer performance on Velar than those on the other two (Labial and Alveolar) target places.

The results of the ANOVAs also revealed a significant effect of target place on the Japanese performance for Voiced Contrasts in Fast Speech. Their performance on Alveolar /d/ was significantly poorer than that on Labial /b/ and also poorer than that on Velar /g/. Their performance on /b/ and on /g/, on the other hand, did not differ significantly, showing the following relationship of performance: /d/ < /g/ ≈ /b/. The results indicate that the Japanese group's performance difference by target place seen in Voiced Contrasts in Fast Speech was attributable to the poorer performance on Alveolar than those on the other two (Labial and Velar) target places.

A significant effect of target place for Japanese performance on Nasal contrasts in Fast Speech was revealed by the ANOVAs as well. Their performance on Alveolar /n/ was significantly worse than that on Labial /m/ and also worse than that on Velar /ŋ/. Their performance on /m/ and on /ŋ/, however, did not differ significantly, showing the following relationship of performance: /n/ < /m/ ≈ /ŋ/. The results indicate that the effect of target place seen in Japanese group's performance in Nasal Contrast in Fast Speech was attributable to the poorer performance on Alveolar than those on the other two (Labial and Velar) target places.

Appendix N

Performance difference between Target Places by AE Listeners
(Friedman's & Wilcoxon Signed Ranks Tests)

Speech Mode	Friedman's				Wilcoxon Signed Ranks			
		df	χ^2	Sig.		*z*	Sig	*r*
Clear Speech (N = 12)	Voiceless	2	10.05	$p < 0.01$**	Labial *vs*. Alveolar (*Md*, 100%, 100%)	-0.32	$p = 0.75$	0.06
					Labial *vs*. Velar (*Md*, 100%, 94.4%)	-2.05	$p = 0.04$	0.42
					Alveolar *vs*. Velar (*Md*, 100%, 94.4%)	-2.72	$p = 0.006$*	0.56
	Voiced	2	2.95	$p = 0.23$	Not tested			
	Nasal	2	8.18	$p < 0.05$*	Labial *vs*. Alveolar (*Md*, 97.2%, 97.2%)	-0.75	$p = 0.45$	0.15
					Labial *vs*. Velar (*Md*, 97.2%, 100%)	-2.11	$p = 0.04$	0.43
					Alveolar *vs*. Velar (*Md*, 97.2%, 100%)	-2.41	$p = 0.016$*	0.49
Fast Speech (N = 24)	Voiceless	2	37.84	$p < 0.01$**	Labial *vs*. Alveolar (*Md*, 91.7%, 100%)	-3.94	$p < 0.001$**	0.57
					Labial *vs*. Velar (*Md*, 100%, 80.6%)	-3.67	$p < 0.001$**	0.53
					Alveolar *vs*. Velar (*Md*, 100%, 80.6%)	-4.20	$p < 0.001$**	0.61
	Voiced	2	21.26	$p < 0.01$**	Labial *vs*. Alveolar (*Md*, 100%, 94.4%)	-3.61	$p < 0.001$**	0.52
					Labial *vs*. Velar (*Md*, 100%, 94.4%)	-3.71	$p < 0.001$**	0.54
					Alveolar *vs*. Velar (*Md*, 94.4%, 94.4%)	-0.88	$p = 0.38$	0.13
	Nasal	2	25.02	$p < 0.01$**	Labial *vs*. Alveolar (*Md*, 95.8%, 98.6%)	-2.16	$p = 0.031$	0.31
					Labial *vs*. Velar (*Md*, 95.8%, 100%)	-3.72	$p < 0.001$**	0.54
					Alveolar *vs*. Velar (*Md*, 98.6%, 100%)	-3.14	$p = 0.002$**	0.45

Bonferroni adjustment applied to Wilcoxon Signed Ranks tests: $\beta = 0.0167$ for $\alpha = 0.05$; $\beta = 0.0033$ for $\alpha = 0.01$. ** indicates $p < 0.0033$, and * indicates $p < 0.0167$.

Appendix O

Performance Difference between Target Places by Japanese Listeners in Clear Speech (Repeated Measures ANOVAs)

Contrast Type			df	F	Sig	η_p^2
Voiceless	Target Place Main Effect		2, 46	6.51	$p = 0.003$**	0.220
	Pair-wise Comparison	Labial *vs*. Alveolar	1, 23	11.72	$p = 0.002$**	0.337
		Labial *vs*. Velar	1, 23	11.16	$p = 0.003$**	0.327
		Alveolar *vs*. Velar	1, 23	0.55	$p = 0.467$	0.023
Voiced	Target Place Main Effect		2, 46	21.52	$p < 0.001$**	0.483
	Pair-wise Comparison	Labial *vs*. Alveolar	1, 23	20.27	$p < 0.001$**	0.468
		Labial *vs*. Velar	1, 23	1.61	$p = 0.217$	0.065
		Alveolar *vs*. Velar	1, 23	32.26	$p < 0.001$**	0.584
Nasal	Target Place Main Effect		1.27, 29.28	14.54	$p < 0.001$**	0.387
	Pair-wise Comparison	Labial *vs*. Alveolar	1, 23	28.77	$p < 0.001$**	0.556
		Labial *vs*. Velar	1, 23	12.95	$p = 0.002$**	0.360
		Alveolar *vs*. Velar	1, 23	6.31	$p = 0.019$	0.215

Bonferroni adjustment for pair-wise comparisons: $\beta = 0.0167$ for $\alpha = 0.05$; $\beta = 0.0033$ for $\alpha = 0.01$. ** indicates $p < 0.0033$, and * indicates $p < 0.0167$.

Appendix P

Performance Difference between Target Places by Japanese Listeners in Fast Speech (Repeated Measures ANOVAs)

Contrast Type			df	F	Sig	η_p^2
Voiceless	Target Place Main Effect		2, 46	16.68	$p < 0.001$**	0.420
	Pair-wise Comparison	Labial *vs*. Alveolar	1, 23	2.78	$p = 0.109$	0.108
		Labial *vs*. Velar	1, 23	43.92	$p < 0.001$**	0.656
		Alveolar *vs*. Velar	1, 23	11.73	$p = 0.002$**	0.338
Voiced	Target Place Main Effect		1.39, 31.70	9.87	$p < 0.001$**	0.300
	Pair-wise Comparison	Labial *vs*. Alveolar	1, 23	13.36	$p < 0.001$**	0.367
		Labial *vs*. Velar	1, 23	1.90	$p = 0.182$	0.076
		Alveolar *vs*. Velar	1, 23	8.45	$p = 0.008$*	0.269
Nasal	Target Place Main Effect		2, 46	15.06	$p < 0.001$**	0.396
	Pair-wise Comparison	Labial *vs*. Alveolar	1, 23	13.80	$p = 0.001$**	0.375
		Labial *vs*. Velar	1, 23	2.34	$p = 0.140$	0.092
		Alveolar *vs*. Velar	1, 23	23.14	$p < 0.001$**	0.501

Bonferroni adjustment for pair-wise comparisons: $\beta = 0.0167$ for $\alpha = 0.05$; $\beta = 0.0033$ for $\alpha = 0.01$. ** indicates $p < 0.0033$, and * indicates $p < 0.0167$.

Appendix Q

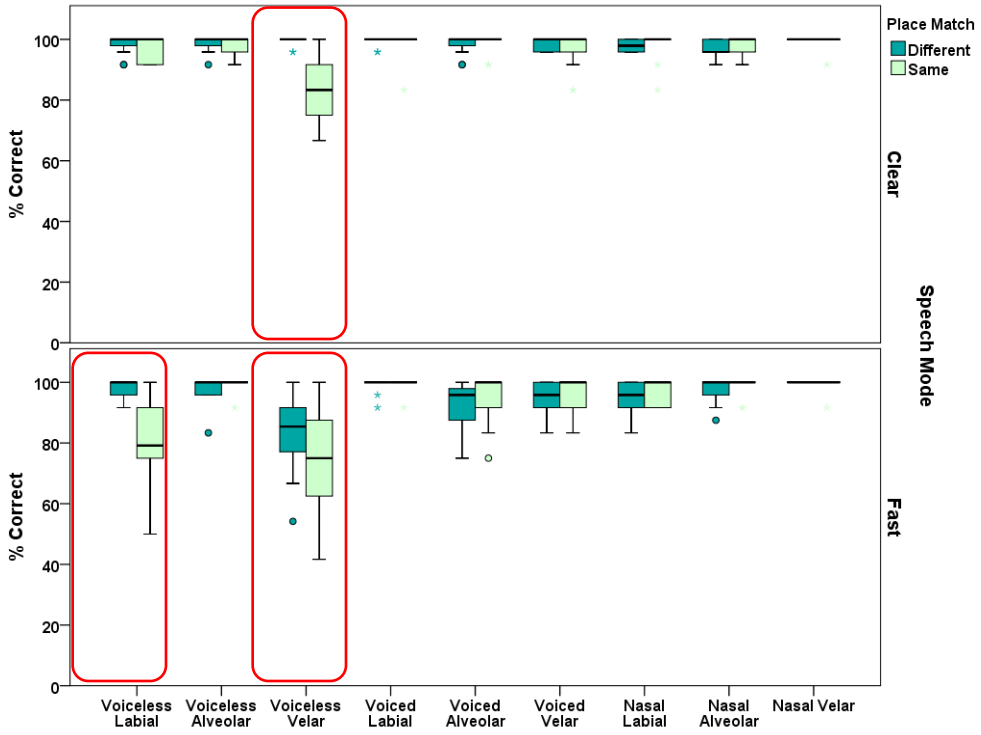
Description of Context Effects on AE performance (Results 3.1.6.)

