Perception of English Word-final Stops in Clearly Articulated Connected Speech by Japanese L2 Learners of English*

Kikuyo Ito
National Institute of Technology, Kagawa College

ABSTRACT. 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/ in running speech by adult Japanese second language (L2) learners of English, using real words. Minimal triplets differing in place of articulation of the word-final stop (e.g., *sip*, *sit*, and *sick*), followed by adverbs starting with /p/, /t/, or /k/ in sentences were presented in clearly articulated speech. Participants chose one of three written options of the target words after listening to a sentence, such as *He said the word sit positively* (or *tauntingly* or *cautiously*).

Results showed that Japanese listeners had difficulty identifying word-final unreleased oral stops, indicating their heavy reliance on the release. Japanese listeners also showed marked difficulty in correctly perceiving word-final nasal stops, in contrast to American listeners' ceiling-level performance.

Keywords: L2 phonetic perception, word-final consonants, connected speech

1. Introduction

English word-final consonants are realized differently in connected speech from the same word-final consonants produced in isolation because of coarticulation with the following segment. Their realization also varies with speaking rate and style (e.g., Gay 1981; Manuel *et al.* 1992). Although some variations are reported to be less intelligible than their base forms (e.g., Householder 1956; Nolan 1992; Gaskell and Marslen-Wilson 2001), first language (L1) 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 and Samuel 2005; Gow 2002, 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. The goal of the present study was to investigate L1 and L2 perception of English word-final oral and nasal stops with different places of articulation in connected speech.

The study examined the identification of English words ending with an oral or nasal stop followed by a word beginning with an oral stop /p/, /t/, or /k/ in clearly articulated connected speech. Native speakers of Japanese (JP) and American English (AE) participated in the experiment.

Since the present study adopted real words for the target stimuli, the lexical influence on word identification was analyzed by examining the correlation of percent accuracy of the experiment with word familiarity and with word frequency of the target words. The correlations between perceptual performance of JP listeners and their length of residence (LOR) and age of arrival (AOA) in English-speaking countries, and their English proficiency were also examined.

2. Methods

2.1. Participants

Twenty-four adult native speakers of JP living in the U.S. and a control group of twelve adult native speakers of AE participated in the experiment. The information of JP and AE participants is presented in Table 1.

Language Group	Gender Distribution	Mean Age (Range)	Mean LOR (Range)	Mean AOA (Range)
JP (N = 24)	3 males 21 females	32.3 (22 – 44)	5y 8m (1m – 17y 8m)	25.5 (17 – 34)
AE (N = 12)	4 males 8 females	30 (23 – 44)		

Table 1 *Biographical information on Japanese and AE participants*

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 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/).

The target words were followed by one of the following three adverbs: *positively*, *tauntingly*, or *cautiously*, creating a total of 162 two-word sequences. 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, a total of 72 filler sequences using word-initial minimal triplets, such as *pick-tick-kick* were included. All two-word combinations including the fillers are presented in the appendix.

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 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." All sentences were digitally recorded using a microphone (SHURE SM 48) in a sound-attenuated room at a sampling rate of 22,050 Hz, monaural, with 16-bit resolution, supported by SOUND FORGE 4.5 software.

A total of 468 stimulus sentences, which included two physically different tokens of 162 target and 72 filler sentences, were presented in the experiment.

The presence or absence of the oral stop release and the magnitude of the release, measured by multiplying the amplitude of each release by its duration, were examined and summarized in Figure 1. The following context (*following place*, hereafter) was categorized into *Different* or *Same* context for each contrast type, depending on whether the following place coincides with the target place.

A clear pattern seen in the data is that the word-final stops of the target words (*target stops*, hereafter) were not released in most cases in Same contexts whereas they were almost always released in Different contexts. This pattern was seen in both Voiceless and Voiced contrasts, and was most clearly seen in Labial in which the stops were never released in Same contexts and were always released in Different contexts.

