Role of amplitude and pitch in the perception of Japanese stop length contrasts

Kaori Idemaru
(University of Oregon)

▌ABSTRACT▐

This study presents experiments which examined the role of amplitude and fundamental frequency (f0) in the phonetic perception of short versus long stop length contrasts in Japanese (e.g., [t] vs. [tt]). Stop length contrasts are normally characterized by differences in the duration of stop closures. However, closure duration can be unreliable as a perceptual cue when one considers variability in the rate at which people speak. Acoustically, the amplitude and f0 of the vowel following stop consonants are known to covary with the length distinction of stops in Japanese. Given this fact, the current study examined amplitude and f0 as potential secondary cues to the distinction. The results indicate that even though both amplitude and f0 are robust correlates, Japanese listeners do not use these cues in categorizing short versus long stops.

Key Words
Japanese, stop length, phonetics, speech perception, prosody

1. An acoustic and perceptual problem in phonetic length distinction

In Japanese, as in many other languages, segmental length is phonemic in both consonants and vowels. Therefore, *obasan* (‘aunt’) and *obaasan* (‘grandmother’) as well as *ita* (‘was there’) and *itta* (‘went’) are two separate lexical words in each pair. Phonemic length distinctions are typically characterized acoustically by differences in the duration of the vowel (short vs. long vowel) and the duration of the stop closure, frication or nasal murmur (short vs. long consonant[1])[2].

However, this characterization faces serious limitations when speaking rate is considered. Naturally, different people speak at different rates, and even in the same speaker and within the same utterance, speaking rate can vary. Speaking rate is a major contextual factor affecting the production and perception of speech (Miller 1981 for review). It has been shown to have a critical influence on the duration of acoustic properties that distinguish phonetic categories (Miller &Liberman, 1979; Summerfield, 1981; Miller & Baer, 1983). When people speak slowly, segmental durations are lengthened, and when people speak fast, they are shortened.One crucial consequence of this variation is that important durational values, such as typical short and longsegment durations and the boundary location between them, also

1) Here, ‘short’ and ‘long’ are used as equivalents of ‘singleton’ and ‘geminate’ when referring to consonant length, to make it consistent with vowel length description.
2) Following Kawahara (2006), ‘length’ is used to refer to the abstract phonological short versus long distinction, whereas ‘duration’ is used to refer to a physical and measurable acoustic property.

change, as demonstrated for several languages including Japanese (Fujisaki, 1979 for Japanese affricates; Idemaru & Guion- Anderson, 2010 and Amano & Hirata, 2010 for Japanese stops; Magen & Blumstein, 1993 for Korean vowels; Kessinger& Blumstein, 1998; Pind, 1999 for Icelandic stops).

For stop length contrast in Japanese, this results in an overlap between contrastingshort and long categoriesin terms of stop closure duration (Fujisaki 1979, Hirata &Whiton 2005, Idemaru&Guion-Anderson 2010). In other words, phonologically short vowels and consonants may actually be physically quite long in a slow utterance; while phonologically long vowels and consonants may be short in a fast utterance. A significant consequence of this variation is ambiguity in the acoustics, with overlapping (short and long) lengthcategories and a lack of fixed acoustic values that can specify phonological length across speech rates. Nonetheless, listeners appear to overcome this ambiguity by interpreting the relevant acoustic cues.

2. Possible solutions

One possible perceptual solution for overcoming this problem is relational-timing or rate normalization. The idea of relational-timing and rate normalization is that listeners adjust perception using the context. Applied to this context, listeners are thought to perceive the length of a segment (e.g., short or long consonant) using a criterion duration of, say, the preceding segment (e.g., the preceding vowel). Typically researchers examining relational-timing have analyzed the durational ratio of the segment in question and the adjacent segment, reporting that durational ratios provide more stable acoustic cuesfor distinguishing short and long consonants and vowels than their raw duration values (Pickett et al., 1999 for Italian short and long stops; Pind 1999 for Icelandic consonant and vowel length; Kohler 1979 for German lenis and fortis stops; Hirata &Whiton 2005, and Idemaru&Guion-Anderson 2010 for Japanese short and long stops). Idemaru & Guion-Anderson (2010) have further demonstrated that Japanese listeners do use relational timing in judging short and long stops.