The response accuracies on Different and Same contexts by the AE group illustrated as box plots are presented in Appendix R. The panels indicating significantly better performance on Different context are marked with solid lines. A set of three 2-tailed Wilcoxon Signed-Rank tests using Bonferroni adjusted alpha levels of 0.0056 per test (0.05/9) was carried out for each contrast type within each speech mode (i.e., a total of 18 tests). The statistic data are presented in Appendix S. The following places showing significantly better performance are indicated in bold face. Because AE performance in Clear Speech was at or near ceiling except for Voiceless Velar, no significant difference between Different and Same contexts was seen in Clear Speech except for Voiceless Velar showing better performance on Different than on Same context. However, the difference became non-significant after disregarding the aforementioned error-prone token *hock cautiously* ($z = -2.57$, $p = 0.01$, $r = 0.52$).

For Fast Speech, the effect of following place is seen only in Voiceless contrasts, with Different context better in Labial and Velar contrasts. No effect was seen in Voiced contrasts, and only Labial showed the effect with better performance on Same in Nasal contrasts.

Appendix R

Percent Correct Accuracy by AE on Following Place



Bonferroni adjustment applied: $\beta = 0.0056$ for $\alpha = 0.05$; $\beta = 0.0011$ for $\alpha = 0.01$.
Red solid lines indicate better performance seen in Different context.

Appendix S

Performance Difference between Following Place by AE Listeners
(Wilcoxon Signed Rank Tests)

Speech Mode	Contrast Type	Target Place	Following Place	*z*	Sig	*r*
Clear Speech (N = 12)	Voiceless	Labial	Different *vs*. Same (*Md* = 100, 100)	-1.34	$p = 0.180$	0.27
		Alveolar	Different *vs*. Same (*Md* = 100, 100)	-0.41	$p = 0.680$	0.08
		Velar	**Different** *vs*. Same (*Md* = 100, 83.3)	-3.00	$p = 0.003$*	0.61
	Voiced	Labial	Different *vs*. Same (*Md* = 100, 100)	-0.45	$p = 0.655$	0.09
		Alveolar	Different *vs*. Same (*Md* = 100, 100)	-0.27	$p = 0.785$	0.06
		Velar	Different *vs*. Same (*Md* = 100, 100)	-0.68	$p = 0.498$	0.14
	Nasal	Labial	Different *vs*. Same (*Md* = 97.9, 100)	-0.18	$p = 0.861$	0.04
		Alveolar	Different *vs*. Same (*Md* = 95.8, 100)	-1.12	$p = 0.262$	0.23
		Velar	Different *vs*. Same (*Md* = 100, 100)	-1.00	$p = 0.317$	0.20
Fast Speech (N = 24)	Voiceless	Labial	**Different** *vs*. Same (*Md* = 100, 79.2)	-3.89	$p < 0.001$**	0.56
		Alveolar	Different *vs*. Same (*Md* = 100, 100)	-2.68	$p = 0.007$	0.39
		Velar	**Different** *vs*. Same (*Md* = 85.4, 75.0)	-2.91	$p = 0.004$*	0.42
	Voiced	Labial	Different *vs*. Same (*Md* = 100, 100)	-0.27	$p = 0.785$	0.04
		Alveolar	Different *vs*. Same (*Md* = 95.8, 100)	-2.18	$p = 0.030$	0.31
		Velar	Different *vs*. Same (*Md* = 95.8, 100)	-1.55	$p = 0.120$	0.22
	Nasal	Labial	Different *vs*. Same (*Md* = 95.8, 100)	-2.47	$p < 0.014$	0.36
		Alveolar	Different *vs*. Same (*Md* = 100, 100)	-2.35	$p = 0.019$	0.34
		Velar	Different *vs*. Same (*Md* = 100, 100)	-1.00	$p = 0.317$	0.14

Bonferroni adjustment for pair-wise (Different *vs*. Same) comparisons: $\beta = 0.0056$ for $\alpha = 0.05$; $\beta = 0.0011$ for $\alpha = 0.01$. ** indicates $p < 0.0011$, and * indicates $p < 0.0056$.

Appendix T

Main Effect of Target Place and Following Place, and Their Interaction by Japanese Listeners (Repeated Measures ANOVAs)

Speech Mode	Contrast Type		df	*F*	Sig	η_p^2
Clear Speech (N = 24)	Voiceless	Target Place Main Effect	2, 46	5.70	$p = 0.006$**	0.20
		Following Place Main Effect	1, 23	109.73	$p < 0.001$**	0.83
		Target Place × Following Place Interaction	2, 46	10.42	$p < 0.001$**	0.31
	Voiced	Target Place Main Effect	2, 46	13.53	$p < 0.001$**	0.37
		Following Place Main Effect	1, 23	8.45	$p = 0.008$**	0.27
		Target Place × Following Place Interaction	2, 46	5.24	$p = 0.009$**	0.19
	Nasal	Target Place Main Effect	1.3, 29.5	11.82	$p < 0.001$**	0.34
		Following Place Main Effect	1, 23	3.82	$p = 0.063$	0.14
		Target Place × Following Place Interaction	2, 46	9.04	$p < 0.001$**	0.28
Fast Speech (N = 24)	Voiceless	Target Place Main Effect	2, 46	14.38	$p < 0.001$**	0.39
		Following Place Main Effect	1, 23	17.88	$p < 0.001$**	0.44
		Target Place × Following Place Interaction	2, 46	48.85	$p < 0.001$**	0.68
	Voiced	Target Place Main Effect	1.4, 31.7	8.97	$p = 0.002$**	0.28
		Following Place Main Effect	1, 23	16.81	$p < 0.001$**	0.42
		Target Place × Following Place Interaction	1.4, 31.7	7.10	$p = 0.007$**	0.24
	Nasal	Target Place Main Effect	2, 46	13.04	$p < 0.001$**	0.36
		Following Place Main Effect	1, 23	15.78	$p = 0.001$**	0.41
		Target Place × Following Place Interaction	2, 46	17.31	$p < 0.001$**	0.43

Appendix U

Error Responses by AE and Japanese Listeners Sorted by Target Place

AE Group

Contrast	Target	Error Response in Clear Speech			Z value	Error Response in Fast Speech			Z value
Type	Place	Labial	Alveolar	Velar		Labial	Alveolar	Velar	
	Labial		6	3	1		**67****		8.18535
Voiceless	Alveolar	1		6	1.88982	5		8	0.83205
	Velar	6	**21****		2.88675	38	**130****		7.09795
	Labial		2	2	0		3	1	1
Voiced	Alveolar	6		1	1.88982	**45****		12	4.37096
	Velar	5	4		0.33333	10	**34****		3.61814
	Labial		**9****		3		**31****	4	4.56383
Nasal	Alveolar	9		3	1.73205	9		10	0.22942
	Velar		1		1		1		1