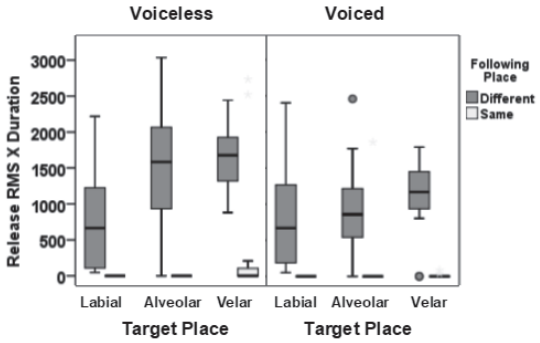

Figure 1 *Magnitude of Oral Stop Release Compared by Following Context*

2.3. Procedures

Participants were tested individually in a sound-attenuated room in the Speech-Language-Hearing Sciences department at the Graduate Center of the City University of New York, using the computer software, Paradigm. 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 the Versant™ English Test (for JP participants only). The entire session was completed within two hours for JP participants and within one and a half hours for AE participants.

2.3.1. Main experiment

The stimuli were presented binaurally through headphones (Telephonics TDH 39) in random order. 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 participants were asked to identify the target word by clicking on one of the three written alternatives appearing on the computer screen (e.g., *sip, sit* and *sick*).

The experiment consisted of 12 blocks, each of which contained 39 trials. 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, 5-minute breaks were inserted after the third, the sixth, and ninth blocks.

2.3.2. Word familiarity rating task

The main experiment was followed by a word familiarity rating task that asked participants to rate their familiarity with the written target words 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).

2.3.3. Versant English test

The Versant™ English Test, a 20-minute computerized telephone test, was administered only to JP participants. It assesses English spoken language skills of non-native speakers by the overall score with four diagnostic subscores (Pronunciation, Fluency, Sentence Mastery, and Vocabulary). It has been reported that extended one-on-one interviews by two experts correlate scores greater than 90% (Bernstein 2009).

3. Results
3.1. Language effect (AE vs. JP)

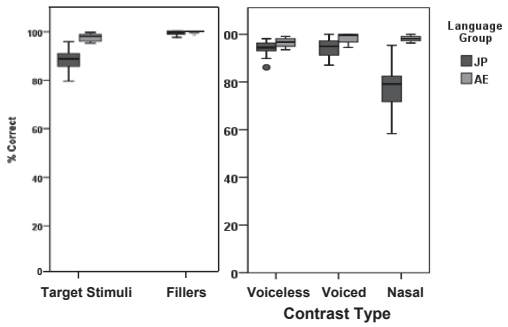

Figure 2 *Percent Correct Accuracy by JP and AE Listeners*

The left panel in Figure 2 illustrates the overall response accuracy by AE and JP listeners. The performance on the filler items by both JP and AE listeners was at ceiling level whereas their performance on the target stimuli exhibited the language effect showing better performance by AE listeners than JP listeners (median response accuracy 88.9% by JP *vs*. 98.5 % by AE). The overall response accuracy by AE listeners was significantly higher than that of JP listeners [Mann Whitney *U*, $U = 3$, $z = -4.74$, $p < 0.001$], with a very large effect size [$r = 0.79$].

The JP and AE performance on each contrast type (Voiceless, Voiced, Nasal) is presented in the right panel of Figure 2. The JP performance was significantly lower than the AE performance on all contrast types [Mann Whitney *U* with Bonferroni adjustments, $U = 63$, $z = -2.74$, $p < 0.01$ for Voiceless, $U = 48$, $z = -3.25$, $p < 0.001$ for Voiced, $U = 0$, $z = -4.84$, $p < 0.001$ for Nasal]. The language effect was larger for Nasal [$r = 0.81$] than for Voiceless [$r = 0.45$] and Voiced [$r = 0.53$] contrasts.

Figure 3 shows the performance on each target place (Labial, Alveolar, Velar) of each contrast type. While the JP performance on Voiceless Alveolar /t/ and Voiced Alveolar /d/ was significantly poorer than the AE performance [Mann Whitney *U* with Bonferroni adjustments, $p < 0.001$ 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 AE listeners. The results indicate that the language effects observed for Voiceless and Voiced contrasts are attributable to the JP group's significantly worse performance only on Alveolar stops. On the other hand, the JP performance on Nasal contrasts of all target places were significantly worse than the corresponding AE performance, with very large to medium effects [$r = 0.77$ for /n/, $r = 0.81$ for /ŋ/, $r = 0.49$ for /m/].

