Is the Vowel Length Contrast in Japanese Exaggerated in Infant-Directed Speech?

Keiichi Tajima[1,2], *Kuniyoshi Tanaka*[2], *Andrew Martin*[2], *Reiko Mazuka*[1,3]

[1] Department of Psychology, Hosei University, Tokyo, Japan
[2] RIKEN Brain Science Institute, Saitama, Japan
[3] Department of Psychology and Neuroscience, Duke University, Durham, NC, USA

tajima@hosei.ac.jp, {takana_k,amartin,mazuka}@brain.riken.jp

10.21437/Interspeech.2013-712

Abstract

Vowel length contrasts in Japanese, e.g., *chizu* "map" vs. *chiizu* "cheese", are cued primarily by vowel duration. However, since short and long vowel durations overlap considerably in ordinary speech, learning to perceive vowel length contrasts is complex. Meanwhile, infant-directed speech (IDS) is known to "exaggerate" certain properties of adult-directed speech (ADS). If so, then it is possible that vowel length contrasts might also be exaggerated in IDS. To investigate this, the present study analyzed vowel durations in the RIKEN Japanese Mother-Infant Conversation Corpus, which contains 11 hours of IDS by 22 mothers talking with their 18-to-24-month-old infants, and 3 hours of ADS by the same mothers. Results indicated that vowel length contrasts were generally not exaggerated in IDS, except at the end of prosodic phrases. Furthermore, several factors that systematically affected vowel duration in IDS were identified, including phrase-final lengthening and "non-lexical lengthening", i.e., the lengthening of vowels for emphatic or other stylistic purposes. These results suggest that vowel duration in Japanese IDS could not only potentially facilitate learning of lexical distinctions, but also signal phrase boundaries, emphasis, or other communicative functions.

Index Terms: phonemic length contrast, motherese, exaggerated speech, speech corpus analysis

1. Introduction

In Japanese, vowel length can be used phonemically to distinguish words, e.g., *chizu* "map" vs. *chiizu* "cheese", *isho* "testament" vs. *ishoo* "costume". The primary acoustic correlate and perceptual cue to such vowel length contrasts is the duration of the vowel [1]. Past studies that measured vowel duration in read speech have reported that phonemically long vowels are approximately 2.0-2.5 times longer than phonemically short vowels [2]. However, it has also been shown that the duration of phonemically short and long vowels overlap considerably when produced at various speaking rates [2]. This implies that perceptual judgment of vowel length cannot be made simply based on absolute vowel duration, but instead must be based on more complex criteria, made in relation to the phonetic environment. Therefore, learning to perceive vowel length contrasts may be a complex process.

How do children acquire the vowel length contrast in Japanese? One study that examined Japanese infants' discrimination of short and long vowels in the nonwords /mana/ vs. /maana/ [3] showed that infants were able to discriminate between short and long vowels at 9 months, but not at 4 or 7.5 months. Meanwhile, infants as young as 4 months were able to detect a change in vowel quality, /mana/ vs. /mina/. These results suggest that perceptual abilities for vowel length contrasts may develop somewhat later than those for vowel quality differences.

An alternative approach to the above question may be to examine the language input that children receive during their development. Children learn language by listening to the language(s) spoken by their caretakers. Thus, one way to understand how children acquire language is to study the nature of the linguistic input that children are exposed to. When adults talk to children, they often change the way they talk. The style of speech used when talking to infants is called "motherese", "baby talk", or "infant-directed speech" (IDS).

IDS differs from adult-directed speech (ADS) in many ways, including phonological, morphological, and syntactic properties [4]. Focusing on phonological differences, for example, IDS has been reported to have acoustically more extreme vowels than ADS, resulting in a "stretching" of the vowel space [5]. IDS also has higher pitch, greater pitch range, shorter utterances, and longer pauses than ADS [6]. There is some evidence that properties of IDS have beneficial effects for children learning language. For example, the use of IDS results in better discrimination of syllable sequences [7] and vowels [8] than the use of ADS.