Japanese Group

Contrast	Target	Error Response in Clear Speech			Z value	Error Response in Fast Speech			Z value
Type	Place	Labial	Alveolar	Velar		Labial	Alveolar	Velar	
	Labial		**26****	4	4.01663		**162****	32	9.33346
Voiceless	Alveolar	27		39	1.4771	**143***		109	2.1418
	Velar	9	**48****		5.16568	132	**266****		6.71681
	Labial		**26****	9	2.87352		**67****	37	2.94174
Voiced	Alveolar	**64****		32	3.26599	122		**162****	2.37356
	Velar	9	15		1.22474	36	**102****		5.61829
	Labial		53	43	1.02062		137	118	1.18983
Nasal	Alveolar	86		**222****	7.74932	94		**319****	11.0715
	Velar	11	**169****		11.7766	40	**167****		8.82711

** = $p < .01$, * = $p < .05$ by Binominal Test

Appendix V

Error Responses by AE and Japanese Listeners Sorted by Following Place

Voiceless Contrasts

Language Group	Target Place	Following Place	Error Response in Clear Speech			Error Response in Fast Speech		
			Labial	Alveolar	Velar	Labial	Alveolar	Velar
AE	Labial	Labial		3	1		**55****	
		Alveolar		2	1		**11****	
		Velar		1	1		1	
	Alveolar	Labial			2	4		
		Alveolar	1		2	1		
		Velar			2			**8***
	Velar	Labial		1		5	**43****	
		Alveolar	1			14	**29***	
		Velar	5	**20****		19	**58****	
Japanese	Labial	Labial		**21****	2		**78****	6
		Alveolar		4	2		**54****	12
		Velar		1			**30***	14
	Alveolar	Labial	**15***		6	**85****		16
		Alveolar	9		**23***	24		18
		Velar	3		10	34		**75****
	Velar	Labial				54	**77***	
		Alveolar	1	1		43	**114****	
		Velar	8	**47****		35	**75****	

Voiced Contrasts

Language Group	Target Place	Following Place	Error Response in Clear Speech			Error Response in Fast Speech		
			Labial	Alveolar	Velar	Labial	Alveolar	Velar
AE	Labial	Labial		2			1	
		Alveolar			1		1	
		Velar			1		1	1
	Alveolar	Labial	2		1	**26****		4
		Alveolar	2			9		5
		Velar	2			10		3
	Velar	Labial				1	**12****	
		Alveolar	3	1		3	**16****	
		Velar	2	3		6	6	
Japanese	Labial	Labial		**16***	5		24	14
		Alveolar		5			**21***	9
		Velar		5	4		22	14
	Alveolar	Labial	**29***		15	**70****		36
		Alveolar	**22****		5	27		**53****
		Velar	13		12	25		**73****
	Velar	Labial	2	1		22	**41***	
		Alveolar	3	4		7	**42****	
		Velar	4	10		7	**19***	

Nasal Contrasts

Language Group	Target Place	Following Place	Error Response in Clear Speech			Error Response in Fast Speech		
			Labial	Alveolar	Velar	Labial	Alveolar	Velar
AE	Labial	Labial		3			**8***	
		Alveolar		4			**14****	2
		Velar		2			**9***	2
	Alveolar	Labial	3		2	6		2
		Alveolar	3			1		1
		Velar	3		1	2		7
	Velar	Velar		1			1	
Japanese	Labial	Labial		**37***	18		**78***	53
		Alveolar		13	21		35	39
		Velar		3	4		24	26
	Alveolar	Labial	34		46	49		**88****
		Alveolar	17		**82****	22		**111****
		Velar	35		**94****	23		**120****
	Velar	Labial	4	**65****		20	**57****	
		Alveolar	4	**50****		11	**46****	
		Velar	3	**54****		9	**64****	

** = p < .01, * = p < .05 by Binominal Test

Appendix W

Error Responses Made for Target Words by Japanese and AE listeners

JP in Clear Speech

Target Word	Error Response			Total	Error Rate (%)
	Labial	Alveolar	Velar		
din	23		50	73	50.7
ban	15		51	66	45.8
kin	38		24	62	43.1
ran	7		51	58	40.3
bang	4	46		50	34.7
rid	32		16	48	33.3
sung	2	43		45	31.3
king	1	33		34	23.6
hock	8	23		31	21.5
run	2		28	30	20.8
bam		15	12	27	18.8
lit	4		18	22	15.3
ram		12	8	20	13.9
ding	1	19		20	13.9
rum		7	12	19	13.2
sun	1		18	19	13.2
rung	3	15		18	12.5
rat	6		9	15	10.4
tad	11		4	15	10.4
sum		8	6	14	9.7
rang		13		13	9.0
sit	8		4	12	8.3
shock	1	10		11	7.6
bid	11			11	7.6
rap		7	2	9	6.3
bib		7	2	9	6.3
rib		5	4	9	6.3
Kim		7	2	9	6.3
dud	3		5	8	5.6
lip		5	2	7	4.9
shop		7		7	4.9
shot	3		4	7	4.9
cod	5		2	7	4.9
lad	2		5	7	4.9
dim		4	3	7	4.9
sat	2		4	6	4.2
rack		6		6	4.2
dub		4	2	6	4.2
tab		5	1	6	4.2
sap		5		5	3.5
sack		5		5	3.5
big	2	3		5	3.5
hot	4			4	2.8
lab		4		4	2.8
cog	1	3		4	2.8
lag	4			4	2.8
rig	2	2		4	2.8
tag		4		4	2.8
sick		3		3	2.1
dug		3		3	2.1
hop		1		1	0.7
sip		1		1	0.7
lick		1		1	0.7
cob		1		1	0.7

AE in Clear Speech

Target Word	Error Response			Total	Error Rate (%)
	Labial	Alveolar	Velar		
hock	5	12		17	23.6
kin	6			6	8.3
big	3	2		5	6.9
rack		4		4	5.6
rid	4			4	5.6
sip		2	1	3	4.2
sack		3		3	4.2
din	2		1	3	4.2
hop		1	1	2	2.8
shop		1	1	2	2.8
hot			2	2	2.8
lit			2	2	2.8
sit	1		1	2	2.8
shock	1	1		2	2.8
lab		2		2	2.8
bid	1		1	2	2.8
tag	1	1		2	2.8
bam		2		2	2.8
Kim		2		2	2.8
ram		2		2	2.8
sum		2		2	2.8
rap		1		1	1.4
sap		1		1	1.4
shot			1	1	1.4
sick		1		1	1.4
bib			1	1	1.4
dub			1	1	1.4
dud	1			1	1.4
lag		1		1	1.4
rig	1			1	1.4
dim		1		1	1.4
ban			1	1	1.4
ran			1	1	1.4
run	1			1	1.4
sung		1		1	1.4

JP in Fast Speech

Target Word	Error Response			Total	Error Rate (%)
	Labial	Alveolar	Velar		
hock	28	105		133	92.4
shock	50	53		103	71.5
kin	25		53	78	54.2
ban	21		56	77	53.5
lit	44		31	75	52.1
bam		51	23	74	51.4
sun	17		56	73	50.7
cog	18	52		70	48.6
ran	11		58	69	47.9
din	11		57	68	47.2
sack	11	56		67	46.5
rid	29		33	62	43.1
sit	44		17	61	42.4
rat	26		30	56	38.9
rap		40	13	53	36.8
sap		48	5	53	36.8
ram		16	35	51	35.4
dud	16		33	49	34.0
rum		27	22	49	34.0
run	9		39	48	33.3
rack	22	25		47	32.6
sung	6	41		47	32.6
bid	22		24	46	31.9
bang	10	35		45	31.3
lad	19		25	44	30.6
tad	12		32	44	30.6
cod	24		15	39	27.1
ding	12	27		39	27.1
king	5	28		33	22.9
shop		24	8	32	22.2
sat	14		18	32	22.2
rung	5	27		32	22.2
big	8	23		31	21.5
sum		17	14	31	21.5
dim		15	15	30	20.8
lick	12	17		29	20.1
bib		24	2	26	18.1
hop		21	4	25	17.4
cob		8	12	20	13.9
Kim		11	9	20	13.9
shot	7		12	19	13.2
sick	9	10		19	13.2
rib		8	10	18	12.5
lip		17		17	11.8
tab		12	5	17	11.8
sip		12	2	14	9.7
lab		10	4	14	9.7
lag	1	13		14	9.7
rig	5	9		14	9.7
rang	2	9		11	7.6
hot	8		1	9	6.3
dub		5	4	9	6.3
tag	1	5		6	4.2
dug	3			3	2.1