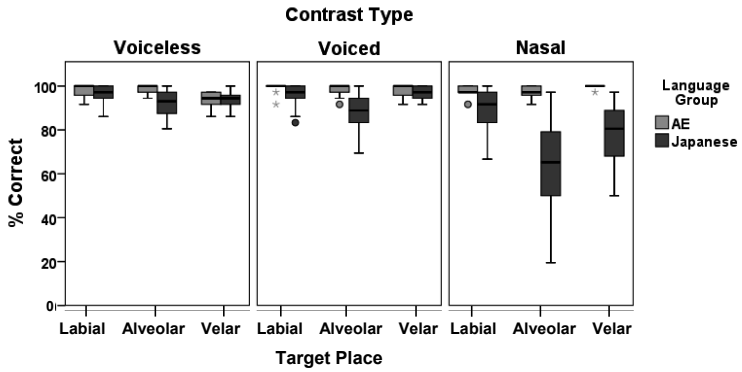

Figure 3 *Percent Correct Accuracy on Target Place by AE and JP Listeners*

3.2. Contrast type comparisons (Voiceless *vs.* Voiced *vs.* Nasal)

The performance differences between contrast types and those between target places within each contrast type were examined within each language group and are briefly summarized below. For the AE group, a non-parametric Friedman's test and three separate Wilcoxon Signed-Rank tests with Bonferroni adjustments were carried out because of the heterogeneity of variance and ceiling effects. A one-way repeated measures ANOVA and three separate repeated measures ANOVAs with Bonferroni adjustments were adopted for the JP group.

The AE 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, showing the following relationship in terms of the goodness of the performance: Voiceless < Voiced ≈ Nasal. The results indicate that word-final voiceless stops are relatively harder to perceive than voiced and nasal counterparts even for native listeners.

The JP performance on Nasal contrasts was significantly poorer than that on both Voiced and Voiceless contrasts with very large effect sizes ($\eta_p^2 > 0.8$ for both contrasts). The difference in their performance between Voiced contrasts and Voiceless contrasts did not differ significantly, showing the following relationship regarding the goodness of the performance: Nasal < Voiceless ≈ Voiced. The results indicate marked difficulty in correctly perceiving the place of articulation of word-final nasal stops in a sentence, showing a very different perceptual pattern from that seen in the AE results.

3.3. Following place comparisons (Different *vs.* Same)

Wilcoxon Signed Rank Tests with Bonferroni adjustments revealed no significant difference between Different and Same contexts in the AE performance except for Voiceless Velar /k/ that showed better performance on Different than on Same context.

The JP performance seemed to be more affected by following place. JP listeners' mean response accuracies on Same and Different contexts with standard errors are presented in Figure 4.

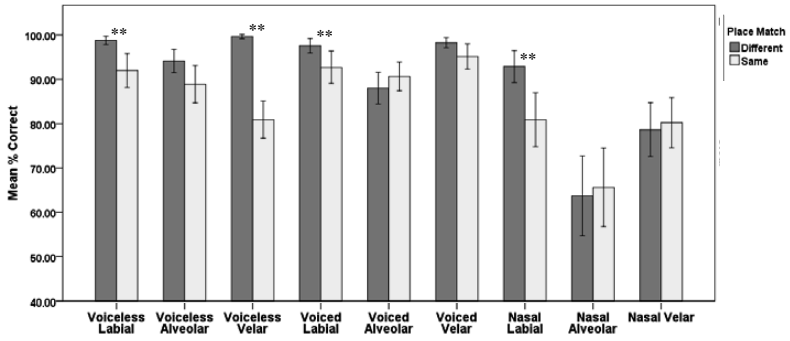

Figure 4 *Mean Percent Correct Accuracy by JP on Following Place*

Repeated measures ANOVAs revealed that four out of nine target stimulus types showed better performance on Different than on Same context and that Same context was never perceived better than Different context in any of the contrast types in the JP performance. Since the release of a word-final oral stop in Same context was either deleted or critically reduced, the JP listeners' pattern showing better performance on Different context than on Same context for oral stops suggests that JP listeners were utilizing stop release cues for place identification of word-final oral stops. The next part examines 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 the stop releases and the corresponding performance.