Given that certain phonological properties of ADS are exaggerated in IDS, the present study investigates whether the vowel length contrast is also exaggerated in Japanese IDS, using a large-scale corpus of IDS in Japanese. The present study also examines what factors other than phonemic length might systematically affect vowel duration in Japanese IDS. One study that used the same Japanese IDS corpus as the present study [9] found that the mean duration of long vowels was clearly longer than that of short vowels, but that the distribution of short and long vowel durations overlapped completely, with no evidence for a bimodal distribution. The authors suggest that it would be difficult to learn vowel length contrasts simply based on the distribution of vowel durations. In the present study, the focus will be on examining whether the vowel length contrast is exaggerated in IDS compared to ADS, and whether factors other than phonemic length, such as phrase-level or stylistic factors, also affect vowel duration.

2. Methods

2.1. The RIKEN Japanese Mother-Infant Conversation Corpus

The present study utilized the RIKEN Japanese Mother-Infant Conversation Corpus [10], which contains recordings of Japanese conversations between mothers and their infants, with accompanying phonetic, prosodic, and morphological annotations. The talkers were 22 native Japanese mothers from the Tokyo metropolitan area, aged 25-43 years (mean = 33 years, 8 months), and their 18-to-24-month-old infants.

From each mother, both IDS and ADS were recorded. To record the IDS, mothers were placed in the recording room with their infants and were asked to play with them using picture books for approximately 15 minutes and then using toys for approximately another 15 minutes. To record the ADS, a female experimenter (a 33-year-old mother of a 23-month-old infant) entered the recording room, and the mother and experimenter engaged in free conversations for approximately 10 minutes. The mother's speech was recorded using a headset dynamic microphone, while the interlocutor's speech was simultaneously recorded using a condenser microphone placed on a table in the recording room. In total, the corpus contains approximately 11 hours (50,000 words) of IDS and 3 hours (30,000 words) of ADS.

Recordings in the corpus are accompanied by the following types of annotations, assigned by trained labelers. Labeling guidelines essentially follow those of the Corpus of Spontaneous Japanese [11].

- *Text*: Katakana transcriptions of the mothers' speech are provided. Also provided are various tags that mark filled pauses, mispronunciations, word fragments, singing, laughter, cough, etc. Pauses longer than 200 ms were automatically labeled as utterance boundaries.
- *Phonetic labels*: Phonemic labels that are time-aligned to the speech signal are provided. Some phonetic details such as vowel devoicing are also provided. The corpus also contained many cases of "non-lexical lengthening" of segments, i.e., lengthening that does not signal a lexical distinction but that is done for emphatic or stylistic purposes, e.g., lengthening of the vowel /e/ in the copula *-desu* as in [deːsu̥]. Segments that were perceptually judged by trained labelers to be non-lexically lengthened were marked with special tags.
- *Prosodic labels*: Intonation labels are given in accordance with X-JToBI (Extended Japanese Tone and Break Indices) guidelines [12]. The corpus assumes the following hierarchically related prosodic units, among others: *word*, *accentual phrase* (AP), *intonation phrase* (IP), and *utterance*. A *word* is a word-size unit that contains at most two morphemes. An *accentual phrase* is a prosodic phrase that contains one or more words. It begins with a phrase-initial pitch rise and contains at most one pitch accent. An *intonation phrase* contains one or more accentual phrases and is a prosodic phrase where pitch range is reset. An *utterance* contains one or more intonation phrases and is followed by a pause longer than 200 ms.
- *Morphological information*: Information about part of speech, conjugation, etc is provided for each morpheme-size unit. Onomatopoetic words, which appear often in Japanese IDS, are treated as an independent part of speech category.

2.2. Data analysis

From the speech corpus, the following portions were excluded from data analysis: speech directed to adults during the IDS recording session, speech directed to infants during the ADS session, singing, laughter, cough, filled pauses, and interjections. Segments that did not have clear acoustic boundaries on either or both sides were also excluded from analysis. As a result, about 24.7% of the data were excluded, for a total of 94,041 vowels that were submitted to analysis.

Then, the duration of short and long vowels was obtained from the phonetic labels. Mean duration of vowels was calculated separately for three phrasal positions: (1) words in non-phrase-final position, (2) words in AP-final position, and (3) words in IP-final position. This was done so as to tease apart the effect of phrase-final lengthening from the analysis of vowel duration.

3. Results

3.1. Short vs. long vowels

Figure 1 shows the mean duration of short and long vowels across the 22 talkers as a function of phrasal position (non-final, AP-final, IP-final) and speech style (ADS, IDS). Also shown in Figure 1 are the long-to-short vowel duration ratios.