AE in Fast Speech

Target Word	Error Response			Total	Error Rate (%)
	Labial	Alveolar	Velar		
hock	27	48		75	52.1
shock	11	54		65	45.1
bid	34		1	35	24.3
rap		34		34	23.6
cog	5	27		32	22.2
sap		21		21	14.6
rack		20		20	13.9
bam		17		17	11.8
rid	9		4	13	9.0
dug	5	4		9	6.3
lip		7		7	4.9
lit			7	7	4.9
Kim		6	1	7	4.9
sum		4	2	6	4.2
sun	1		5	6	4.2
sack		5		5	3.5
rum		4	1	5	3.5
ran	1		4	5	3.5
din	4			4	2.8
hop		3		3	2.1
dud			3	3	2.1
lad	2		1	3	2.1
shop		2		2	1.4
hot	2			2	1.4
shot	1		1	2	1.4
lick		2		2	1.4
tab		2		2	1.4
cod			2	2	1.4
lag		2		2	1.4
kin	2			2	1.4
run	1		1	2	1.4
sat	1			1	0.7
sit	1			1	0.7
sick		1		1	0.7
dub		1		1	0.7
lab			1	1	0.7
tad			1	1	0.7
rig		1		1	0.7
sung		1		1	0.7

Appendix X

Error Responses Made for Target Words Sorted by the Following Contexts

(Japanese in Clear Speech)

Target Word	Following Stop	Error Response				Error Rate (%)
		Labial	Alveolar	Velar	Total	
hock	K	8	23		31	64.6
ban	K	9		20	29	60.4
din	K	12		15	27	56.3
din	T	6		20	26	54.2
kin	K	11		14	25	52.1
kin	P	19		5	24	50.0
ran	T	1		22	23	47.9
ran	K	3		20	23	47.9
din	P	5		15	20	41.7
ban	T	1		19	20	41.7
bang	K	2	17		19	39.6
rid	K	8		10	18	37.5
lit	T	2		15	17	35.4
rid	T	14		3	17	35.4
ban	P	5		12	17	35.4
run	K			17	17	35.4
bam	P		11	5	16	33.3
bang	P	1	15		16	33.3
sung	P	1	15		16	33.3
sung	T	1	15		16	33.3
king	P	1	14		15	31.3
bang	T	1	14		15	31.3
rid	P	10		3	13	27.1
kin	T	8		5	13	27.1
sung	K		13		13	27.1
tad	P	9		3	12	25.0
ram	P		8	4	12	25.0
ran	P	3		9	12	25.0
rum	P		6	4	10	20.8
king	T		10		10	20.8
shock	K		9		9	18.8
bam	T		4	5	9	18.8
sun	T			9	9	18.8
ding	P	1	8		9	18.8
king	K		9		9	18.8
ram	T		4	4	8	16.7
run	T	1		7	8	16.7
sun	K			8	8	16.7
rung	P		8		8	16.7
lip	P		5	2	7	14.6
sit	T	3		4	7	14.6
rat	K	2		5	7	14.6
dud	P	3		4	7	14.6
lad	P	2		5	7	14.6
Kim	P		6	1	7	14.6
sum	P		4	3	7	14.6
rum	T			7	7	14.6
rack	K		6		6	12.5
rib	P		4	2	6	12.5
ding	K		6		6	12.5
rap	P		5		5	10.4
shop	P		5		5	10.4
sit	P	5			5	10.4
sack	K		5		5	10.4
bib	K		4	1	5	10.4
bid	T	5			5	10.4
big	K	2	3		5	10.4
sum	T		3	2	5	10.4
run	P	1		4	5	10.4
rang	P		5		5	10.4
ding	T		5		5	10.4

Target Word	Following Stop	Error Response				Error Rate (%)
		Labial	Alveolar	Velar	Total	
rung	T	2	3		5	10.4
rang	K		5		5	10.4
rung	K	1	4		5	10.4
sap	P		4		4	8.3
rat	P	2		2	4	8.3
shot	P	3		1	4	8.3
rat	T	2		2	4	8.3
bib	P		3	1	4	8.3
lab	P		4		4	8.3
tab	P		4		4	8.3
bid	K	4			4	8.3
cog	K	1	3		4	8.3
dim	T		2	2	4	8.3
rap	T		1	2	3	6.3
sat	P	1		2	3	6.3
hot	P	3			3	6.3
sat	T	1		2	3	6.3
lit	K	1		2	3	6.3
shot	K			3	3	6.3
sick	K		3		3	6.3
dub	P		1	2	3	6.3
cod	P	3			3	6.3
cod	T	2		1	3	6.3
tag	T		3		3	6.3
dug	K		3		3	6.3
dim	P		2	1	3	6.3
rang	T		3		3	6.3
shop	T		2		2	4.2
lit	P	1		1	2	4.2
shock	T	1	1		2	4.2
dub	T		2		2	4.2
rib	K			2	2	4.2
bid	P	2			2	4.2
tad	T	1		1	2	4.2
lag	T	2			2	4.2
rig	T	1	1		2	4.2
bam	K			2	2	4.2
rum	K		1	1	2	4.2
sum	K		1	1	2	4.2
sun	P	1		1	2	4.2
hop	P		1		1	2.1
sip	P		1		1	2.1
sap	T		1		1	2.1
rap	K		1		1	2.1
hot	T	1			1	2.1
lick	K		1		1	2.1
cob	T		1		1	2.1
rib	T		1		1	2.1
tab	T		1		1	2.1
dub	K		1		1	2.1
tab	K			1	1	2.1
cod	K			1	1	2.1
dud	K			1	1	2.1
tad	K	1			1	2.1
lag	P	1			1	2.1
rig	P	1			1	2.1
tag	P		1		1	2.1
lag	K	1			1	2.1
rig	K		1		1	2.1
Kim	T			1	1	2.1
Kim	K		1		1	2.1

Appendix Y

Error Responses Made for Target Words Sorted by the Following Contexts

(AE in Clear Speech)

Target Word	Following Stop	Error Response				Error Rate (%)
		Labial	Alveolar	Velar	Total	
hock	K	4	12		16	66.7
rack	K		4		4	16.7
sack	K		3		3	12.5
big	T	2	1		3	12.5
kin	P	3			3	12.5
hop	P		1	1	2	8.3
sit	T	1		1	2	8.3
shock	K	1	1		2	8.3
lab	P		2		2	8.3
rid	P	2			2	8.3
big	K	1	1		2	8.3
Kim	P		2		2	8.3
bam	T		2		2	8.3
kin	T	2			2	8.3
sap	P		1		1	4.2
sip	P		1		1	4.2
rap	T		1		1	4.2
shop	T		1		1	4.2
sip	T			1	1	4.2
shop	K			1	1	4.2
sip	K		1		1	4.2
lit	P			1	1	4.2
shot	P			1	1	4.2
hot	T			1	1	4.2
hot	K			1	1	4.2
lit	K			1	1	4.2

Target Word	Following Stop	Error Response				Error Rate (%)
		Labial	Alveolar	Velar	Total	
sick	P		1		1	4.2
hock	T	1			1	4.2
bib	T			1	1	4.2
dub	K			1	1	4.2
bid	P			1	1	4.2
bid	T	1			1	4.2
rid	T	1			1	4.2
dud	K	1			1	4.2
rid	K	1			1	4.2
tag	T	1			1	4.2
lag	K		1		1	4.2
rig	K	1			1	4.2
tag	K		1		1	4.2
ram	P		1		1	4.2
dim	T		1		1	4.2
sum	T		1		1	4.2
ram	K		1		1	4.2
sum	K		1		1	4.2
ban	P			1	1	4.2
din	P			1	1	4.2
din	T	1			1	4.2
din	K	1			1	4.2
kin	K	1			1	4.2
ran	K			1	1	4.2
run	K	1			1	4.2
sung	K		1		1	4.2

Appendix Z

Error Responses Made for Target Words Sorted by the Following Contexts

(Japanese in Fast Speech)