While relational-timing provides an effective solution to the problem of variability of segmental duration for length distinction in Japanese, there is another potentially viable solution to the problem. As noted earlier, phonemic length distinctions are typically characterized acoustically by differences in the duration of the segment. However, there is a possibility that other acoustic properties are involved in defining length categories in Japanese. In general, speech categories are multi-dimensional, with multiple acoustic properties covarying with phonetic category distinctions (Coleman 2003; Dorman, Studdert-Kennedy, & Raphael, 1977 for stop place of articulation; Jongman, Wayland, & Wong, 2000 for fricative place of articulation; Hillenbrand, Clark, &Houde, 2000 for tense and lax vowels; Kluender, & Walsh, 1992 for fricative/affricate distinction; Lisker 1986 for stops voicing; Polka & Strange, 1985 for liquids). For example, as many as 16 acoustic properties covary with the distinction of stop voicing in English (i.e., [b] vs. [p]) (Lisker 1986). It has been demonstrated that native listeners use secondary acoustic cues in addition to the primary one in perception (Hillenbrand, Clark, &Houde, 2000; Kluender, & Walsh, 1992; Whalen, Abramson, Lisker&Mody, 1993; Francis et al 2000). It is therefore possible that Japanese listeners use secondary acoustic cues for phonemic segmental length distinction.

Stop length distinction is often associated with multiple covariants in production cross-linguistically (Lisker 1958,Abramson 1987, Ham 2001,Payne 2005, 2006; Idemaru&Guion, 2008). The amplitude of the syllable following stops is known to covary with short versus long stops in Pattani Malay and Bengali (Abramson 1992; Hankamer, Lahini, &Koreman 1989). For Japanese, the amplitude and

fundamental frequency (f0) of the following vowel have been reported to covary with stop length distinction[3] (Kawahara 2006; Idemaru&Guion, 2008). In the production of a short stop, the amplitude of the vowels was nearly consistent across the stop or was slightly larger in the following vowel. In contrast, in the production of a long stop, the amplitude of the following vowel was consistently lower than that of the preceding vowel (Idemaru&Guion 2008). As for f0, the previous work used test words with high-low (HL) or high-low-low (HLL) pitch patterns, with falling pitch across the medial stop. Within this environment, f0 fell consistently lower across a long stop than across a short stop (Kawahara, 2006; Idemaru&Guion, 2008).

Production of stops in Japanese thus presents multiple acoustic covariants that reliably differentiate short and long stops. It has been proposed that there are two aspects in the sound systems of languages that make them robust signaling devices despite various sources of variability such as speech rate: maximal acoustic/perceptual distance between contrasts and redundancy of information (Diehl et al 1991). The presence of multiple acoustic covariants can serve to provide redundant information, thus increasing the distance between contrasting sounds (i.e., short and long stops in this case). Given the ambiguity of category distinction solely on the basis of raw duration, the documented covariants, amplitude and f0, may play an important role in the perceptual distinction of the stop length in Japanese.

The goal of the current study was to conduct a preliminary investigation examining the perceptual role of amplitude (Experiment 1) and f0 (Experiment 2) in distinguishing short and long consonants in Japanese. The role of these perceptual cues was evaluated, while the speaking rate was controlled for, as a baseline investigation. In this study, listeners heard Japanese words, synthesized *seta* and *setta,* in which amplitude or f0were systematically manipulated,and indicated which word they heard.

3. Experiment 1 – Effect of amplitude

3.1 Method

3.1.1 Participants

Twenty two native Japanese speakers (13 female and 9 male) participated for a small fee. Participants were born in various regions of Japan(with the largest group, N=14, from Tokyo). All resided in the US at the time of testing. Length of residency in the US ranged from one month to nine years (mean = 2 years and 10 months). All reported normal hearing.

3.1.2 Stimuli

The experiment used Japanese words *seta* and *setta* (Idemaru&Guion, 2008; Idemaru&Guion-Anderson, 2010). The stimuli were synthesized using KlattWorks (McMurray, 2000) in order to systematically control for the duration of stop closure ([t]) and amplitude of the vowels ([e] and [a]). The durations of [s], [e] and [a] were set to be 125 ms, 70 ms, and 70 ms respectively, based on the production data in Idemaru andGuion-Anderson (2010). The duration of [t] varied from 50 to 250 ms in five 50-ms steps, ranging from an extreme value of short stop to an extreme value of long stop (Idemaru&Guion-Anderson, 2010).

