Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Duration of Japanese singleton and geminate obstruents in two- to four-mora words

Yukari Hirata[a], Hiroaki Kato[b] and Keiichi Tajima[c]

[a]Colgate University, 13 Oak Drive, Department of East Asian Languages and Literatures, Hamilton, NY 13346, USA
[b]ATR-Promotions, 2-2-2 Hikaridai, Seika-cho Soraku-gun, 619-0288 Kyoto, Japan
[c]Hosei University, 2-17-1 Fujimi, Chiyoda-ku, 102-8160 Tokyo, Japan
yhirata@mail.colgate.edu

Hirata and Whiton [J. Acoust. Soc. Am. 118, 1647-1660 (2005)] found an invariant durational structure for Japanese stop quantity distinction in two- and three-mora words across different speaking rates. The present study examined whether their finding extends to include fricative and affricate quantity distinctions and three- and four-mora words. Stimuli were stop, fricative, and affricate contrasts in four types of words, including (1) a long vowel (e.g., ka.so.o vs. ka.s.so.o), (2) a moraic nasal (e.g., ji.se.n vs. ji.s.se.n), (3) CV sequence (e.g., ho.so.ku vs. ho.s.so.ku), and (4) shorter words (e.g., i.shi vs. i.s.shi), spoken in isolation at three speaking rates by four native Japanese speakers. Duration of contrasting obstruents, words, and the interval between the onsets of the first and the second vowels (VOI1) was measured. Results indicated durational patterns similar to those found previously for stop contrasts. In addition, duration of words, regardless of their segmental composition, reflected well the number of moras they contained. Finally, the ratio of the contrasting obstruent to the word and the ratio of the VOI1 to the mean mora duration were useful in classifying the singleton and geminate categories across rates. [Supported by Grant-in-Aid for Scientific Research, JSPS]

1 Introduction

This study aims at best characterizing the durational structure of Japanese singleton and geminate consonants in various lengths of words spoken in isolation at three speaking rates. The study searches for an invariant durational structure in spite of variability and specific effects caused by factors such as speaking rate, word length, and syllable types. Previous studies have reported various durational effects of consonants, vowels, and syllable types, as well as effects of pitch accent and other non-durational variables on the acoustic structures of singleton and geminate consonant distinctions. However, it is not clear whether there is a single invariant durational structure despite these various effects. Searching for an invariant structure is important as it addresses not only theoretical issues such as testing the acoustic invariance theory [13], but also issues related to non-native speakers' acquisition of Japanese speech timing [4, 8, 14]. If there are only specific effects and ad hoc rules, we might wonder whether non-native speakers have to learn all these specific rules separately, or whether one single general rule exists and this rule can encompass minimal and necessary tactics that enable them to make these quantity distinctions accurately. The latter case would be more useful and parsimonious from non-native speakers' practical points of view. It is possible that effective training methods for non-native speakers may simply fall out of our complete understanding of native Japanese speakers' production and perception mechanisms.

The present study examined two durational ratios as well as absolute duration of obstruents (stops, affricates, and fricatives) contrasting in quantity. The first ratio examined was duration of contrasting consonants over the entire word (C/W ratio). The major purpose of examining C/W ratio was to determine whether it can serve as a reliable indicator for the distinction of singletons and geminates. In [5], C/W ratio was the best indicator of stop quantity distinction among the several measures examined. They proposed that C/W ratio is a linguistically meaningful unit that is predictable from the traditional mora weighting. For example, C/W ratio of a singleton consonant in a CV.CV word should be 0.25 (= 0.5 mora/4 moras), and that of a geminate consonant in a CV.C.CV word should be 0.5 (= 1.5 moras/3 moras; a geminate is expected to weigh one mora plus half of the next mora [2]). Thus, the C/W ratio boundary that best classifies the two categories is expected to be somewhere in between. The actual data showed that the C/W ratios were 0.24 for singletons and 0.48 for geminates, and that the optimal boundary of 0.35 classified the two categories with 96-98% accuracy. However, [5] examined only quantity distinction for stops (/t/-/tː/ and /k/-/kː/) and only in disyllables (i.e., two- and three-mora words) spoken in sentences. The present study sought to find whether the results are replicable in different contexts, examining a wider variety of obstruents contrasting in quantity (/t(ː) k(ː) ʧ(ː) ʤ(ː) ʃ(ː) s(ː)/) in longer words containing three- and four-moras. According to the traditional mora weighting, we would expect C/W ratio to be 0.167 (= 0.5 moras/3 moras) for singletons in three-mora words (CV.CV.CV, CV.CV.V, CV.CV.N; N=moraic nasal), and 0.375 (= 1.5 mora/4 moras) for geminates in four-mora words (CV.C.CV.CV, CV.C.CV.V, CV.C.CV.N). We would expect, then, that the optimal boundary to be around 0.27. The present study examined whether these predicted ratios were indeed found in our data, and whether this unit can serve as a good indicator of quantity distinction across different speaking rates.

