RELATIONAL TIMING OR ABSOLUTE DURATION? CUE WEIGHTING IN THE PERCEPTION OF JAPANESE SINGLETON-GEMINATE STOPS

Kaori Idemaru and Lori L Holt

Carnegie Mellon University
idemaru@cmu.edu, lholt@andrew.cmu.edu

ABSTRACT

Relational timing has been proposed as a solution to the problem of variability across durational properties of speech arising with changes in speaking rate. The current study investigates the role of absolute and relational timing cues in perception of Japanese stop length (singleton/ geminate) categorization. Absolute (stop duration) and relational (ratio of stop duration to preceding mora duration) duration cues were independently varied in a categorization test. Although Ratio was shown previously to classify speakers' productions more accurately (Idemaru, 2005), listeners' category responses showed strong individual differences in cue use. These results demonstrate that a highly reliable acoustic cue in the distribution of cues available in speech production does not necessarily predict its primacy in speech perception.

Keywords: perception, cue weighting, Japanese

1. INTRODUCTION

Speaking rate exerts a powerful influence on the absolute duration of critical temporal cues that distinguish phonetic categories, yet listeners maintain perceptual constancy across speaking rate changes. The perceptual consequences of changes in speaking rate have been studied extensively, demonstrating that phonetic categorization of categories distinguished by temporal cues shifts with varying speaking rate such that the observed category boundary changes as a function of the speaking rate (Miller & Baer, 1983; Miller & Liberman, 1979; Summerfield, 1981). The temporal cues for phonetic categories appear to be perceived in a wholly context-dependent manner, dependent on speaking rate. Some researchers have noted that this context-dependent perception, combined with overlapping phonetic categories could weaken the category differences.

Further investigation on this problem led researchers to attempt to discover more stable acoustic properties across varying speaking rates in higher-order relational timing which are often expressed in the form of durational ratios. For example, the ratio of consonant closure to the duration of the preceding vowel successfully distinguishes the voicing contrast in English stops (Port & Dalby, 1982), as well as Italian single vs. geminate stop productions (Pickett, Blumstein, & Burton, 1999) and this property remains stable across different speaking rates. Relational timing has been demonstrated to be a more reliable property of the acoustic signal across changes in speaking rate for durational phonetic contrasts such as the consonant and vowel length categorization in Icelandic (Pind, 1999), the distinction of vowel plus lenis stop and vowel plus fortis stop in Standard German (Kohler, 1979), and the stop length contrast in Japanese (Hirata & Whiton, 2005; Idemaru, 2005). Pind (1986) suggested that these types of relational timing may provide an acoustic invariant across changes in speaking rate. Some of these studies have examined the role of relational timing in perception (Kohler, 1979; Pickett *et al*., 1999; Pind, 1999).

Investigating stop length contrasts in four languages, Ham (2001) reported interesting cross-linguistic differences in the acoustic realization of the durational contrast. Comparing stop closure duration of singletons versus geminates revealed substantially greater differentiation between the two categories for mora-timed languages (e.g., Japanese) than syllable-timed languages (e.g., Italian). This indicates that the singleton-geminate phonetic categories overlap along the duration dimension to a greater degree in syllable-timed languages than in mora-timed languages, perhaps making it a less reliable perceptual cue within syllable-timed languages. A consequence of this may be that there are cross-linguistic differences in

the perceptual cue weighting (Holt & Lotto, 2006) in singleton-geminate stop categorization.

Even though relational timing has been proposed for the categorization of stop length contrast in Japanese (Hirata & Whiton, 2005; Idemaru, 2005), a careful perceptual investigation is yet to be conducted. Given Ham's analysis, Japanese listeners may not necessarily use relational timing because, as a mora-timed language, the differentiation of singleton and geminate stops in Japanese is relatively robust. In fact, Idemaru (2005) showed that the singleton and geminate stops in Japanese produced in three different speaking rates can be categorized with 87% accuracy using only stop duration, while using stop to preceding mora duration ratio produced 93% accuracy. Given this, the current study attempts a direct comparison of stop duration and relational timing, as expressed in the durational ratio of stop consonant to the preceding vowel, in categorization of singleton geminate stops in Japanese.