Target Word	Following Stop	Error Response				Error Rate (%)
		Labial	Alveolar	Velar	Total	
hock	T	10	36		46	95.8
hock	K	13	31		44	91.7
hock	P	5	38		43	89.6
shock	P	22	18		40	83.3
shock	T	19	21		40	83.3
lit	P	33		2	35	72.9
sack	T	3	30		33	68.8
cog	P	10	23		33	68.8
ram	P		9	24	33	68.8
sap	P		31	1	32	66.7
lit	K	5		26	31	64.6
ban	K	8		22	30	62.5
cog	T	5	24		29	60.4
bam	P		25	4	29	60.4
ban	P	9		20	29	60.4
sun	P	13		15	28	58.3
kin	P	10		18	28	58.3
ran	K	1		27	28	58.3
rid	P	17		10	27	56.3
din	T	4		23	27	56.3
sit	K	17		9	26	54.2
bam	T		13	13	26	54.2
kin	K	9		17	26	54.2
sit	P	21		3	24	50.0
rum	P		14	10	24	50.0
kin	T	6		18	24	50.0
shock	K	9	14		23	47.9
sun	T	1		22	23	47.9
ran	T	4		19	23	47.9
rat	K	3		19	22	45.8
din	P	6		16	22	45.8
sun	K	3		19	22	45.8
rap	P		18	3	21	43.8
rap	T		14	7	21	43.8
rat	P	15		6	21	43.8
rid	T	8		13	21	43.8
sack	P	6	14		20	41.7
rack	T	8	12		20	41.7
bang	P	4	16		20	41.7
bid	K	5		14	19	39.6
dud	K	4		15	19	39.6
bam	K		13	6	19	39.6
din	K	1		18	19	39.6
dim	P		11	7	18	37.5
ran	P	6		12	18	37.5
ban	T	4		14	18	37.5
run	T	3		15	18	37.5
run	K	1		17	18	37.5
bid	P	12		5	17	35.4
dud	P	9		8	17	35.4
lad	P	14		3	17	35.4
lad	K	2		15	17	35.4
sung	T	1	16		17	35.4
rung	K	2	15		17	35.4
sat	K	4		12	16	33.3
sum	P		12	4	16	33.3
rum	T		10	6	16	33.3
sung	P	5	11		16	33.3
tad	P	8		7	15	31.3
tad	T	1		14	15	31.3
cod	K	7		8	15	31.3
big	P	5	10		15	31.3
rack	P	10	4		14	29.2
sack	K	2	12		14	29.2
rid	K	4		10	14	29.2
tad	K	3		11	14	29.2
bang	T	5	9		14	29.2
ding	K	5	9		14	29.2
king	K		14		14	29.2
sung	K		14		14	29.2
hop	T		10	3	13	27.1
shop	K		9	4	13	27.1
sat	P	9		4	13	27.1
rat	T	8		5	13	27.1
rack	K	4	9		13	27.1
cod	P	10		3	13	27.1
dud	T	3		10	13	27.1
ding	P	4	9		13	27.1
run	P	5		7	12	25.0
king	P	4	8		12	25.0
ding	T	3	9		12	25.0
rap	K		8	3	11	22.9
sap	K		7	4	11	22.9
sit	T	6		5	11	22.9
shot	K	3		8	11	22.9
cod	T	7		4	11	22.9
Kim	P		7	4	11	22.9
sum	T		4	7	11	22.9
bang	K	1	10		11	22.9
hop	P		10		10	20.8
sap	T		10		10	20.8
shop	T		8	2	10	20.8
lick	P	7	3		10	20.8
lick	K	3	7		10	20.8
bib	K		10		10	20.8
bid	T	5		5	10	20.8
lad	T	3		7	10	20.8
big	K	3	7		10	20.8
ram	K		4	6	10	20.8
rung	P	2	8		10	20.8
shop	P		7	2	9	18.8
lit	T	6		3	9	18.8
lick	T	2	7		9	18.8
sick	T	1	8		9	18.8
dub	P		5	4	9	18.8
bib	T		8	1	9	18.8
rum	K		3	6	9	18.8
rib	K		3	5	8	16.7
cog	K	3	5		8	16.7
ram	T		3	5	8	16.7
lip	T		7		7	14.6
bib	P		6	1	7	14.6
cob	T		3	4	7	14.6
cob	K		3	4	7	14.6
rig	T	1	6		7	14.6
dim	T		2	5	7	14.6
king	T	1	6		7	14.6
lip	P		6		6	12.5
sip	P		6		6	12.5
sick	K	4	2		6	12.5
cob	P		2	4	6	12.5
rib	P		3	3	6	12.5
tab	P		4	2	6	12.5
tab	K		4	2	6	12.5
rig	P	4	2		6	12.5
big	T		6		6	12.5
lag	K	1	5		6	12.5
Kim	T		3	3	6	12.5
rang	P	1	5		6	12.5
sip	T		5		5	10.4
lab	T		4	1	5	10.4
tab	T		4	1	5	10.4
lab	K		2	3	5	10.4
lag	P		5		5	10.4
dim	K		2	3	5	10.4
rung	T	1	4		5	10.4
lip	K		4		4	8.3
hot	P	4			4	8.3
shot	P	3		1	4	8.3
shot	T	1		3	4	8.3
sick	P	4			4	8.3
lab	P		4		4	8.3
rib	T		2	2	4	8.3
tag	T	1	3		4	8.3
sum	K		1	3	4	8.3
sip	K		1	2	3	6.3
sat	T	1		2	3	6.3
hot	K	2		1	3	6.3
dug	P	3			3	6.3
lag	T		3		3	6.3
Kim	K		1	2	3	6.3
rang	K	1	2		3	6.3
hop	K		1	1	2	4.2
hot	T	2			2	4.2
rang	T		2		2	4.2
tag	P		1		1	2.1
rig	K		1		1	2.1
tag	K		1		1	2.1

Appendix AA

Error Responses Made for Target Words Sorted by the Following Contexts

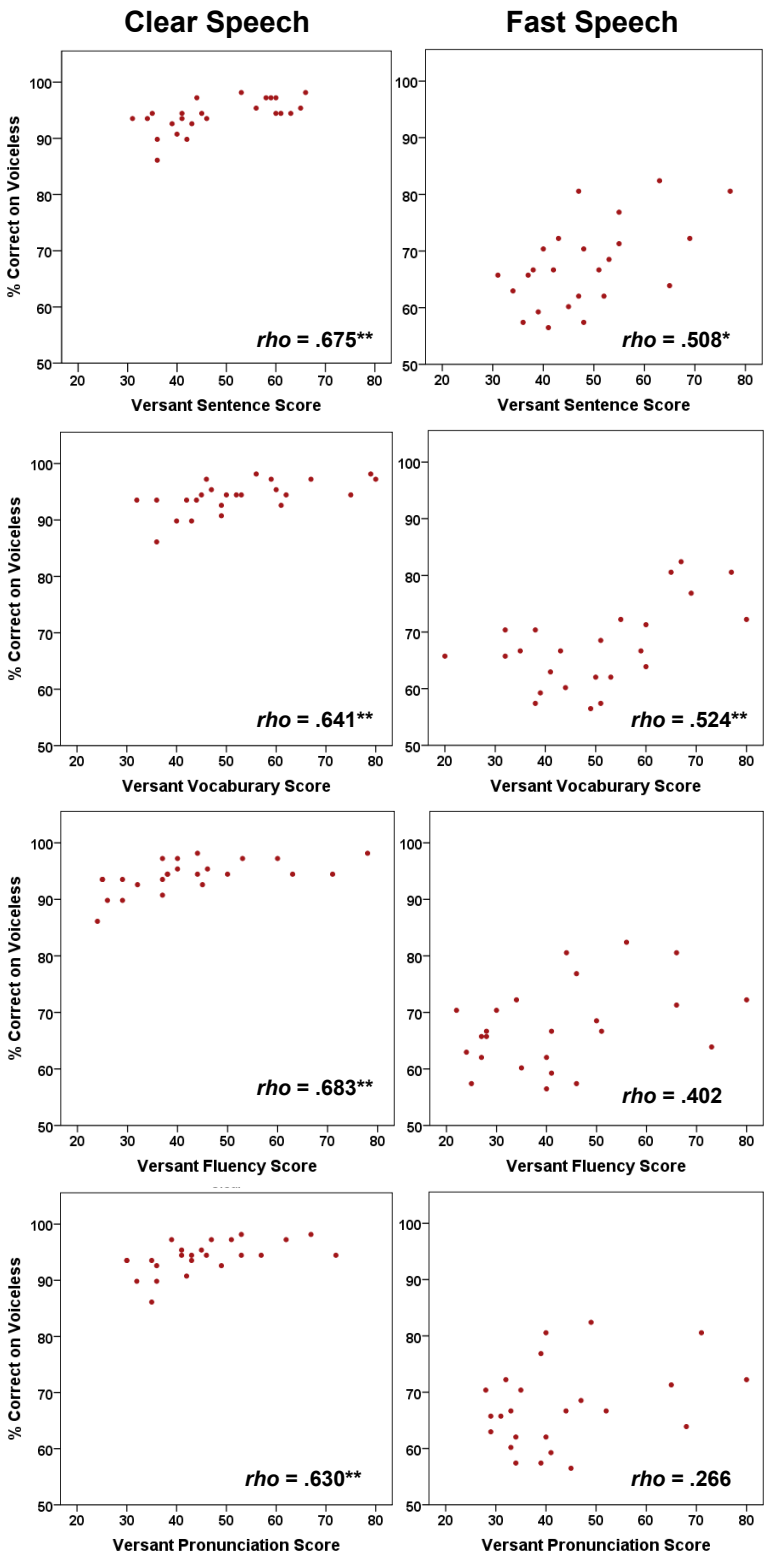
(AE in Fast Speech)