The second ratio examined was the duration of a vowel onset interval over the mean mora duration (VOI1/Mmora). The mean mora duration was calculated by the word duration divided by the number of moras it contained. If vowel onsets form the fundamental unit that clearly indicates equal beats of moras in Japanese, the VOI1/Mmora values should roughly correspond to 1 and 2 for those including singletons and geminates, respectively. In this vowel onset mora hypothesis, "VC" is a basic unit of timing in CVCV… sequence, and contrasts with the traditional notion that "CV" is a unit of mora that should indicate equal length. There are several reasons for investigating this "VC" unit as a possible index of durational invariance [9, 10]. In [9] vowel onsets, but not offsets, were found to be used in native speakers' detection of speaking rates when hearing Japanese words. [10] also noted the crucial role of VC as a perceptual timing unit for Japanese. Other studies [1, 3] examined the vowel onset intervals in native speakers' production of Japanese disyllables, and found that they reflected the structure of moras clearly. The present study extended this investigation to include words of longer lengths.

In addition to the two ratios described above, the present study examined absolute duration of words and consonants contrasting in quantity. Absolute duration is useful to measure because it is expected to reflect speaking rate variations, as well as to indicate the amount of overlap between the contrasting consonant duration.

2 Methods

2.1 Stimuli

Stimuli were 58 words that contrast in consonant quantity. These words fell into four word types: CV, VV, N, and SH.

CV included 16 three-syllable words: CV.(C.)CV.CV. The singleton words in this word type included three moras in them, and the geminate words included four moras. The geminate consonant always appeared between the first and the second syllables, e.g., /hos:oku/.

VV included 24 two-syllable words in which the last vowels were always long: (C)V.(C.)CV.V. The singleton words in this word type had three moras, e.g., /kaso:/, and the geminate words had four moras, e.g., /kas:o:/.

N also included 8 two-syllable words in which the last segment was always the moraic nasal: (C)V.(C.)CV.N. The singleton words had three moras, e.g, /iken/, and the geminate words had four moras, e.g., /ik:en/.

SH (which stands for shorter numbers of moras; [(C)V.(C.)CV]) included 10 two-syllable words, and the singleton and geminate words had two and three moras, respectively, e.g., /satʃi/ and /satʃ:i/.

All of these four types had pairs of words that contrast in stops, affricates, and fricatives, except for type N lacking affricate pairs. These words were spoken by two female and two male native speakers of Japanese who are professionally trained. These speakers spoke all of the above words in isolation at three speaking rates: slow, normal, and fast. Thus, a total of 696 tokens (= 29 pairs x 2 lengths x 4 speakers x 3 rates) were analyzed.

2.2 Analyses

For each of the above 696 tokens, duration of words, contrasting consonants, and the interval between the onsets of the first and the second vowels (VOI1) was measured. For details of measurement criteria, see [6]. The criteria were similar to those used in [5] to which this study was compared, but there was one notable difference between the two studies. Duration of stop consonants /t(:), k(:)/ in the present study included the stop closure and voice onset time (VOT), but in [5], it only included the stop closure. VOT does not significantly differ between singleton and geminate stops, and Japanese VOTs are middle-lag [12]. Therefore, this measurement difference would yield only small differences, if any, in comparing the two studies.