2. METHODS

Twenty four native listeners of Japanese participated for a small payment. All listeners were residing in the US at the time of testing and reported normal hearing.

The perception experiment used nonce words *seta* and *setta* and methods from auditory category-learning experiments (Holt & Lotto, 2006) whereby sounds drawn from a two-dimensional acoustic space were used as stimuli. The two-dimensional acoustic space was defined by stop duration varying from 50 to 250 ms in 50 ms steps and the Ratio of preceding mora duration to stop duration, varying from .20 to 1.4 in .15 steps. These endpoints spanned values typical of Japanese singleton and geminate stops (Idemaru, 2005).

Stimuli were synthesized using KlattWorks (McMurray, in prep). The 2-d acoustic space determined the duration specifications for the [se] and [t]. The [a] duration was determined by the consonant-to-vowel durational ratio (2.00) reported by Idemaru (2005) as the value unbiased either for singleton or geminate. The F1, F2, and F3 values for the vowels [e] and [a] were taken from Keating and Hoffman (1984); those for [s] were taken from Klatt (1979).

Table 1: Target frequencies used in synthesis (Hz).

	[s]	[e]	[a]
F1	320	476	632
F2	1390	1715	1374
F3	2530	2500	2383

The formant-frequency varied across the first 20 ms, rising from 276 to 476 Hz and 1515 to 1715 Hz for [e] F1 and F2. For [a], F1 increased from 432 to 632Hz and F2 decreased from 1663 to 1374 Hz, characteristic of vowels following a [t]. This formant transition was determined using the locus equation of Sussman, McCaffrey, and Matthews. (1991). Amplitude was 40 dB for the duration of [s], then increased linearly to 60 dB across the first 20ms of [e] and decreased back to 40 dB in the last 20ms of the [e]. Amplitude then transitioned to 0 dB where it remained for the duration of the stop, then increased linearly to 60 dB across the first 20 ms of [a] and decreased back to 40 dB in the last 20 ms of the [a]. Fundamental frequency was 160 Hz for [e] and 100 Hz for [a], typical of male values (Idemaru, 2005). A 10 ms voice onset time (VOT), including a stop burst was excised from a natural production of *seta* by a male native speaker, and was inserted before [a].

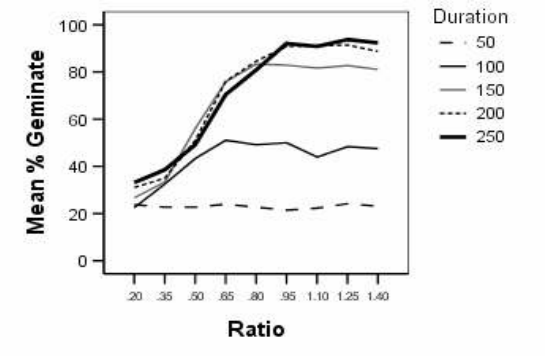
Seated in individual sound-attenuated booths and wearing headphones, listeners categorized 20 repetitions of each of the 45 stimuli by pressing response buttons labeled "seta" and "setta." The stimuli were presented by E-Prime (Psychology Software Tools, Inc.).

3. RESULTS

Figure 1 shows percent *setta* (geminate) responses as a function of the Duration and Ratio acoustic cues defining the 2-d stimulus space. Japanese listeners' singleton/geminate categorization appears to be influenced by both Ratio and Duration cues.

The typical acoustic singleton value, boundary value and typical geminate value of Ratio and Duration reported by Idemaru (2005) were selected as test levels (.20, .65, 1.40 for Ratio, and 50 ms, 150 ms, 250 ms for Duration), and a 3 x 3 (Ratio x Duration) repeated-measure ANOVA was run on the percent geminate responses averaged across 20 repetitions. Significant main effects were found for Ratio, $F(2, 46)=47.08$, $p<.0001$ and for Duration, $F(2,46)=46.24$, $p<.0001$. The interaction between the factors was also significant, $F(4, 92)=54.67$, $p<.0001$.

Figure 1: Categorization responses for five Duration values across nine Ratio values.

