THE BOUNDARY OF THE QUANTITATIVE CONTRAST BETWEEN SINGLE/GEMINATE CONSONANTS IN PRODUCTION AND PERCEPTION

Yoko Otaki

School of Culture, History and Language, The Australian National University, Australia
yoko.otaki@anu.edu.au

ABSTRACT

The current paper investigated correlativity of acoustic cues to distinguish between single and geminate consonants both in native Japanese production and perception. An analysis revealed that the robust parameters that discriminate the longer length categories from their shorter counterparts in production do not play the same role as decisive factors that they do in perception. Native speakers' production of the quantitative contrast exhibited more standardized manner, whereas their perception reflected a wider range of individual variation than in the production. This indicates that the production and perception are not completely identical.

Keywords: quantitative contrast, relative durational value, native Japanese production and perception

1. INTRODUCTION

On the basis of the isochronisms of Japanese morae, a geminate stop is assumed to be twice as long as their shorter counterpart, a single consonant, in the production. Many research contributions [1, 3, 4] have been made to clarify the mechanism of the production and perception of the contrast.

It has been consented that the closure duration is a fundamental parameter to distinguish a single stop and a geminate stop. For example, Hirata and Whiton [5] analyzed native Japanese production of single and geminate stops in di-syllabic words embedded in a carrier sentence over three speech rates (fast, normal and slow). Although, the contrast of actual durational values between single and geminate stops was clearly maintained *within* one speech rate, the contrast was not robust *across* speech rates. That is, overlaps in the durational values of the two length categories were observed. This confirmed that the contrast is maintained relative to the speech rate. Several parameters were measured from the target stimuli such as closure duration, preceding vowel duration, word duration to determine the most robust parameter that distinguishes a single stop from the longer counterpart. They concluded that closure to word duration ratio serves as the most reliable parameter to differentiate the two length categories. The closure-word ratio of 0.35 classified the two length categories with 95.7-98.0% accuracy.

There have been many studies on the perception of the quantitative contrast, however only a few studies investigated this perception in relation to the production. Assuming that the production and perception are closely related to each other is natural; as Liberman and Mattingly [6] argued that 'if speech perception and speech production share the same set of invariants, they must be intimately linked'. However, contrary to this assumption, Amano, et al. [1] found that the production boundary between a single and a geminate stop obtained from Hirata and Whiton [5] did not coincide with the closure-word ratio obtained at the perceptual boundary. This indicates that mapping between production and perception is not straightforward in terms of the quantitative contrast in Japanese vowels and consonants. Notwithstanding the disaccord of the closure-word ratio of the production and perceptual boundary, the closure-word ratio at the perceptual boundary was consistent within each stimulus word. Moreover, since Amano, et al. [1] used a systematically modified speech rate, what we do not know is how production and perception are related to each other as a function of natural speech rate.

Thus, the current study addresses the question of whether the closure-word ratio at the perceptual boundary is consistent for each word item even in more natural speech environments over different speech rates.

2. METHOD

2.1. Participants

30 native Japanese (17 females and 13 males), their ages range from 24 to 48, participated in the experiment. All of them stayed in Australia for certain periods of time. However, they all claimed that Japanese was their first language regardless of the duration of their stay in Australia.

2.2. Materials

2.2.1. Original materials

Two di-syllabic nonsense words, /seta/ and /kake/ (CVCV) and their variant (CVCCV [three-mora with a geminate consonant]) were selected as stimulus set. Each word (/seta/ and /kake/) for both variants (CVCV and CVCCV) were read with two types of pitch contours (either falling or raising) embedded in the carrier sentence: */korewa ____ dayo (this is ___)/*, in order to prevent the final vowels of these target words from being devoiced or weakened in pronunciation. As it is shown in Table 1, 8 words were used to elicit sound stimuli.

Table 1: Variations of /seta/ and /kake/ for elicitation.

Types / Words	CVCV	CVCCV
/seta/	/seta/ HL	/setta/ HLL
	/seta/ LH	/setta/ LHH
/kake/	/kake/ HL	/kakke/ HLL
	/kake/ LH	/kakke/ LHH

These words were read by a native Japanese speaker (24-year old female) in three different speech rates (fast, normal and slow) whose definitions are quoted from Port [7]. The 'fast' rate was indicated to be the 'fastest tempo possible without making an excessive number of errors'. The 'normal' rate was instructed to be a rate at which a speaker can feel 'relaxed and comfortable'. The 'slow' rate was defined as the 'slowest tempo possible that the speaker could produce' without having a strange pause between segments. The sound stimuli were recorded and digitized with *Audacity 1.3.12.* with sampling rate of 16000Hz in a soundproof room.