To synthesize [s], frication noise was created using parameter values proposed by Klatt (1979).

3) The test words in these studies were produced in high-low (HL) or high-low-low (HLL) pitch patterns. Other pitch patterns, namely LH and LHH, are possible. However, this study focuses on the HL and HLL.

The F1 through F6 frequencies were 320, 1390, 2530, 3250, 3700 and 4900 Hz, with the parallel tract amplitude (A1 – A6) set at zero for the first five formants and 52 dB for F6. The amplitude of frication noise (AF) was set at 70 dB for the duration of the [s].

The vowels [e] and [a] were created with the steady state F1, F2, and F3 frequencies taken from the acoustic study of Japanese vowels by Keating and Hoffman (1984). In each stimulus, the F1 and F2 frequencies varied across the first 20 ms, rising from 276 to 476 Hz and 1515 to 1715 Hz respectively for [e]. For [a], F1 increased from 432 to 632 Hz and F2 decreased from 1663 to 1374 Hz, which is characteristic of vowels following [t]. This formant transition was determined using the locus equation of Sussman, McCaffrey, and Matthew (1991). The F3 frequencies, 2500 Hz for [e] and 2383 Hz for [a], were steady-state across the vowel.

Amplitude was 40 dB at the onset of [e], then increased linearly to 75 dB across the first 20 ms of [e], and decreased to 40 dB in the last 20 ms of the [e]. It then transitioned to 0 dB, where it remained for the duration of the stop, after which it increased linearly to the peak point across the first 20 ms of [a] and decreased to 40 dB in the last ms of the [a]. Critically to the experiment, the peak amplitude of [a] varied as 70, 73, 75, 77, and 80 dB. Consequently, the amplitude differences between the first vowel [e] (75 dB) and the second vowel [a] were -5, -2, 0, +2, and +5, simulating the amplitude scale between typical short and long stop production in Japanese (Idemaru&Guion, 2008).

For these vowels, the fundamental frequency (f0) was set at 160 Hz for [e] and 100 Hz for [a], within the typical range of male values (Idemaru&Guion, 2008). In Japanese production, F0 covaries with stop length such that there is a slight drop in f0 from the previous to the following vowel across a short stop, whereas the drop is more prominent across a long stop. The drop from 160 Hz at [e] to 100 Hz at [a] (i.e., difference of 60 Hz) was midway between the dropacross a short and a long stop,so there was no acoustic bias due to f0.

A 10-ms stop burst was excised from a natural production of *seta* by a male native Japanese speaker and was inserted before [a].Varying the closure duration and amplitude of [a] in five levels each resulted in 25 unique stimuli.

3.1.3 Procedure

Seated in individual sound-attenuated booths and wearing headphones (Beyer DT-150), listeners responded tofive repetitions each of the 25 stimuli (125 trials) by pressing response buttons labeled “seta” and “setta” in Japanese orthography. Stimulus presentation and response collection were under the control of E-Prime (Psychology Software Tools, Inc.).

3.2. Analysis and Results

In order to examine the influence of closure duration and amplitude in the perception of Japanese stop length contrasts, percent long-stop ([tt]) responses were submitted to a 5 x 5 (closure duration x amplitude) repeated-measures ANOVA. The test returned a significant main effect of closure duration [$F(4, 84) = 448.384$, $p < .001$], no effect of amplitude [$F(4, 84) = 78.493$, $p = .438$] and no interaction between the two factors [$F(16, 336) = 104.083$, $p = .142$]. Figure 1 illustrates the main effects of closure duration (a) and amplitude (b). Given the lack of significant interaction, variation due to amplitude was collapsed in (a), and variation due to closure duration was collapsed in (b).