3.4. Correlations between magnitude of stop release and performance

The correlations between the magnitude of the oral stop releases of the target words with the corresponding performance by JP and AE listeners were examined, by multiplying the RMS amplitude of the release by duration as the indicator of the magnitude of the release. Spearman rank order correlations revealed positive correlations of the JP performance on both Voiceless [*rho*=0.55, $p < 0.01$] and Voiced [*rho*=0.34, $p < 0.01$] tokens, indicating JP listeners' heavy reliance, especially for Voiceless stops, on the place information in the stop release. As for AE listeners, significant correlations were seen in performance on Voiceless tokens only [*rho*=0.300, $p < 0.01$], showing similar but much weaker patterns of correlations than those of the JP performance.

3.5. Lexical effects on performance: word frequency and word familiarity

The lexical effects were analyzed by examining the correlation of performance on the main experiment with scores on the word familiarity task and with word frequency of the target words based on the SUBTLEXUS corpus. The SUBTLEXUS norms are reported to predict lexical decision times quite consistently (Brysbaert and New 2009). Spearman rank-order correlations showed no correlations between the percent correct accuracy on target words and the corresponding word familiarity scores by either of the language groups. None of the correlations between the response accuracy and word frequency scores were significant either, indicating that there were no lexical effects on the performance.

3.6. Correlations with LOR, AOA, and language proficiency.

The correlations of the JP listeners' performance with their LOR, AOA, and English proficiency measured by the Versant™ English Test (*Versant Test*, hereafter) were examined by adopting Spearman rank order correlations. The scatter plots showing the correlations of

the overall accuracy of JP listeners with their LOR, AOA, and Versant Test scores are presented in Figure 5. Strong positive correlations of LOR were seen not only with the overall performance but also with the performance on all three contrast groups.

The correlations of the JP listeners' performance with their AOA were much weaker than those with LOR. A significant negative correlation was only seen with the performance on Voiceless contrasts.

The correlations of the JP performance with the overall scores of the Versant Test were almost as strong as those of LOR. Among the subscores of the Versant Test, Pronunciation and Fluency scores showed higher correlations with performance than the other subscores.

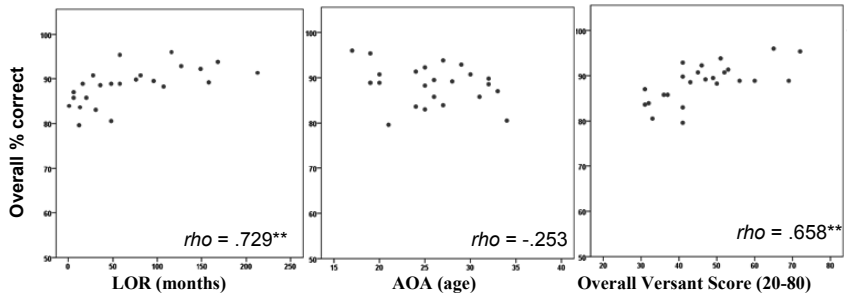

Figure 5 *Correlations of JP Performance with LOR, AOA, and Language Proficiency*

4. Discussion

Results of the present study indicated that the language effect predicting less accurate perception by JP than AE listeners was evident in the performance of the main experiment on all contrast types. No correlations between performance and word familiarity or word frequency were found in either of the language groups, indicating no lexical effects on their place perception of word-final stops in connected speech. Perceptual patterns by JP and AE listeners observed in this study are discussed below.

4.1. JP listeners' perceptual difficulty in identifying word-final nasal stop

The JP listeners' difficulty in identifying word-final nasal stops was clearly indicated in the result showing the JP group's much lower performance on Nasal contrasts than the corresponding AE performance, as well as in the result showing the JP group's much poorer performance on Nasal contrasts than their performance on Voiceless and on Voiced contrasts.