Target Word	Following Stop	Error Response				Error Rate (%)
		Labial	Alveolar	Velar	Total	
hock	K	17	17		34	70.8
rap	P		27		27	56.3
shock	K	2	21		23	47.9
hock	P	2	20		22	45.8
shock	P	3	19		22	45.8
sap	P		21		21	43.8
shock	T	6	14		20	41.7
hock	T	8	11		19	39.6
bid	P	18		1	19	39.6
cog	T		15		15	31.3
rack	K		14		14	29.2
cog	P		12		12	25.0
bam	T		11		11	22.9
bid	T	9			9	18.8
rid	P	6		2	8	16.7
lit	K			7	7	14.6
bid	K	7			7	14.6
lip	P		6		6	12.5
rap	T		6		6	12.5
cog	K	5			5	10.4
rack	T		4		4	8.3
sack	K		4		4	8.3
dug	T	3	1		4	8.3
dug	K	1	3		4	8.3
Kim	P		4		4	8.3
bam	K		4		4	8.3
sun	K			4	4	8.3
rid	K	3			3	6.3
rum	K		2	1	3	6.3
hop	T		2		2	4.2
shop	T		2		2	4.2
hot	P	2			2	4.2
rack	P		2		2	4.2
lick	K		2		2	4.2
lad	P	2			2	4.2
cod	T			2	2	4.2
rid	T			2	2	4.2
lag	K		2		2	4.2
bam	P		2		2	4.2

Target Word	Following Stop	Error Response				Error Rate (%)
		Labial	Alveolar	Velar	Total	
sum	P		2		2	4.2
rum	T		2		2	4.2
sum	T		1	1	2	4.2
Kim	K		2		2	4.2
sum	K		1	1	2	4.2
sun	P	1		1	2	4.2
din	P	2			2	4.2
ran	P	1		1	2	4.2
ran	K			2	2	4.2
hop	P		1		1	2.1
lip	T		1		1	2.1
rap	K		1		1	2.1
shot	P	1			1	2.1
sit	P	1			1	2.1
sat	T	1			1	2.1
shot	K			1	1	2.1
sack	P		1		1	2.1
sick	P		1		1	2.1
tab	P		1		1	2.1
dub	T		1		1	2.1
lab	K			1	1	2.1
tab	K		1		1	2.1
dud	P			1	1	2.1
dud	T			1	1	2.1
dud	K			1	1	2.1
lad	K			1	1	2.1
tad	K			1	1	2.1
dug	P	1			1	2.1
rig	K		1		1	2.1
Kim	T			1	1	2.1
kin	P	1			1	2.1
run	P	1			1	2.1
din	T	1			1	2.1
ran	T			1	1	2.1
din	K	1			1	2.1
kin	K	1			1	2.1
run	K			1	1	2.1
sung	K		1		1	2.1

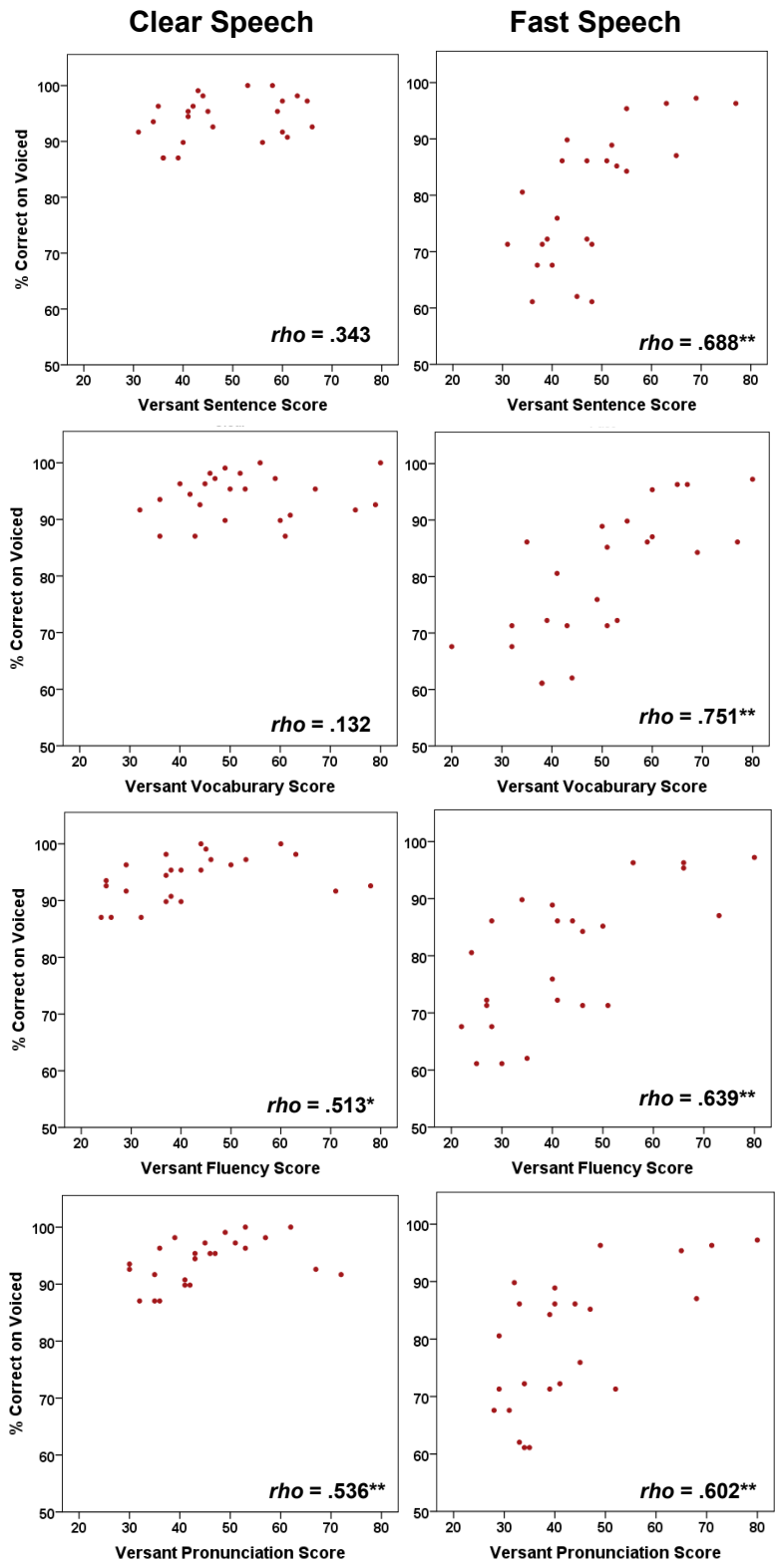
Appendix AB

Correlation of Japanese Performance on Voiceless Contrasts with Versant Subscores