Two ratios were calculated: C/W ratio and VOI1/Mmora. C/W ratio was the durational ratio of the contrasting consonants to the entire words. VOI1/Mmora was the ratio of VOI1 to the mean mora duration. The mean mora duration was calculated by dividing the word duration by the number of moras in the word.

3 Results

3.1 Word duration

Port et al. [11] and others have shown that duration of words in Japanese corresponds closely with the number of moras the words contain. There have been abundant data showing that duration of words increases proportionately to the number of moras, and not to the number of syllables. From these previous studies, we predicted for our data that word duration should be longer for the greater numbers of moras in the words (i.e., four-mora words > three-mora words > two-mora words). We also predicted that duration should be similar within the words of the same number of moras, despite the differences in their segmental compositions. That is, duration of three-mora words, for example, (C)V.C.CV (including geminates), CV.CV.CV (all short segments), CV.CV.V (including a long vowel), and CV.CV.N (including the moraic nasal), should all be similar. These predictions were indeed found to be the case, as Fig. 1 indicates. Within the three-mora words, word type VV, i.e., (C)V.CV.V, was longer than the other word types, and word type N had the shortest word duration. A similar pattern was found for the four-mora words. However, these differences across word types were rather small, compared to the differences by the number of moras of the words.

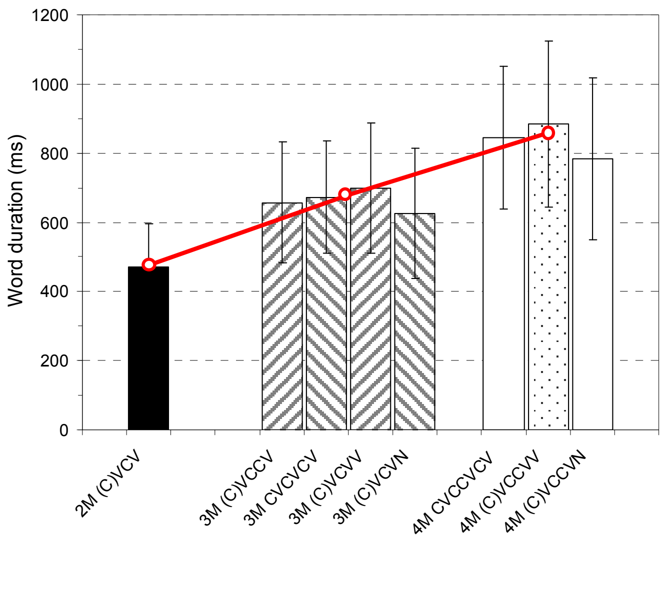

Fig.1 Duration of words as a function of the number of moras. Dark, striped, and white bars represent duration of words with two, three, and four moras (2M, 3M, and 4M), respectively. Vertical bars represent one standard deviation from the mean. The diagonal line across the entire panel shows the word duration mean for each number of moras.

3.2 Contrasting consonant duration

3.2.1. Effects of rate, quantity, and word type

Analysis of variance (ANOVA) was conducted on duration of consonants that contrast in quantity. Word Type (CV, VV, N, SH), Rate (slow, normal, fast), and Quantity (singleton, geminate) were repeated-measures factors. A significant main effect of Quantity assured that geminate consonants were, on average, significantly longer than singletons. A main effect for Rate was found, indicating that consonant duration reflected the three distinct speaking rates: duration of consonants was longer for the slow than normal rate, and longer for the normal than fast rate. In addition, a significant Rate x Quantity interaction was found. The durational difference between the geminates and singletons was greater for slower speaking rates (Fig. 2). The geminate consonants lengthened with slower speaking rates to a greater extent than the singletons did.