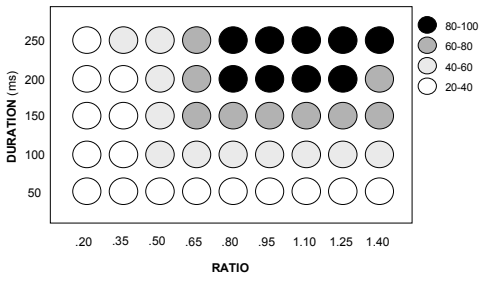


Post-hoc tests revealed that at 0.2, a Ratio value typical of singleton consonants, there was no significant difference between any pairs of the three Duration values. At Ratio of .65, differences in percent geminate responses between Duration of 50 and 150 ms, as well as Duration of 50 and 250 ms were significant, $t(23)=.-11.74$, $p=.000$ and $t(23)=-8.76$, $p=.000$; but the difference between Duration of 150 and 250 ms was not. At Ratio of 1.40, all pairs of the three Duration values showed significant difference, $t(23)=.-11.23$, $p=.000$ for Duration 50 and 150 ms, $t(23)=.-12.38$, $p=.000$ for Duration 50 and 250 ms, and $t(23)=.-3.20$, $p=.004$ for Duration 150 and 250 ms. These results indicate that there was a general trend for geminate responses to increase with increasing Ratio value. Moreover, for higher Ratio values (.65 and 1.40), Duration values affected the listeners categorization as well.

This is illustrated in Figure 2, showing listeners' categorization responses across the acoustic space defined by Duration and Ratio. No boundary line capturing listeners' categorization responses can be drawn simply in the Duration dimension or in the Ratio dimension, indicating that both cues influenced singleton/geminate categorization.

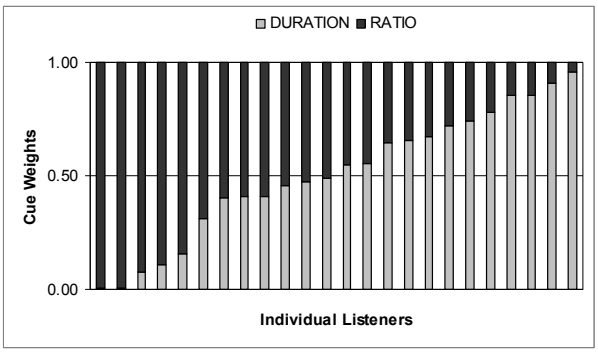
Following the methods of Holt and Lotto (2006), perceptual cue weights were computed for each listener as the correlation between Duration and Ratio values and the percent geminate responses. The absolute values of the correlation coefficients were normalized to sum to one. These values provided a quantitative estimate of the relative perceptual weight Japanese listeners give to the Duration and Ratio cues in singleton/geminate categorization.

Figure 2: Categorization responses plotted in the experimental acoustic space, indicating % geminate responses.



The mean cue weights for Duration (0.51) and Ratio (0.49) across all listeners were fairly unbiased; however, inspection of individual cue weights revealed substantial individual differences, as shown in Figure 3. Some listeners relied only on the Duration cue (1.0) without making use of the Ratio cue (0.0); others were characterized by the opposite pattern. Other listeners made use of both cues in their categorization responses. These cue weights did not correlate with the listeners' length of residency in the US, $r = .221$, $p = .299$.

Figure 3: Relative cue weights for the Duration and Ratio dimensions for each listener.



4. DISCUSSION AND CONCLUSION

Relational timing has been proposed as a solution to the problem of variability in durational properties of speech sounds caused by changes in speaking rate. Relational timing, often expressed in the form of a ratio of the durations of two segments of an utterance, has been shown to exhibit more stability across changes in speaking rate than raw durations (e.g., Picket *et al.*, 1999; Hirata & Whiton. 2005; Idemaru, 2005).

In previous work, this was tested for the stop length contrast in Japanese by examining the acoustic characteristics of singleton/geminate productions. The durational ratio of the stop and the preceding mora was more stable than stop duration when rate of speech varied and the ratio

was better correlated with category identity than stop duration (Idemaru, 2005). However, the current study has shown that the presence of a more reliable acoustic property does not necessarily predict its primacy in perception. Japanese listeners, as a group, did not exhibit a strong perceptual weighting for the ratio cue in any unified way in the current study. Whereas some listeners relied heavily on the ratio cue, others relied primarily on the duration cue and yet others used both cues.