2.2.2. Synthesized materials

The sound stimuli for /setta/ and /kakke/ (stimulus words that contain a geminate stop) with two different tonal contours (HLL and LHH) were chosen individually from the three speech rates (slow, normal and fast) (yields 12 sets). Each stimulus was synthesized with *Praat* [2] for the geminate closure durations (/setta/ and /kakke/). The target durations were shortened from the original durations systematically with the *Praat* durational control panel from 0.9 (90% length of the original duration) to 0.1 (10%) at intervals of 0.1, and thus produced 9 modified sounds with different target durations (shortened at 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 and 0.1). According to the result obtained from a pilot study, seven modified sounds (including the original unmodified sounds) at which natives' perception of the quantity shifts were selected out of the nine modified stimuli. These sound stimuli were recorded in three CDs and the stimulus presentation order was randomized for each CD.

2.3. Procedures

The participants were asked to circle their answer on a paper questionnaire form. Response options were two minimal (one pair with double consonants and one without) pairs such as /seta//kake/ (singleton), and /setta//kakke/ (geminate), and the choices were given according to the sound stimulus the participants listened to. For example, if the sound stimulus they listened to is one of the derivations of /seta/, the two minimal pairs of /seta/ sets (/seta/ and /setta/) were provided for the options. The participants were divided into three groups and a different CD set was assigned to each group. Please note that number of people in each group is not equal. Six participants were assigned CD1, nine participants were assigned CD2 and 15 participants were assigned CD3. The three separate experiments were conducted in a quiet room and each CD was played with two speakers (*Sony SRS-A202*).

3. RESULTS

The results of the questionnaire were analyzed for all the participants en bloc and translated into graphs with logistic-curves. In the analysis, the main focus was on the shift in native speakers' perceptual boundary between single and geminate stops. The perceptual boundary between the two length categories was obtained where more than 50% of the participants' (more than 15 out of the 30 participants) perception shifts on the logistic curve for each word type and modified duration

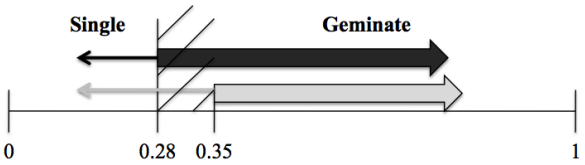
rate was recorded as shifting from a geminate to a single consonant. As for the perceptual boundary between single and geminate closure duration, three parameters and two ratios were measured; duration of a target closure, duration of the preceding vowel of the target closure, word duration, the closure-preceding vowel ratio and the closure-word ratio shown in Table 2.

Table 2: Possible parameters measured to distinguish a single stop and a geminate stop

Word	Pitch	Speech rate	Preceding vowel duration (ms)	Closure duration at perceptual boundary (ms)	Word duration (ms)	Closure-vowel ratio	Closure-word ratio
/seta/	HLL	fast	56	62	276	1.11	0.22
		normal	79	82	355	1.04	0.23
		slow	126	123	506	0.98	0.24
	LHH	fast	46	71	263	1.54	0.27
		normal	72	89	363	1.24	0.25
		slow	118	146	402	1.24	0.36
/kake/	HLL	fast	53	64	200	1.21	0.32
		normal	74	72	247	0.97	0.29
		slow	165	124	477	0.75	0.26
	LHH	fast	59	65	202	1.10	0.32
		normal	76	67	266	0.88	0.25
		slow	137	132	459	0.96	0.29

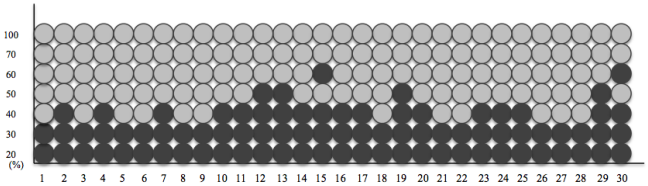
The result shows that the perceptual boundary between a single and geminate stop observed in the current experiment contradicts the production boundary obtained from Hirata and Whiton [5]. The average closure-word ratio at the perceptual boundary is 0.28, whereas that of the production boundary was determined as 0.35. This indicates that the closure-word ratio of the perceptual boundary does not coincide with that of the production boundary, which was also confirmed by Amano, et al. [1]. More importantly, closure duration has to be shorter to be perceived as a single stop in perception than in production. As Figure 1 shows, a sound stimulus lying in the gray area (closure-word ratio between 0.28 and 0.35) will be produced as a single stop and perceived as a geminate stop.