The ANOVA also showed a significant effect of Word Type. The mean duration of the contrasting consonants was 242 ms for VV, 241 ms for SH, 229 ms for N, and 206 ms for CV. Multiple comparisons (LSD) indicated that the consonant duration for word type CV was significantly shorter than that in all the other word types. The consonant duration in word type VV was longest among the four, and it was significantly longer than that in N and CV, but not significantly longer than that of word type SH. There was no significant difference between the consonant durations for word types SH and N. However, the durational differences among the word types were smaller, compared to the other effects. For example, all of the mean durations were longer for VV than CV (the left 12 bars), but this difference was small, compared to the robust Rate and Quantity interaction.

Another point worthy of note in Fig. 2 is that absolute duration of geminates and singletons can overlap when comparing across the three rates. For example, the mean duration overlaps between the fast geminates and the slow singletons approximately in the 180-220 ms range. Thus, if we are to search for invariant durational structures that are reliably associated with singletons, on the one hand, and geminates, on the other, absolute duration is not a good indicator. This problem motivates examination of durational ratios as other (and perhaps better) indices.

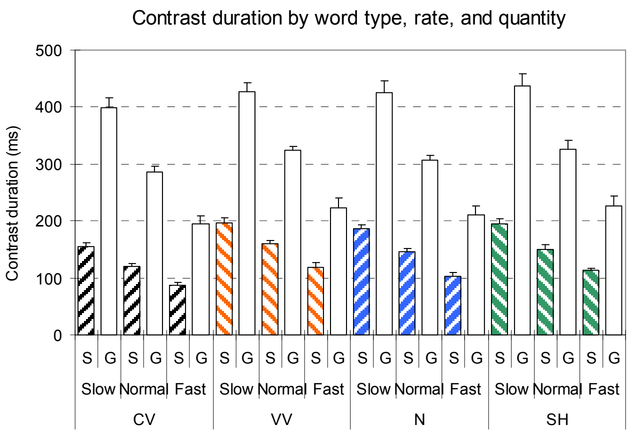

Fig.2 Contrast duration by word type (CV/VV/N/SH), rate (slow/normal/fast), and quantity (S=single / G=geminate).

3.2.2. Optimal boundaries for contrasting consonant duration

Before moving on to the examination of ratios, it is important here to determine how well the contrasting consonant duration alone can serve as an indicator for classifying the singleton and geminate categories across different speaking rates. For this, optimal boundaries were computed by counting the number of misclassified tokens according to each criterion millisecond value within the range of overlap between singletons and geminates. We determined optimal boundaries separately for each word type in order to find whether similar values could be obtained. The results are shown in the top third rows of Table 1. Classification accuracy was in the range of 87.8-90.6% if absolute duration alone was used to classify the single and geminate categories. The optimal boundaries differed across different word types. These results will be compared with those based on ratio values in subsequent sections.

	Word Type	Optimal Boundary	Percentage Misclassified	Classification Accuracy
Conso-nant Dur. (ms)	CV	148-150	9.4%	90.6%
	VV	203-204	12.2%	87.8%
	N	205-208	10.4%	89.6%
	SH	250-253	11.7%	88.3%
C/W ratio	CV	0.25	1.0%	99.0%
	VV	0.30	5.6%	94.4%
	N	0.31-0.34	2.1%	97.9%
	SH	0.41 or 0.42	11.7%	88.3%
VOI1/ Mmora	CV	1.32-1.51	0%	100%
	VV	1.48-1.50	0.7%	99.3%
	N	1.54-1.59	1.0%	99.0%
	SH	1.63	0%	100%

Table 1 Optimal boundaries for contrasting consonant duration, C/W ratio, and VOI1/Mmora computed in each of the four word types