(a) (b)

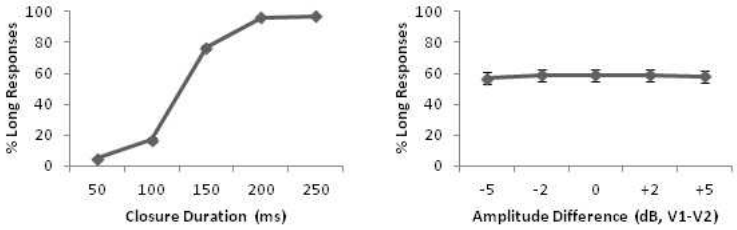

Figure 1. Effect of closure duration (a) and amplitude (b) in the perception of Japanese stop lengths.

As Figure 1 shows and the statistics confirm, Japanese listeners' perception of stop length was heavily influenced by the changes in closure duration (Figure 1(a)). This was expected as closure duration is the primary cue when the speaking rate is consistent, as it was in this experiment.

More importantly, we report here that Japanese listeners do not use amplitude as a secondary perceptual cue to the contrast (Figure 1(b)). Secondary perceptual cues often exert their strongest influence when the primary perceptual cue is ambiguous (Abramson &Lisker, 1985; Francis et al, 2008). The lack of interaction between closure duration and amplitude indicates that even when closure duration was ambiguous, there was no influence of amplitude. These results demonstrate that Japanese listeners do not use amplitude information in judging short and long stop lengths.

4. Experiment 2 – Effect of pitch

4.1 Method

The same native Japanese speakers who participated in Experiment 1 participated in this experiment. The procedure was identical to that of Experiment 1.

4.1.1 Stimuli

As in Experiment 1, synthesized Japanese words *seta* and *setta*were used. The duration of the segment as well as the spectral properties of each segment were the same as in the stimuli used in Experiment 1. The critical difference was that here, f0 of the vowels varied systematically, simulating the pattern of variation from typical short to long stop production in Japanese, while their amplitude remained at fixed values.

For the vowel [e], f0 was set at 160 Hz, whereas f0 of [a] varied from 130 to 70 Hz in five 15-Hz steps. As a result, the f0 differences between the first vowel (160 Hz) and the second were -30, -45, -60, -75, and -90 Hz, simulating the pitch scale between typical short and long stop production in Japanese (Idemaru&Guion, 2008).

Amplitude was 40 dB at the onset of [e], then increased linearly to 75 dB across the first 20 ms of [e] and decreased to 40 dB in the last 20 ms of the [e]. Amplitude then transitioned to 0 dB where it remained for the duration of the stop, after which it increased linearly to 75 dB across the first 20 ms of [a] and decreased to 40 dB in the last ms of the [a]. Varying the closure duration and f0 of [a] in five levels each resulted in 25 unique stimuli.

4.2 Analysis and Results

Percent long-stop ([tt]) responses were submitted to a 5 x 5 (closure duration x f0) repeated-measures ANOVA. The test returned a significant main effect of closure duration [$F(4, 84) = 286.104$, $p < .001$], no effect of f0 [$F(4, 84) = .515$, $p = .725$] and no interaction between the two factors [$F(16, 336) = 1.004$, $p = .452$]. Figure 2 illustrates the effect of closure duration (a) and f0 (b). Given the lack of significant interaction, variation due to f0 was collapsed in (a), and variation due to closure duration was collapsed in (b).

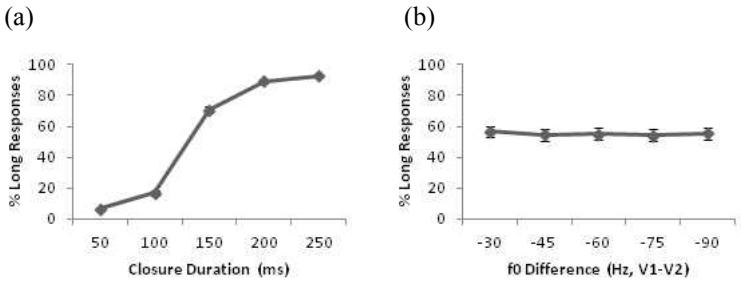

Figure 2. Effect of closure duration (a) and f0 (b) in the perception of Japanese stop lengths.

The results here parallel those of Experiment 1. Japanese listeners' perception of stop length is heavily dependent on closure duration (Figure 2(a)), and they do not use pitch (f0) as a secondary cue to this contrast (Figure 2(b)).