In addition to these perceptual measures, speech productions were collected from the same set of native-Japanese participants to determine whether listeners' perceptual cue weights are reflected in their own patterns of cue use in speech production. Preliminary acoustic analysis of multiple productions of *seta* and *setta* spoken by two listeners who relied primarily on the duration cue in perception and two listeners who relied primarily on the ratio cue in perception suggest that ratio was better associated with category membership. Using only the ratio cue as a predictor, it was possible to classify, the correct category among 93% (Duration users) and 91% (Ratio users) of the produced tokens. Duration was somewhat less reliable, classifying 87% (Duration users) and 85% (Ratio users) of the tokens. Although based on a small sample, these results seem to suggest that the listeners' perceptual cue weights may not be well-predicted by their own cue use in production. Further analyses will be necessary to determine the reliability of this preliminary observation.

The results presented here extend the proposal presented by Ham (2001) regarding the cross-linguistic differences in phonetic realization of the stop length contrast. Ham noted that mora-timed languages show more robust differentiation of singleton and geminate stop length whereas syllable-timed languages exhibit more overlap between the two categories. Given this, one may predict that relational timing cue may be perceptually weighted more among syllable-timed languages than among mora-timed languages. The current study has shown that listeners of Japanese, a mora-timed language, do not exhibit a strong perceptual tendency to make use of the more reliable ratio cue in categorization.

Further work is needed to compare perceptual cue weights of Japanese listeners with cue weights of listeners of syllable-timed languages and to test the hypothesis that relational timing is perceptually more relevant among listeners of syllable-timed languages. Research on this issue will provide insight into perceptual cue weighting and its interaction with the distributional characteristics of acoustic cues in speech production.

[This work was supported by NSF 0345773. The authors thank Christi Gomez and Sung-Joo Lim for help in conducting the experiments.]

5. REFERENCES

[1] Ham, W. H. (2001). *Phonetic and phonological aspects of geminate timing*. New York: Routledge.

[2] Hirata, Y., Whiton, J. (2005). Effect of speaking rate on the single/geminate stop distinction in Japanese. *The Journal of the Acoustical Society of America, 118*(3), 1647-1660.

[3] Holt, L.L., Lotto, A. J. (2005). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America, 119*(5), 3059-3071.

[4] Idemaru, K. (2005). *An acoustic and perceptual investigation of the geminate and singleton stop contrast in Japanese.* Unpublished dissertation, University of Oregon.

[5] Keating, P., Hoffman, M. (1984). Vowel variation in Japanese. *Phonetica: International Journal of Speech Science, 41*, 191-207.

[6] Kohler, K. J. (1979). Dimensions in the Perception of Fortis and Lenis Plosives. *Phonetica: International Journal of Speech Science, 36*, 332-343.

[7] Klatt, D. H. (1979). Speech perception: A model of acoustic phonetic analysis and lexical access. *Journal of Phonetics, 7,* 279-312.

[8] Miller, J. L., Baer, T. (1983). Some Effects of Speaking Rate on the Production of /b/ and /w. *Journal of the Acoustical Society of America, 73*(5), 1751-1755.

[9] Miller, J. L., Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25*(457-465).

[10] McMurray, B. (in preparation). KlattWorks: A [somewhat] new systematic approach to formant-based speech synthesis for empirical research.

[11] Pickett, Blumstein, S. E., Burton, M. W. (1999). Effects of speaking rate on the singleton/geminate contrast in Italian. *Phonetica, 56*(3-4), 135-157.

[12] Pind, J. (1986). The perception of quantity in Icelandic. *Phonetica, 43*, 116-139.

[13] Pind, J. (1999). Speech segment durations and quantity in Icelandic. *Journal of Acoustic Society of American, 106*(2), 1045-1053.

[14] Summerfield, A. Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*(5), 1074-1095.

[15] Sussman, H. M., McCaffrey, H. A., Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America, 90*(3), 1309-1325.