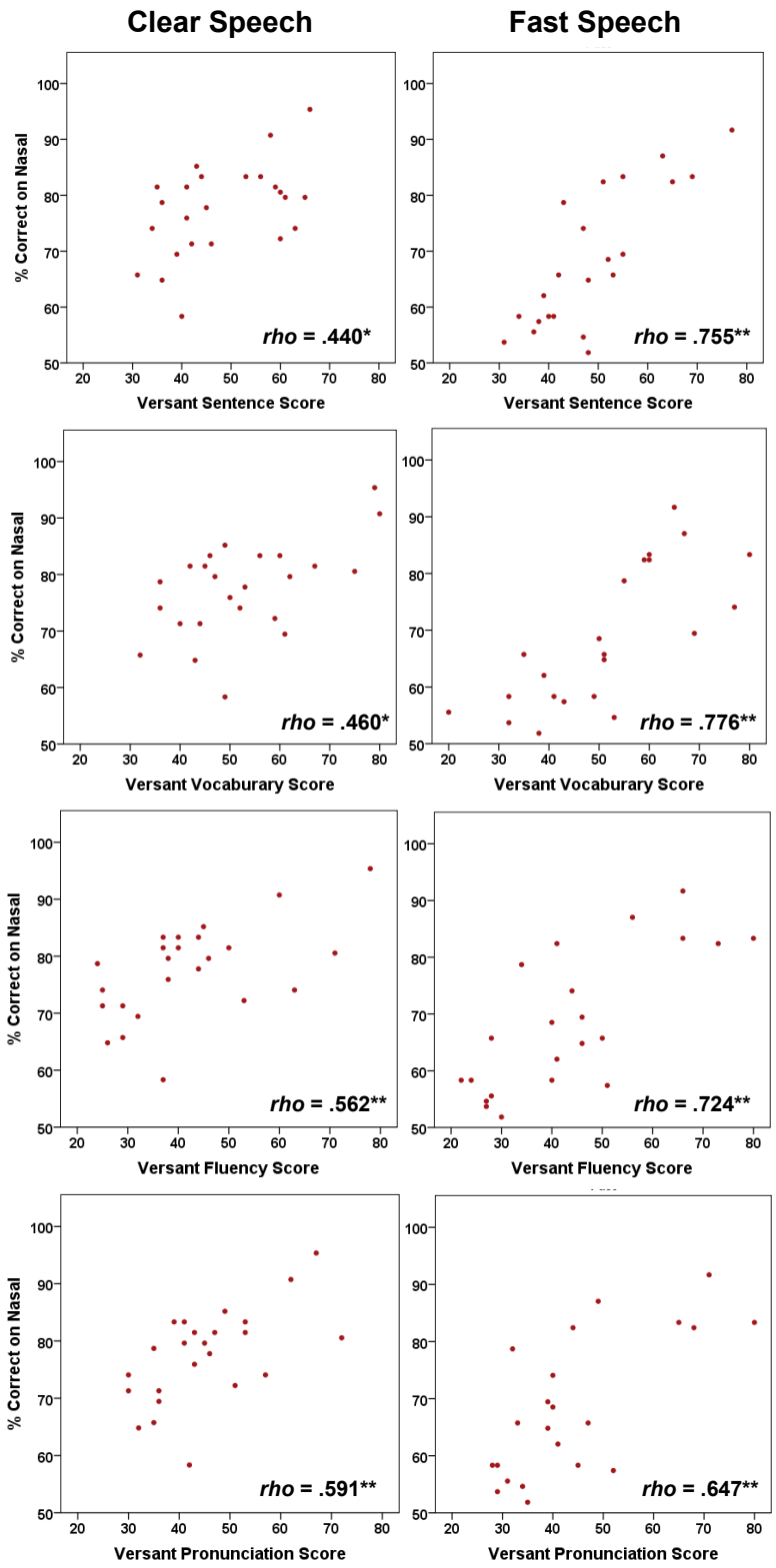
Appendix AC

Correlation of Japanese Performance on Voiced Contrasts with Versant Subscores



Appendix AD

Correlation of Japanese Performance on Nasal Contrasts with Versant Subscores



Appendix AE

Correlation of Japanese Performance with Language Backgrounds (Spearman's *rho*)

	Clear Speech				Fast Speech			
	% Correct on Voiceless	% Correct on Voiced	% Correct on Nasal	% Correct on Overall	% Correct on Voiceless	% Correct on Voiced	% Correct on Nasal	% Correct on Overall
Age	-0.001	0.183	0.181	0.195	-0.205	0.066	0.002	0.02
AOA	-.423*	-0.146	-0.186	-0.253	-.425*	-.434*	-.463*	-.411*
LOR	.625**	.491*	.651**	.729**	0.313	.662**	.609**	.584**
Versant Overall	.747**	.423*	.568**	.658**	.470*	.753**	.817**	.770**
Versant Sentence	.675**	0.343	.440*	.528**	.508*	.688**	.755**	.687**
Versant Vocabulary	.641**	0.132	.460*	.466*	.524**	.751**	.776**	.761**
Versant Fluency	.683**	.513*	.562**	.680**	0.402	.639**	.724**	.666**
Versant Pronunciation	.630**	.536**	.591**	.692**	0.266	.602**	.647**	.572**
L2 Use Speech/Listen Home	0.308	0.163	0.313	0.295	-0.139	0.118	0.092	0.078
L2 Use Speech/Listen Work/School	-0.412	0.025	-0.347	-0.285	0.238	0.054	-0.025	-0.01
L2 Use Speech/Listen other places	-0.216	-0.187	-0.195	-0.19	0.308	0.097	0.073	0.146
L2 Use Speech/Listen Mean	-0.025	0.06	-0.024	-0.009	0.192	0.224	0.156	0.176
L2 Use Read/Write Home	0.16	0.139	0.225	0.204	-0.147	0.07	0.038	0.026
L2 Use Read/Write Work/School	-.437*	-0.054	-0.269	-0.349	-0.014	-0.153	-0.152	-0.169
L2 Use Read/Write Other Places	-0.188	-0.064	-0.114	-0.109	0.064	-0.036	-0.06	-0.026
L2 Use Read/Write Mean	-0.131	-0.086	-0.13	-0.172	-0.082	-0.154	-0.152	-0.175

References

Abramson, A. S. & Tingsabadh, K. (1999). Thai final stops: cross-language perception. *Phonetica*, *56*, 111-122.

Amanuma, Y., Otsubo, K. and Mizutani, O. 1983: *Nihongo onseigaku* [*Japanese phonetics*]. Tokyo: Kuroshio.

Aoyama, K. (2003). Perception of syllable-initial and syllable-final nasals in English by Korean and Japanese speakers. *Second Language Research*, *19*, 251–265

Barry, M. C. (1985). A palatographic study of connected speech processes. In *Cambridge papers in phonetics and experimental linguistics (Vol. 4)*. Department of Linguistics, University of Cambridge.

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America*, *114*, 1600-1610.

Bent, T., Bradlow, A. R., & Smith, B. L. (2007). Segmental errors in different word positions and their effects on intelligibility of non-native speech: all's well that begins well. In O. Bohn & M.J. Munro (Eds.) *Language experience in Second Language Speech Learning: In honor of James Emil Flege* (pp. 331-347). Amsterdam: John Benjamins Publishing Company.

Bernstein, J. (2009). Proficiency instrumentation for cross language perception studies. J*ournal of the Acoustical Society of America*, *125*, 2753-2753.

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

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*, 775-794.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O. -S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production.* Amsterdam: John Benjamins.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, *121*, 2339-2349.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, *112*, 272-284.

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America*, *106*, 2074-2085.

Bradlow, A. R., Torreta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: global and fine-gained acoustic-phonetic talker characteristics. *Speech Communication*, *20*, 255-272.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977-990.

Byrd, D. (1993). 54,000 American stops. *UCLA Working Papers in Phonetics*, 83, 97-116.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15*, 39-54.

Byrd, D., & Tan, C. C. (1996). Saying consonant clusters quickly. *Journal of Phonetics*, *24*, 263-282.

Coenen, E., Zwitserlood, P., & Bölte, J. (2001). Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes*, *16*, 535-564.

Cutler, A., & Otake, T. (1998). Assimilation of place in Japanese and Dutch. *Proceedings of the Fifth International Conference on Spoken Language Processing*, 1751-1754.

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, *33*, 824-844.

Crystal, T. H., & House, A. S. (1982). Segmental durations in connected-speech signals: preliminary results. *Journal of the Acoustical Society of America*, *83*, 1553-1573.

Crystal, T. H., & House, A. S. (1988a). The duration of American-English stop consonants: an overview. *Journal of Phonetics*, *16*, 285-294.

Crystal, T. H., & House, A. S. (1988b). Segmental durations in connected-speech signals: current results. *Journal of the Acoustical Society of America*, *83*, 1553-1573.

Daniloff, R., & Hammarberg, R. (1973). On defining coarticulation. *Journal of Phonetics*, *1*, 239-248.

Davidson, L. (2006). Schwa elision in fast speech: segmental deletion or gestural overlap? *Phonetica*, *63*, 79-112.

Davidson, L. (2011). Characteristics of stop releases in American English spontaneous speech. *Speech Communication*, *53*, 1042-1058.

Deelman, T. & Connine, C. (2001). Missing information in spoken word recognition: nonreleased stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 656-663.

Flege, J. (1988) Factors affecting degree of perceived foreign accent in English sentences, *Journal of the Acoustical Society of America*, *84*, 162–177.