5. Discussion and Conclusion

This study explored the role of amplitude and pitch (f0)as secondary cues in the perception of Japanese stop length contrasts. This work was motivated by the finding that closure duration of the stop, which is the primary cue to the length contrast, can vary due to changes in speaking rate (Hirata &Whiton 2005; Idemaru&Guion-Anderson 2010). The variability in closure duration causes acoustic overlap between short and long stop categories, resulting in perceptual ambiguity if perception is solely based on closure duration. However, the current study demonstrated that Japanese listeners do not use either amplitude or pitch as a secondary cue.

There are at least two possible explanations for these findings. First, it may be that these non-durational secondary cues are not necessary and listeners do not exploit them because contextual duration (i.e., relational timing) can provide a robust perceptual cue to the contrast. Japanese listeners do use relative timing to compensate for variability due to speaking rate (Idemaru&Guion-Anderson, 2010). This finding demonstrates that the presence of acoustic covariants to a phonetic contrast does not necessarily mean that they are automatically used in perception. It is important to distinguish acoustic cues (covariants or correlates) and perceptual cues.

Second, note that the speaking rate of the stimuli did not vary in this study, as reflected in the fixed durations of non-target segments (i.e., [s], [e], and [a]). In this acoustic environment, closure duration remained a robust acoustic cue. It is possible that processing words at a consistent speaking rate led listeners to focus on the robust closure duration and not exploit other potential cues. This second point leaves open the possibility that either amplitude or pitch, or both, may only be recruited as secondary cues when the primary cue, closure duration, is ambiguous. To investigate this hypothesis,

further work is needed that examines the role of these properties when speaking rate varies and closure duration is unreliable as perceptual cue.*

❖ References

Abramson, A. S. "Amplitude as a Cue to Word-initial Consonant Length: Pattani Malay." *Proceedings of the 12th International Congress of Phonetic Sciences*. 1991. 98–101. Print.

_____. "Word-initial Consonant Length in Pattani Malay." *Proceedings of the Xlth International Congress of Phonetic Sciences*. 1987. 66–70. Print.

Amano, S., and Y. Hirata. "Perception and Production Boundaries Between Single and Geminate Stops in Japanese." *The Journal of the Acoustical Society of America* 128. 2010. 2049. Print.

Coleman, J. "Discovering the Acoustic Correlates of Phonological Contrasts." *Journal of Phonetics* 31.3-4. 2003. 351–372. Print.

Diehl, R. L, M. A Walsh, and K. R Kluender. "On the Interpretability of Speech/nonspeech Comparisons: A Reply to Fowler." *The Journal of the Acoustical Society of America* 89. 1991. 2905. Print.

Dorman, M. F., M. Studdert-Kennedy, and L. J. Raphael. "Stop-consonant Recognition: Release Bursts and Formant Transitions as Functionally Equivalent, Context-dependent Cues." *Perception & Psychophysics* 22.2 1977. 109-122. Print.

Francis, A. L., K. Baldwin, and H. C. Nusbaum."Effects of Training on Attention to Acoustic Cues."*Percept Psychophys* 62.8. 2000. 1668-80. Print.

Fujisaki, H., and T. Kawashima."On the Modes and Mechanisms of Speech Perception."*Annual Report of the Engineering Research Institute* 28 1969. 67–73. Print.

Ham, W. H. *Phonetic and Phonological Aspects of Geminate Timing*. Routledge, 2001. Print.

Hankamer, J., A. Lahiri, and J. Koreman. "Perception of Consonant Length: Voiceless Stops in Turkish and Bengali." *Journal of Phonetics* 17. 1989. 283-298. Print.

Hillenbrand, J. M., M. J. Clark, and R. A. Houde."Some Effects of Duration on Vowel Recognition."*The Journal of the Acoustical Society of America* 108. 2000. 3013. Print.

Hirata, Y., and J. Whiton."Effects of Speaking Rate on the Single/geminate Stop Distinction in Japanese."*The Journal of the Acoustical Society of America* 118. 2005. 1647. Print.

Idemaru, K., and S. G. Guion. "Acoustic Covariants of Length Contrast in Japanese Stops." *Journal of the International Phonetic Association* 38.02. 2008. 167-186. Print.