The observed difficulty by JP listeners is explicable by attributing it to the negative L1 influence caused by the place assimilatory nature of the Japanese moraic nasal. The place of articulation of the moraic nasal /N/ is underspecified because of its obligatory place assimilation to the following segment, resulting in several allophonic variations including [m], [n] and [ŋ] (Amanuma *et al.* 1983; Vance 1987; Nakajo 1990). This archiphonemic nature of Japanese moraic /N/ create L1-based perceptual patterns of syllable-final nasals by JP listeners (e.g., Otake *et al.* 1993, 1996), which are likely to be applied to non-native inputs (e.g., Cutler and Otake 1994, 1998). Aoyama (2003) reported that these L1-based perceptual patterns cause confusion in identifying the place of articulation of English word-final nasals produced in isolation. The results of the present study support the Aoyama study, extending the argument to the perception of English word-final nasals followed by other consonants in connected speech.

4.2. Influence of oral stop releases on performance

Results comparing the performance on Same and Different contexts of Voiceless and Voiced contrasts revealed that the JP performance in Same contexts was less accurate than that in Different contexts for the target places /p/, /k/, and /b/. No significant trend in the opposite direction was found. The AE performance showed no difference between Different and Same contexts except for /k/, which showed better performance on Different than on Same context.

The JP listeners' less accurate perception of oral stop contrasts in Same context than in Different context suggests their heavy reliance on acoustic information in stop releases because the release of a word-final oral stop in Same context was either deleted or critically reduced. The positive correlations between the magnitudes of the oral stop releases and the JP performance on the corresponding tokens strongly supported this assumption. The AE performance on Voiceless contrasts showed a weaker positive correlation, but their performance on Voiced contrasts did not show a correlation, indicating that AE listeners were much less dependent on the acoustic cues in stop releases. The results are 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 English word-final stops (Warren and Marslen-Wilson 1987, 1988).

4.3. Correlations of JP Performance with language experience and proficiency

The JP listeners' overall performance as well as their performance on all contrast types was more strongly correlated with their LOR than AOA, indicating that LOR is a better predictor of performance. In the scatter plots, however, the improvement of the performance as a function of LOR is not clearly seen after four to five years of LOR, which may indicate a possible cut-off point of the effects of L2 immersion on perceptual accuracy. This observation is in line with the notion that a long LOR alone may not guarantee the acquisition of difficult L2 contrasts, as has often been pointed out by Flege and colleagues (e.g. Flege and Liu 2001).

The JP performance on all contrast types was also positively correlated with their language proficiency as measured by the Versant Test. Among the subscores, the pronunciation score and fluency score were better correlated with the performance than the sentence mastery score and the vocabulary score, suggesting a close relationship between L2 phonetic perception in connected speech and L2 production skills.

4.4. Future direction

The present study is based on the author's dissertation study, which investigated the JP listeners' place perception of English word-final stops followed by word-initial stops in clearly articulated speech (*clear speech*, hereafter), as well as in casually produced fast speech (*fast speech*, hereafter). The current study only reported the results from clear speech, setting a starting point for the entire dissertation study and the following extension studies. The next paper will report findings from comparisons of the data for clear speech with another set of data for fast speech. In addition, the same experiment using the fast speech stimuli were administered to the same number of Korean listeners to confirm the notion that JP listeners' difficulty in identifying word-final nasals is due to the negative influence of L1 phonology. The report of the Korean study will follow the dissertation study. Furthermore, supported by a Grant-in-Aid for Scientific Research, or KAKENHI, a new set of follow-up

studies using the same experimental paradigm with the stimulus materials from new recordings by multiple speakers of AE are in preparation to establish a series of studies, the accumulated findings of which would make a contribution to the area of cross-language speech perception.

Note

*The present study is written based on my dissertation study submitted to the Ph.D. program in Speech-Language-Hearing Sciences at the Graduate Center of the City University of New York. I would like to thank the members of the supervisory committee of the dissertation, Dr. Winifred Strange, Dr. Klara Marton, Dr. Valerie Shafer, Dr. Lisa Davidson, and Dr. Douglas Whalen. This material is based upon work supported by the National Science Foundation under Grant No. BCS-1023192.