3.3 C/W ratio

3.3.1. Effects of rate, quantity, and word type

The first two goals of examining C/W ratio were (1) to determine whether and how this unit is affected by the speaking rate and the word type, and (2) to determine whether the predicted ratios of 0.17 and 0.38 were indeed found in our actual data of three- and four-mora words, as well as the predicted ratios of 0.25 and 0.50 for two- and three-mora words. ANOVA was conducted with Word Type (CV, VV, N, SH), Rate (slow, normal, fast), and Quantity (singleton, geminate) as repeated-measures factors. Results indicated that the two factors, Quantity and Word Type, significantly affected the C/W ratios. As expected, values were consistently greater for geminates than singletons. Since word type SH was of two- and three-moras and the other word types were of three- and four-moras, we had expected that the C/W ratios were greater for SH than for the others, and this was found to be the case. Among word types, CV, VV, and N, the mean C/W ratios were greatest for N (0.32), second greatest for VV (0.30), and smallest for CV (0.26). A significant Word Type x Quantity interaction was found. Fig. 3 shows the mean C/W ratios for singletons and geminates for each word type. The dashed lines show the predicted ratios from the traditional mora weighting for two/three-mora words (right; 0.25 for singletons; 0.50 for geminates) and for three/four-mora words (left; 0.17 for singletons; 0.38 for geminates). The values for singletons were somewhat greater than expected, but those for geminates were close to the expected values. The solid horizontal lines in Fig. 3 show the expected boundaries between the singleton and geminate categories for two/three-mora words (right) and for three/four-mora words (left). ANOVA also revealed no significant main effect for Rate, indicating that speaking rate did not have a major role in characterizing the quantity distinction with C/W ratios. This is a good thing if we are to look for an invariant, reliable classifier in the face of speaking rate variations. A Rate x Quantity interaction was found to be significant, indicating that the C/W ratios were affected by speaking rates more for geminate words (0.42, 0.40, and 0.39 for slow, normal, and fast rates, respectively) than for singleton words (0.24, 0.24, and 0.25). However, as these values indicate, the difference was quite small.

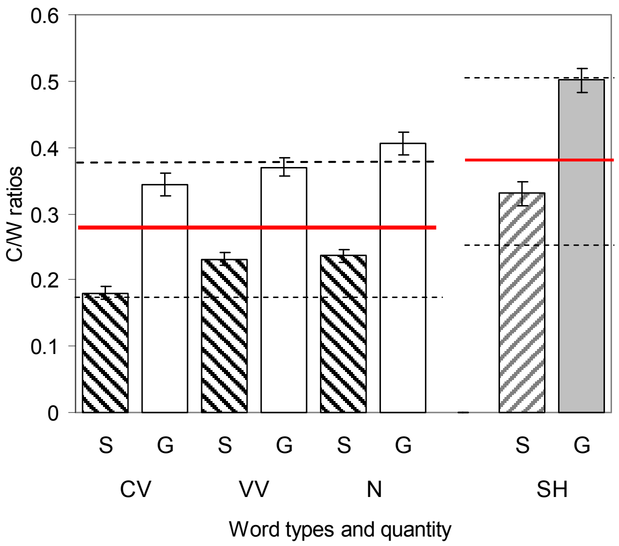

Fig.3 C/W ratios for each word type. The two horizontal dashed lines above CV, VV, and N represent the predicted ratios of 0.17 and 0.38 for singletons and geminates in three- and four-mora words, respectively, and the solid line represents the predicted boundary of 0.27. The two horizontal dashed lines above SH represent the predicted ratios of 0.25 and 0.50 for singletons and geminates in two- and three-mora words, respectively, and the solid line represents the predicted boundary of 0.38.

3.3.2. Optimal boundaries for C/W ratio

The third goal of examining C/W ratio was to determine whether this unit can serve as a good indicator for accurately classifying the singleton and geminate categories across different speaking rates. Just as done for absolute duration in the previous section, optimal boundaries for C/W ratio were computed. Table 1 (in the middle rows) shows the classification accuracy with C/W ratio optimal boundaries for four word types. The highest accuracy of 99.0% was obtained for CV word type. Compared with the accuracy based on absolute duration, C/W ratio served as a better indicator of the quantity distinction for word types CV, VV, and N, and this result is consistent with those in [5]. Compared with the predicted C/W optimal boundary of 0.27 for three- and four-mora words, the obtained optimal boundaries were somewhat greater for VV (0.30) and N (0.31-0.34), and smaller for CV (0.25).