Flege, J. E. (1989). The perception of /t/ and /d/ by native and Chinese listeners. *Journal of the Acoustical Society of America*, *84*, 1639-1652.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and Problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E. & Fletcher, K. (1992) Talker and listener effects on the perception of degree of foreign accent, *Journal of the Acoustical Society of America*, *91*, 370–389.

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

Flege, J. E., & Wang, C. (1989). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/ - /d/ contrast. *Journal of Phonetics, 17*, 299-315.

Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. Language and Speech, 36, 171-195.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 144–158.

Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 380–396.

Gaskell, M. G., & Marslen-Wilson, W. D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language*, *44*, 325–349.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*, *63*, 223-230.

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica*, *38*, 148-158.

Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, *45*, 133–159.

Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journalof Experimental Psychology: Human Perception and Performance*, *28*, 163–179.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception and Psychophysics*, *65*, 575–590.

Greisbach, R. (1992). Reading aloud at maximal speed. *Speech Communication*, *11*, 469-473.

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

Holst, T. & Nolan, F. (1995). The influence of syntactic structure on [s] to [ʃ] assimilation. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence: papers in laboratory phonology IV* (pp. 315 – 333). Cambridge: Cambridge University Press.

House, A. S., Williams, C. E., Hecker, M. H. L. & Kryter, K. D. (1965). Articulation-testing methods: consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America*, *37*, 158-166.

Householder, F. W. Jr. (1956). Unreleased PTK in American English. In M. Halle, H. G. Lunt, H. McLean, and C. H. Van Schooneveld (Eds), *For Roman Jakobson: Essays on the Occasion of His Sixtieth Birthday, 11 October 1956.* pp. 235-244. The Hague, the Netherlands: Monton & Co.

Ito, K. & Strange, W. (2009). Perception of allophonic cues to English word boundaries by Japanese second language learners of English. *Journal of the Acoustical Society of America*, *125*, 2348-2360.

Kent, R. D., & Minifie, F. D. (1977). Coarticulation in recent speech production models. *Journal of Phonetics*. *5*, 295-320.

Kent, R. D., & Read, C. (2001). *The acoustic analysis of speech*. San Diego: Singular.

Kerswill, P. E. (1985). A sociophonetic study of connected speech processes in Cambridge English: An outline and some results. *Cambridge Papers in Phonetics and Experimental Linguistics*, *4*, 1–39.

Koster, C. (1987). Word recognition in foreign and native language: Effects of context and assimilation. Dordrecht: Foris Publications.

Kreul, E. J., Nixon, J. C., Kryter, K. D., Bell, D. W., Lang, J. S. & Schubert, E. D. (1968). A proposed clinical test of speech discrimination. *Journal of Speech and Hearing Research*, *11*, 536-552.

Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Labov, W. (2007). Transmission and Diffusion. *Language*, *83*, 344-387.

Lahiri, A., & Marslen-Wilson, W.D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, *38*, 245–294.

Lamothe, P., & Horowitz, A. (2006). StoryCorps. The *Journal of American History*, *93*, 171–174.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modeling* (pp. 403-439). Kluwer: Dordrecht.

Lindblom, B. (1996). Role of articulation in speech perception: Clues from production. *Journal of the Acoustical Society of America*, *99*, 1683-1692.

Lisker, L. (1999). Perceiving final voiceless stops without release: effects of preceding monophthongs versus nonmonophthongs. *Phonetica*, *56*, 44-55.

Manuel, S. Y. (1995). Speakers nasalize /ð/ after /n/ but listeners still hear /ð/. *Journal of Phonetics*, *23*, 453-476.

Manuel, S. Y., Shattuck-Hufnagel, S. Huffman, M. K., Stevens, K. N., Carlson, R., & Hunnicutt, S. (1992). Studies of vowel and consonant reduction. *ICSLP-1992*, 943-946.

MacKay, I. R. A., Meador, D., & Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica*, *58*, 103-125.

Matthies, M., Perrier, P., Perkell, J. & Zandipour, M. (2001). Variation in anticipatory coarticulation with changes in clarity and rate. *Journal of Speech, Language and Hearing Research*, *44*, 340-353.

Max, L. & Caruso, A. J. (1997). Acoustic measures of temporal intervals across speaking rates: variability of syllable- and phrase-level relative timing. *Journal of Speech, Language and Hearing Research*, *40*, 1097-1110.

Miller, J. L. (1981) Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the Study of Speech*. (pp. 39-74). Hillsdale: Lawrence Erlbaum Associates.

Moon, S.-J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America*, *96*, 40-54.

Munhall, K. & Löfqvist, A. (1992). Gestural aggregation in speech: laryngeal gestures. *Journal of Phonetics*, *20*, 111-126.

Nakajo, O. (1990). *Nihongo no onin to akusento* [*Japanese phonology and accent*]. Tokyo: Keisoo shoboo.

Nix, A., Gaskell, G., & Marslen-Wilson, W. D. (1993). Phonological variation and mismatch in lexical access. In *Proceedings of the 3rd European Conference on Speech Communication and Technology* (pp. 67-71). Berliln: ESCA.

Nolan, F. (1992) The descriptive role of segments: evidence from assimilation . In G. J. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: gesture, segment, prosody* (pp. 261–280). Cambridge: CUP.

Nolan F., Holst T., & Kühnert B.(1996). Modelling [s] to [ʃ] accommodation in English. *Journal of Phonetics*, 24, 113-137.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, *32*, 258-278.

Otake, T., Yoneyama, K., Cutler, A., & van der Lugt, A. (1996). The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America*, *100*, 3831-3842.

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

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, *28*, 96-103.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*, 434-446.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1989). Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research*, *32*, 600-603.

Pickett, J. M. (1999). *The acoustics of speech communication: fundamentals, speech perception theory, and technology*. Needham Heights: Allyn & Bacon.

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

Raphael, L. J., Borden, G. J., & Harris, K. S. (2011). *Speech science primer: Physiology, acoustics,and perception of speech* (6th ed.). Philadelphia: Lippincott Williams & Wilkins.

Repp, B. H. (1986). Some observations on the development of coarticulation. *Journal of the Acoustical Society of America*, *79*, 1616-1619.

Shaiman, S. (2001). Kinematics of compensatory vowel shortening: the effect of speaking rate and coda composition on intra- and inter-articulatory timing. *Journal of Phonetics*, *29*, 89-107.

Shaiman, S., Adams, S. G., & Kimelman, M. D. Z. (1995). Timing relationships of the upper lip and jaw across changes in speaking rate. *Journal of Phonetics*, *23*, 119-128.

van Son, R. J. J. H. & Pols, L. C. W. (1990). Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, *88*, 1683-1693.

van Son, R. J. J. H. & Pols, L. C. W. (1992). Formant movements of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, *92*, 121-127.

Surprenant, A., & Goldstein, L. (1988). The perception of speech gestures. *Journal of the Acoustical Society of America*, *104*, 518–529.

Strange, W. (2006). Second-language speech perception: The modification of automatic selective perceptual routines. *Journal of the Acoustical Society of America*, *120*, 3137-3137.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: a working model. *Journal of Phonetics*.

Strange, W. & Shafer, V. L. (2007). Speech perception in second language learners: The reeducation of selective perception. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Sumner, M. & Samuel, A. G. (2005). Perception and representation of regular variation: the case of final /t/. *Journal of Memory and Language*, *52*, 322-338.

Takata, Y. & Nábělek, A. K. (1990). English consonant recognition in noise and in reverberation by Japanese and American listeners. *Journal of the Acoustical Society of America*, *88*, 663-666.

Tsukada, K. (2004). Cross-language perception of final stops in Thai and English: a comparison of native and non-native listeners. *Proceedings of 10th Australian International Conference on Speech Science & Technology*, Sydney, Australia, pp. 563-568.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., & Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech and Hearing Research*, *39*, 494-509.

Vance, T. J. (1987). *An introduction to Japanese Phonology*. New York: State University of New York Press.

Warren, P., & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, *41*, 262-275.

Warren, P., & Marslen-Wilson, W. (1988). Cues to lexical choice: Discriminating place and voice. *Perception & Psychophysics*, *43*, 21-30.

Zsiga, E. C. (1994). Acoustic evidence for gestural overlap in consonant sequences. *Journal of Phonetics*, *22*, 121-140.

Zsiga, E. C. (2000). Phonetic alignment constraints: consonant overlap and palatalization in English and Russian. *Journal of Phonetics*, *28*, 69–102.