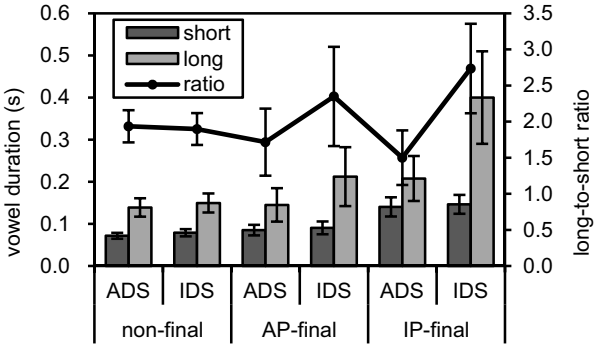

Figure 1: *Mean duration of short and long vowels and mean long-to-short vowel duration ratio as a function of phrasal position and speech style. Error bars indicate standard deviations across the 22 talkers.*

For short vowels, mean duration was 72 ms (ADS) vs. 79 ms (IDS) for non-final words, 85 ms (ADS) vs. 91 ms (IDS) for AP-final words, and 140 ms (ADS) vs. 146 ms (IDS) for IP-final words. A 2-way repeated-measures analysis of variance (ANOVA) with phrasal position (non-final, AP-final, IP-final) and speech style (ADS, IDS) as within-subjects factors and short vowel duration as the dependent variable showed significant main effects of phrasal position [$F(2,42)=299.42$; $p<.001$] and speech style [$F(1,21)=5.73$; $p<.05$].

For long vowels, mean duration was 139 ms (ADS) vs. 150 ms (IDS) in non-final words, 145 ms (ADS) vs. 212 ms (IDS) for AP-final words, and 208 ms (ADS) vs. 400 ms (IDS) for IP-final words. A similar 2-way ANOVA with long vowel duration as the dependent variable revealed significant main effects of phrasal position [$F(2,42)=98.92$; $p<.001$] and speech style [$F(1,21)=66.50$; $p<.001$] and a significant position-by-style interaction [$F(2,42)=30.17$; $p<.001$]. Further analysis of the interaction using simple effects tests indicated no significant difference between ADS and IDS for non-final words, but significantly greater duration in IDS than ADS for AP-final words ($p<.001$) and IP-final words ($p<.001$).

For long-to-short vowel duration ratios, they were 1.93 (ADS) vs. 1.90 (IDS) for non-final words. For AP-final and IP-final words, ADS and IDS showed opposite trends. In ADS, the ratio *decreased* to 1.72 for AP-final words and further to 1.50 for IP-final words. In contrast, in IDS, the ratio

increased to 2.35 for AP-final words and further to 2.74 for IP-final words. The difference between ADS and IDS grew larger as the prosodic boundary strength increased from non-final to AP-final to IP-final. A 2-way ANOVA with long-to-short vowel duration ratio as the dependent variable revealed a significant main effect of speech style [$F(1,21)=41.64$; $p<.001$] and a significant position-by-style interaction [$F(2,42)=19.72$; $p<.001$]. Further analysis of the interaction using simple effects tests indicated no significant difference between ADS and IDS for non-final words, but a significantly higher ratio in IDS than ADS for AP-final words ($p<.001$) and IP-final words ($p<.001$). Moreover, multiple comparisons indicated that the gradually increasing ratios in IDS were all significantly different from one another as follows: non-final < AP-final < IP-final ($p<.05$). For the gradually decreasing ratios in ADS, the following significant difference was found: non-final < IP-final ($p<.05$).

3.2. Other factors affecting vowel duration

Close examination of vowel durations in the corpus suggested that, in addition to phonemic length (short vs. long), there were several factors that substantially affected vowel duration.

3.2.1. Phrase-final lengthening

One such factor was phrase-final lengthening. As witnessed in Figure 1, vowels were slightly longer in AP-final words than non-final words, and considerably longer in IP-final words than AP-final words. In ADS, the magnitude of lengthening was comparable between phonemically short and long vowels. That is, vowels were longer by 13 ms (short vowels) vs. 8 ms (long vowels) in AP-final than non-final words, and by 55 ms (short vowels) vs. 63 ms (long vowels) in IP-final than AP-final words. By contrast, in IDS, the magnitude of lengthening was much greater in long vowels than short vowels. That is, vowels were longer by 12 ms (short vowels) vs. 63 ms (long vowels) in AP-final than non-final words, and by 56 ms (short vowels) vs. 188 ms (long vowels) in IP-final than AP-final words. Note that the standard deviations also increased from non-final to AP-final to IP-final words, indicating that the duration of long vowels was highly variable in phrase-final words. Put together, these results suggest that phrase-final lengthening was generally more pronounced in IDS than ADS.