The results of SH word type differed from the other three word types in that its classification accuracy did not improve with C/W ratios, compared to the accuracy based on absolute duration values (both 88.3%). With the optimal boundary of 0.41, eight out of 60 singleton tokens would be incorrectly classified because their values were above 0.41. Those eight were tokens of the word /iʃi/. The inherently long duration of /ʃ/ and the inherently short duration of /i/ in this word must have contributed to the greater-than-average C/W ratios. Some of these discrepancies between the present results and those of [5] might be the contexts in which these words were spoken (isolated words in the present study vs. words in a carrier sentence in [5]). Overall, a follow-up study with a greater number of words with more balanced segments might clarify this inconsistency.

3.4 Vowel onset interval / mean mora

3.4.1. Effects of rate, quantity, and word type

For VOI1/Mmora, a significant main effect of quantity was found, and indicated that the ratio was greater for geminates than singletons: 2.01 vs 1.11. Note that these mean values, respectively, roughly correspond to their numbers of moras in the intervals, i.e., 2 vs. 1. A main effect of rate was not significant, indicating that the above mean ratios were not affected by speaking rates. There was a significant main effect for Word Type. The VOI1/Mmora ratios were 1.68 for word type N and 1.61 for word type SH, significantly greater than 1.56 for word type VV. The ratio for word type VV, in turn, was significantly greater than word type CV (1.41). Note that these mean VOI1/Mmora ratios were the average of geminate and singleton words, and therefore, falling between 2 and 1. There was also a significant Word Type x Quantity interaction. Fig. 4 shows the mean VOI1/Mmora values separately for singleton and geminate words for each word type. The difference between the values of singleton and geminate words was greatest for word type N (0.96), then second greatest for word type CV (0.92), third for word type SH (0.90), and the smallest for word type VV (0.84). As seen in Fig. 4, however, all of the word type values for singleton words were close to 1 (this predicted value is marked with the lower dashed line), and those for geminate words were close to 2 (the upper dashed line), and the differences were those of a small degree. Fig. 4 also shows the boundary of 1.5 (marked with a solid line) between the singleton and geminate words. The boundary line clearly distinguished the two categories.

3.4.2. Optimal boundaries for VOI1/Mmora

To further investigate how accurately VOI1/Mmora can distinguish the two categories of singleton and geminate words, optimal boundaries were computed. Investigation of optimal boundaries was concerned with accurately classifying not only the means but also every token into the two categories. Table 1 (the bottom rows) shows the results of optimal boundary for each of the word type. The optimal boundaries for all of the four word types were close to 1.5, and with those boundary values, the two categories of singletons and geminates were classified with 99.0-100% accuracy. This accuracy is higher than those obtained from C/W ratio and absolute duration of contrasting consonants (upper rows in Table 1). This means that VOI1/Mmora is the most useful measure in characterizing the singleton and geminate categories across different speaking rates and different word types.

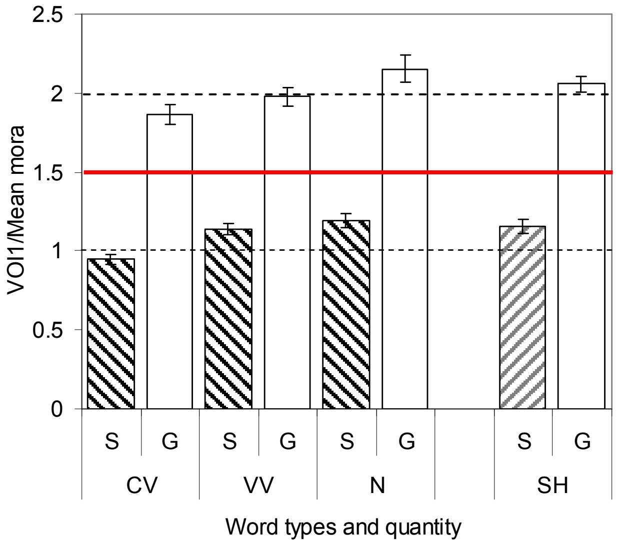

Fig.4 The mean values of VOI1/Mean Mora for all word types. The lower and upper dashed lines represent the predicted ratios of 1 and 2 for singletons and geminates, respectively, and the middle solid line represents the predicted boundary of 1.5.