Appendix

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**	**tauntingly**
/m/-/n/-/ŋ/	/ɪ/	**dim-din-ding** **Kim-kin-king**	
	/æ/	**ram-ran-rang** **bam-ban-bang**	
	/ɑ/ or /ʌ/	**sum-sun-sung** **rum-run-rung**	**cautiously**
Fillers (word-initial minimal triplets: 24 words)			
contrast type		**minimal triplet stimuli**	
/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)	

References

Amanuma, Yasushi, Kazuo Otsubo and Osamu Mizutani. 1983. *Nihongo onseigaku* [Japanese phonetics]. Tokyo: Kuroshio Shuppan.

Aoyama, Katsura. 2003. Perception of syllable-initial and syllable-final nasals in English by Korean and Japanese speakers. *Second Language Research* 19.251–265

Bernstein, Jared. 2009. Proficiency instrumentation for cross language perception studies. *Journal of the Acoustical Society of America* 125.2753-2753.

Brysbaert, Marc and Boris New. 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.

Cutler, Anne and Takashi Otake. 1998. Assimilation of place in Japanese and Dutch. *Proceedings of the Fifth International Conference on Spoken Language Processing* 1751-1754.

Cutler, Anne and Takashi Otake. 1994. Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language* 33.824-844.

Flege, James Emil and Serena Liu. 2001. The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition* 23.527-552.
Flege, James Emil. 1989. The perception of /t/ and /d/ by native and Chinese listeners. *Journal of the Acoustical Society of America* 84.1639-1652.
Flege, James Emil and Chipin Wang. 1989. Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/ - /d/ contrast. *Journal of Phonetics* 17.299-315.
Gaskell, Gareth M. and William Marslen-Wilson. 2001. Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language* 44.325–349.
Gay, Thomas. 1981. Mechanisms in the control of speech rate. *Phonetica* 38.148-158.
Gow Jr., David W. 2002. Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance* 28.163–179.
Gow Jr., David W. 2003. Feature parsing: Feature cue mapping in spoken word recognition. *Perception and Psychophysics* 65.575–590.
Householder Jr., Fred W. 1956. Unreleased PTK in American English. *For Roman Jakobson: Essays on the occasion of his sixtieth birthday, 11 October 1956,* ed. by M. Halle, H. G. Lunt, H. McLean and C. H. Van Schooneveld, 235-244. The Hague: Mouton.
Manuel, Sharon Y. 1995. Speakers nasalize /ð/ after /n/ but listeners still hear /ð/. *Journal of Phonetics* 23.453-476.
Manuel, Sharon Y., Stefanie Shattuck-Hufnagel, Marie. K. Huffman, Kenneth N. Stevens, Rolf Carlson, and Sheri Hunnicutt. 1992. Studies of vowel and consonant reduction. *ICSLP-1992* 943-946.
Nakajo, Osamu. 1990. *Nihongo no onin to akusento.* Tokyo: Keisō shobō.
Nolan, Francis. 1992. The descriptive role of segments: evidence from assimilation. *Papers in laboratory phonology II: Gesture, segment, prosody*, ed. by G. J. Docherty and D. R. Ladd, 261–280. Cambridge: CUP.
Otake, Takashi, Giyoo Hatano, Anne Cutler and Jacques Mehler. 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32.258-278.
Otake, Takashi, Kiyoko Yoneyama, Anne Cutler and Arie van der Lugt. 1996. The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America* 100.3831-3842.
Sumner, Meghan and Arthur G. Samuel. 2005. Perception and representation of regular variation: the case of final /t/. *Journal of Memory and Language 52*.322-338.
Vance, Timothy J. 1987. *An introduction to Japanese Phonology*. New York: State University of New York Press.
Warren, Paul and William Marslen-Wilson. 1987. Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics* 41.262-275.
Warren, Paul and William Marslen-Wilson. 1988. Cues to lexical choice: Discriminating place and voice. *Perception & Psychophysics* 43.21-30.