Idemaru, K., and S. Guion-Anderson."Relational Timing in the Production and Perception of Japanese Singleton and Geminate Stops." *Phonetica* 67.1-2. 2010. 25–46. Print.

Jongman, A., R. Wayland, and S. Wong."Acoustic Characteristics of English Fricatives." *The Journal of the Acoustical Society of America* 108. 2000. 1252. Print.

Kawahara, S. "A Faithfulness Ranking Projected from a Perceptibility Scale: The Case of[+ Voice] in Japanese." *Language(Baltimore)* 82.3 2006. 536-574. Print.

Keating, P. A. and M. K. Huffman."Vowel Variation in Japanese." *Phonetica* 41.4. 1984. 191-207. Print.

Kessinger, R. H., and S. E. Blumstein. "Effects of Speaking Rate on Voice-onset Time and Vowel Production: Some Implications for Perception Studies." *Journal of Phonetics* 26.2. 1998. 117-128. Print.

Klatt, D. H. "Synthesis by Rule of Segmental Durations in English Sentences." *Frontiers of Speech Communication Research*. 1979. 287-299. Print.

Kluender, K. R., and M. A. Walsh. "Amplitude Rise Time and the Perception of the Voiceless Affricate/fricative Distinction." *Perception and Psychophysics* 51.4. 1992. 328-333. Print.

Kohler, K. J. "Dimensions in the Perception of Fortis and Lenis Plosives." *Phonetica* 36.4-5. 1979. 332-43. Print.

Lisker, L. "Closure Duration and the Intervocalic Voiced-voiceless Distinction in English." *Language* 33.1. 1957. 42–49. Print.

_____. "'Voicing'in English: A Catalogue of Acoustic Features Signaling/b/ versus/p/in Trochees." *Language and Speech* 29.1. 1986. 3-11. Print.

Magen, H. S., and S. E. Blumstein."Effects of Speaking Rate on the Vowel Length Distinction in Korean." *The*

* Acknowledgements: I wish to thank the three anonymous reviewers for their valuable comments on an earlier version of this paper. Thanks also to Lucy Gubbins for help conducting the experiments.

Journal of the Acoustical Society of America 89. 1991. 1918. Print.

McMurry, Bob. *KlattWorks: A [somewhat] New Systematic Approach to Formant-based Speech Synthesis for Empirical Research*. 2000. Print.

Miller, J. L. "Effects of Speaking Rate on Segmental Distinctions." *Perspectives on the Study of Speech*. 1981. 39-74. Print.

Miller, J. L., and T. Baer. "Some Effects of Speaking Rate on the Production Of/b/and/w." *Journal of the Acoustical Society of America* 73.5. 1983. 1751-1755. Print.

Miller, J. L., and A. M. Liberman. "Some Effects of Later-occurring Information on the Perception of Stop Consonant and Semivowel." *Percept Psychophys* 25.6. 1979. 457-65. Print.

Payne, E. M. "Phonetic Variation in Italian Consonant Gemination." *Journal of the International Phonetic Association* 35.02. 2005. 153–181. Print.

Pickett, Emily R., Sheila E. Blumstein, and Martha W. Burton."Effects of Speaking Rate on the Singleton/Geminate Consonant Contrast in Italian."*Phonetica* 56.3-4. 1999. 135-157. Print.

Pind, Jorgen. "Speech Segment Durations and Quantity in Icelandic."*The Journal of the Acoustical Society of America* 106.2. 1999. 1045-1053. Print.

Polka, L., and W. Strange. "Perceptual Equivalence of Acoustic Cues That Differentiate/r/and/l." *The Journal of the Acoustical Society of America* 78. 1985. 1187–1197. Print.

Summerfield, Q. "Articulatory Rate and Perceptual Constancy in Phonetic Perception." *Journal of Experimental Psychology: Human Perception and Performance* 7.5. 1981. 1074. Print.

Sussman, H. M., H. McCaffrey, and S. Matthews."An investigation of locus equations as a source of relational invariance for stop place categorization." *The Journal of the Acoustical Society of America* 90. 1991. 1309–1325. Print.

Whalen, D. H et al. "F0 Gives Voicing Information Even with Unambiguous Voice Onset Times." *The Journal of the Acoustical Society of America*93. 1993. 2152–2152. Print.