4 Conclusions

In summary, absolute duration of words and contrasting consonant was well structured within each rate, as well as reflecting their speaking rates. There was also a significant amount of overlap between singleton and geminate words, although this overlap was not as much as that found in [5]. An overlap in absolute duration is problematic in characterizing two phonemic categories of singletons and geminates. However, less overlap was found when C/W and VOI1/Mmora ratios were examined. Accuracy with C/W ratio was higher than that with absolute duration of contrasting consonants in three of the four word types (CV, VV, and N). Furthermore, VOI1/Mmora classified the two phonemic categories with highest accuracy (99.0-100%) for all of the four word types. Although small differences of VOI1/Mmora were found across word types, it worked well to classify the two categories clearly. Further studies are necessary to examine this production results in comparison with perceptual proclivity [7].

Acknowledgments

This study was supported by Grant-in-Aid for Scientific Research, JSPS.

References

[1] M. C. Brady, R. F. Port, K. Nagao, "Effects of speaking style on the regularity of mora timing in Japanese", *J. Acoust. Soc. Am.* 120(5), 3208 (2006)

[2] M. Han, "The feature of duration in Japanese", *Onsei no Kenkyuu* 10, 65–80 (1962)

[3] Y. Hirata, C. Forbes, "Vowel-onset interval as a timing unit for singleton/geminate stop distinction in Japanese", *Proc. 16th International Congress of Phonetic Sciences* 349-354 (2007)

[4] Y. Hirata, L. Whitehurst, E. Cullings, "Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates". *J. Acoust. Soc. Am.* 121(6), 3837-3845 (2007)

[5] Y. Hirata, J. Whiton, "Effects of speaking rate on the single/geminate stop distinction in Japanese", *J. Acoust. Soc. Am.* 118(3), 1647-1660 (2005)

[6] T. Kamiya, "ラベリング作業のための分節の定義", Unpublished document at ATR (2007)

[7] H. Kato, M. Tsuzaki, "Evidence for functional differences between rise and fall markers in discrimination of auditory filled durations", *J. Acoust. Soc. Japan* 19, 73-76 (1998)

[8] H. Kato, K. Tajima, A. Rothwell, R. Akahane-Yamada, K. Munhall, "Perception of phonemic length contrasts in Japanese with or without a carrier sentence by native and non-native listeners", *Proc. 18th International Congress on Acoustics* I-609-612 (2004)

[9] H. Kato, M. Tsuzaki, Y. Sagisaka, "Functional differences between vowel onsets and offsets in temporal perception of speech: Local-change detection and speaking-rate discrimination", *J. Acoust. Soc. Am.* 113(6), 3379-3389 (2003)

[10] E. Ofuka, Y. Mori, S. Kiritani, "Perception of a Japanese geminate stop: the effect of the duration of the preceding/following vowel", *J. Phonetic Soc. Japan* 9(2), 59-65 (2005)

[11] R. F. Port, J. Dalby, M. O'Dell, "Evidence for mora timing in Japanese", *J. Acoust. Soc. Am.* 81(5), 1574–1585 (1987)

[12] T. J. Riney, N. Takagi, K. Ota, Y. Uchida, "The intermediate degree of VOT in Japanese initial voiceless stops", *J. Phonetics* 35, 439-443 (2007)

[13] K. H. Stevens, S. E. Blumstein, "Invariant cues for place of articulation in stop consonants", *J. Acoust. Soc. Am.* 64(5), 1358-1368 (1978)

[14] K. Tajima, H. Kato, A. Rothwell, R. Akahane-Yamada, K. G. Munhall, "Training English listeners to perceive phonemic length contrasts in Japanese", *J. Acoust. Soc. Am.* 123(1), 397-413 (2008)