3.2.2. Non-lexical lengthening

A second factor that affected vowel duration was "non-lexical lengthening". Many vowels in the corpus were lengthened, not to signal phonemically long vowels or phrase final position, but to achieve emphasis or other stylistic purposes. For example, *sotto* "softly, gently" was occasionally produced as [so:tto] with a lengthened /o/ to emphasize its meaning, and *kudasai* "(please) give me" was produced as [kudasa:i] with a lengthened /a/ to convey playfulness or friendliness.

When the percentage of such non-lexically lengthened vowels was calculated for each talker separately for ADS and IDS, the mean percentage was 5.8% (s.d. = 1.1) for ADS and 7.1% (s.d. = 1.6) for IDS. A matched-sample *t*-test indicated that the percentage was significantly higher in IDS than ADS [$t(21)=4.02$; $p<.001$].

To examine how much longer the non-lexically lengthened vowels were in comparison with other vowels, Figure 2 shows the mean duration of phonemically short vowels that did not undergo non-lexical lengthening and those that did undergo lengthening, and phonemically long vowels that did not undergo lengthening. The figure does not include data for phonemically long vowels that were non-lexically lengthened because the entire corpus contained only six such tokens in IDS and zero such tokens in ADS.

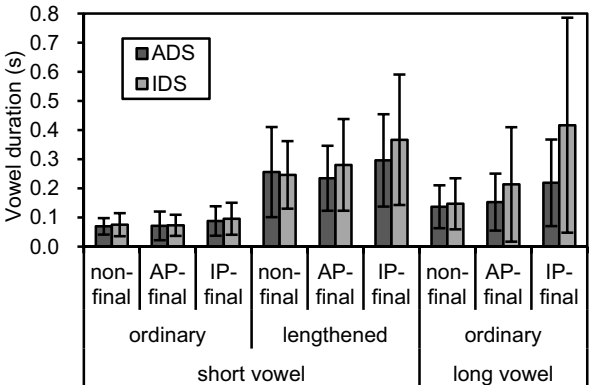

Figure 2: *Mean duration of phonemically short vowels that did not undergo non-lexical lengthening (ordinary) and those that did (lengthened), and phonemically long vowels that did not undergo lengthening (ordinary), as a function of phrasal position and speech style. Error bars indicate standard deviations across all vowel tokens.*

Figure 2 shows that the mean durations of non-lexically lengthened short vowels (mean duration = 280 ms) were clearly longer than those of ordinary short vowels (mean duration = 79 ms), by a factor greater than three. Furthermore, a noteworthy result is that the mean duration of non-lexically lengthened short vowels (280 ms) was longer than the mean duration of phonemically long vowels (214 ms). Standard deviations were also very large, suggesting that the duration of non-lexically lengthened vowels was high variable. These trends were observed in both ADS and IDS.

Further analysis of non-lexically lengthened vowels indicated differences between ADS and IDS regarding where the lengthened vowels tended to occur within a word. Table 1 shows the number (and percentage) of non-lexically lengthened vowels that occurred in word-final position and those that occurred in a non-final (word-initial or word-medial) position, separately for ADS and IDS. In ADS, 95.3% of the 1,877 vowels that were non-lexically lengthened occurred in word-final position, while only 4.7% of the vowels occurred in a non-final position. In contrast, in IDS, non-lexically lengthened vowels occurred very frequently in non-final position; of the 5,437 vowels that were non-lexically lengthened, 39.5% occurred in non-final position, and 60.5% occurred in word-final position. A Chi-square test on the count data in Table 1 revealed that the row and column factors were not independent [$\chi^2(1)=794.04$; $p<.001$], indicating that the percentages of non-lexically lengthened vowels in non-final vs. word-final positions were significantly different between ADS and IDS. Thus, in ADS, non-lexical lengthening occurred predominantly in word-final vowels, whereas in IDS, it occurred almost as frequently in non-final vowels as it did in word-final vowels.