Figure 1: A gap in the closure-word ratio at the production [gray arrows] and perceptual boundary [black arrows]



However, consistency was not observed in the closure-word ratio within each word category *across* speech rates. This result contradicts with the findings from Amano, et al. [1] that closure-word ratio at the perceptual boundary was very consistent over time-stretch rates within each word category (0.2 difference at most). Amano et al. [1] controlled speech rates systematically by shortening and extending the duration of the vowels which precede the target closure duration (preceding vowels) and the target closure duration. Additionally, stimuli were presented without a carrier sentence in Amano, et al. [1]'s study. In natural speech in contrast with synthetic speech, segmental durations in an utterance change in accordance with speech rate and neighboring phonological and phonetic environment. The change is neither controllable nor predictable due to differences in intrinsic duration and acoustic characteristics of a segment, etc. Therefore, the result obtained from the current experiment accurately reflects a fundamental aspect of the dynamic nature of speech production and perception.

Figure 2: Individual perceptual boundary for /setta (LHH)/. (*x axis = individual participants; y axis = ratio of the target duration to original duration*)



The inconsistency in the perceptual boundaries largely reflects individual differences. Figure 2 shows the individual responses for the variables of the /setta (LHH)/ set. Black dots indicate that the token was perceived as having a single stop. Although these sound stimuli were randomized in their order, each participant demonstrated consistency and reliability of their judgment in distinguishing a single stop and a geminate stop. Exceptionally, participant 15 and 30 did not show

a consistent perceptual boundary, however the perceptual boundaries between a single and geminate stop were clearly consistent *within* individuals regardless of rather large differences in the perceptual boundaries observed *between* individuals.

4. DISCUSSION

The results of the current study indicate that the closure-word ratio of the perceptual boundary between a single and a geminate stop do not agree completely with that of the production boundary. This collision in these boundaries was also confirmed by Amano, et al. [1]. Moreover, the closure-word ratio did not appear as a robust parameter to distinguish the quantitative difference across speech rates in perception. This result contradicts with what has been observed in Amano, et al.'s [1] study. However, at the individual level, a consistency in the judgment on the quantitative distinction was observed in spite of a rather large difference between the individual judgments. This indicates that individual differences in their perception between a single and a geminate consonant is not negligible that they are the indicative of the inconsistency in the perceptual boundary overall.

One possible cause of the discrepancy in the production and the perceptual boundary is differences in segments of stimuli words employed for the current experiment and Hirata and Whiton's [5] study. Segmentals undergo a durational change in accordance with the surrounding phonological environments in order to maintain the isochronisms of the morae in Japanese. This phenomenon explains why the consistency observed in the closure-word ratio was restricted within each word item in Amano, et al.'s [1] study. Segmental durations in different phonological environments are not comparable in a strict sense.

In the current experiment, the closure-word ratio at the perceptual boundary did not show consistency across speech rates even within each word set. Hirata and Whiton [5] found that the durational difference between a single and a geminate stop was extensively enhanced in slow speech as compared to fast speech in the production. That is, the durational ratio of a geminate stop to a single stop needs to be bigger in slow speech than in fast speech. This possibly accounts for the inconsistency of perceptual boundaries across speech rates.

Interestingly enough, the results of the current experiment lead to the conclusion that there is an inconsistency in human perception and significant individual differences in perception. The inconsistency indicates the possibility of miscommunication such as a speech production intended to contain a geminate stop being perceived as containing a single stop. This indicated that there is a clear need to provide more precise teaching for the pronunciation in Japanese as a second language education, considering that even native Japanese have difficulties in differentiating the quantitative contrast. Further study is necessary to explain explicitly the mechanism of speech production and perception.

5. ACKNOWLEDGEMENTS

I would like to thank the two anonymous reviewers for their sound advice to improve this paper. I would also like to give my utmost and sincere thanks to Drs. Peter Hendriks and Shunichi Ishihara for their supervision and support for my Master's thesis, which this current paper is based on. Finally, I would like to thank the College of Asia and the Pacific of ANU for supporting me financially to attend this conference and present this paper.

6. REFERENCES

[1] Amano, S., Mugitani, R., Kobayashi, T. 2007. Perceptual boundary between a single and a geminate stop in Japanese. *ICPhS XVI* Saarbrücken, 733-736.

[2] Boerma, P., Weenink, D. 2009. *Praat: Doing phonetics by computer* (version 5.1.07).

[3] Han, M. 1992. The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers. *Phonetica* 49, 102-127.

[4] Hardison, M.D. Saigo, M.M. 2010. Development of perception of second language Japanese geminates: Role of duration, sonority, and segmentation strategy. *Applied Psycholinguistics* 31, 81-99.

[5] Hirata, Y., Whiton, J. 2005. Effects of speaking rate on the single/geminate stop distinction in Japanese. *Journal of Acoustical Society of America* 118, 1647-1660.

[6] Liberman, A.M., Mattingly, I.G. 1985. The motor theory of speech perception revised. *Cognition* 21, 1-36.

[7] Port, F. 1977. *The Influence of Speaking Tempo on the Duration of Stressed Vowel and Medial Stop in English Trochee Words.* Bloomington: Indiana University.