Furthermore, additional analysis of the data in Table 1 suggested differences between the kinds of words that showed

non-lexical lengthening in non-final vowels and those that showed lengthening in word-final vowels. For words that showed non-lexical lengthening in word-final vowels (the third column of Table 1), the following characteristics were found: (1) Most of the words (93.7% in ADS and 91.2% in IDS) were short words of 1 or 2 moras in length. (2) Most words (88.7% in ADS and 76.2% in IDS) were function words such as particles, auxiliaries, and interjections, e.g., *ne* "isn't it?, you know?", *te* (gerundive particle of verbs), *kedo* "but" (underlined vowels were non-lexically lengthened). (3) Most words (93.8% in ADS and 95.3% in IDS) occurred in phrase-final position (AP-final or IP-final). The fact that the vast majority of these words occurred in phrase-final position suggests that many of the vowels in word-final position that were marked as non-lexically lengthened may be cases of extremely strong phrase-final lengthening.

Table 1. *Number (and percentage) of non-lexically lengthened vowels that occurred in non-final and word-final positions, separately for ADS and IDS..*

Speech style	Non-final vowel	Word-final vowel	Total
ADS	88 (4.7%)	1789 (95.3%)	1877
IDS	2146 (39.5%)	3291 (60.5%)	5437

On the other hand, for words that contained non-lexical lengthening in non-final (word-initial or word-medial) vowels, (the second column of Table 1), the following characteristics were found: (1) Fewer words (39.5% in ADS and 40.5% in IDS) were short words of 1 or 2 moras in length. (2) Most words (80.2% in ADS and 74.4% in IDS) were content words such as nouns, adjectives, and adverbs, e.g., *panda* "panda", *nani* "what", *sugoi* "terrific, great". (3) Fewer words (11.1% in ADS and 51.0% in IDS) appeared in phrase-final position. The fact that many of these words did not occur in phrase-final position suggests that they were not lengthened as a result of phrase-final lengthening, but were instead lengthened for other purposes (see discussion in section 4).

Put together, these results suggest that vowels that were marked as non-lexically lengthened in ADS were observed primarily in 1- or 2-mora function words that underwent very strong phrase-final lengthening. In IDS, however, in addition to this type of non-lexical lengthening, it was also frequently observed in word-initial or word-medial vowels within content words; these vowels were presumably lengthened for reasons other than phrase-final lengthening.

4. Discussion

Analysis of vowel durations in a large-scale speech corpus of Japanese IDS conducted in the present study indicated that vowel length contrasts in Japanese were in general *not* exaggerated in IDS compared to ADS. Mothers therefore do not seem to make short and long vowels more acoustically distinct from each other across the board when talking to their infants. However, one notable exception is that the durational distinction between short and long vowels was greater in magnitude in IDS when the vowels occurred in phrase-final words. The long-to-short vowel duration ratio increased in IDS as the prosodic boundary strength increased from non-final to AP-final to IP-final. This trend was opposite from that found in ADS, wherein the same ratio *decreased* as the prosodic boundary strength increased. Such a trend for the vowel length contrast to be weakened in final position has been pointed out previously in ordinary (adult-directed) speech, e.g., [13]. When talking to infants, what the mothers seem to be doing is to counteract this tendency toward weakening in phrase-final words and to acoustically enhance the length contrast just where it tends to be weakened. In fact, all 22 mothers showed greater long-to-short vowel duration ratios in IDS than ADS for IP-final words. Therefore, these results suggest that within certain specific contexts, vowel length contrasts in Japanese are indeed more acoustically distinct in IDS than ADS.

A second finding in the present study is that in addition to phonemic length, vowel duration was systematically influenced by several factors. One factor was phrase-final lengthening. Vowels were increasingly longer as the prosodic boundary strength increased from non-final to AP-final to IP-final. This was observed in both ADS and IDS, but the magnitude of lengthening was stronger for IDS especially for phonemically long vowels.

Another factor was non-lexical lengthening, i.e., the lengthening of vowels for emphatic or stylistic purposes. Many vowels in the present corpus were marked by trained labelers as having undergone non-lexical lengthening. Some of these occurred in word-final vowels in phrase-final position. These may be viewed as instances of extreme phrase-final lengthening, and therefore may not necessarily represent a separate factor influencing vowel duration. However, other cases of lengthening were observed in word-initial or word-medial position, and in phrase-medial positions. These cases occurred more frequently in IDS than ADS, and may reflect attempts by the mothers to place emphasis or achieve other stylistic goals, as a way to draw the attention of the interlocutor (infant) and engage them in communicative interactions. For example, mothers often placed emphasis or focus on words by lengthening one or more vowels in a word, e.g., *minna* "everyone", *miruku* "milk". Also, onomatopoeic words that mimic natural sounds also frequently exhibited lengthening, e.g., *gatan-goton* "clickety-clack (train sound)", *paon* (elephant cry). Finally, mothers often used chanting or other stylized forms of speech that exhibited lengthening, e.g., *itadakimasu* "let's eat", *kuma-chan* ("bear" + diminutive suffix, produced with a vocative chant). These vowels are often realized with a longer duration than phonemically long vowels. This could be problematic when trying to accurately convey the phonemic length of vowels to infants. Further research is needed to determine precisely what factors trigger non-lexical lengthening and how they interfere with accurate transmission of words to the listener.

In conclusion, vowel duration in Japanese IDS seems to serve (at least) a dual function. The first is to facilitate language development, e.g., by exaggerating vowel length distinctions in contexts where they tend to be weakened in ADS, i.e., phrase-final words. The second is to draw infants' interest and engage them in communicative interactions. The two functions often mutually conflict, in which case the latter function often seems to predominate.

5. Acknowledgements

This research was supported by a Grant-in-Aid for Scientific Research (C), No. 23520474, by the Japan Society for the Promotion of Science, granted to the first author.

6. References

[1] Fujisaki, H., Nakamura, K., and Imoto, T., "Auditory perception of duration of speech and non-speech stimuli", in G. Fant. and M. A. A. Tatham [Eds.], Auditory Analysis and Perception of Speech, 197–219, Academic, London, 1975.

[2] Hirata, Y., "Effects of speaking rate on the vowel length distinction in Japanese", J. Phonetics, 32:565–589, 2004.

[3] Sato, Y., Sogabe, Y., and Mazuka, R., "Discrimination of phonemic vowel length by Japanese infants", Devel. Psych., 46:106-119, 2010.

[4] Soderstrom, M., "Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants", Devel. Rev., 27:501-532, 2007.

[5] Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kzhevnikova, E. V., Ryskina, V., L., Stolyarova, E. I., Sundberg, U., and Lacerdo, F., "Cross-language analysis of phonetic units in language addressed to infants", Science, 277:684-686, 1997.

[6] Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I., "A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants", J. Child. Lang., 16:477-501, 1989.

[7] Karzon, R. G., "Discrimination of polysyllabic sequences by one- to four-month-old infants", J. Exp. Child Psych., 39:326-342, 1985.

[8] Trainor, L. J. and DesJardins, R. N., "Pitch characteristics of infant-directed speech affects infants' ability to discriminate vowels", Psychonom. Bullet. and Rev., 9:335-340, 2002.

[9] Bion, R. A. H., Miyazawa, K., Kikuchi, H., and Mazuka, R. "Learning phonemic vowel length from naturalistic recordings of Japanese infant-directed speech", PLoS ONE, 8:e51594, 2013.

[10] Mazuka, R., Igarashi, Y., & Nishikawa, K. "Input for learning Japanese: RIKEN Japanese Mother-Infant Conversation Corpus", Tech. Rep. Inst. Electron. Info. Commun. Eng., TL2006-15, 11-15, 2006.

[11] Maekawa, K., Koiso, H., Furui, S., and Isahara, H., "Spontaneous speech corpus of Japanese," Proc. LREC2000 (Second Int'l Conf. on Lang. Resources and Evaluation) 2:947-952, 2000.

[12] Maekawa, K., Kikuchi, H., Igarashi, Y., and Venditti, J., "X-JToBI: An extended J-ToBI for Japanese spontaneous speech", Proc. ICSLP 2002 (7th Int'l Conf. of Spoken Lang. Proc., 1545-1548, 2002.

[13] Kubozono, H., "Temporal neutralization in Japanese", in C. Gussenhoven and N. Warner [Eds.], Papers in Laboratory Phonology VII, Mouton, Berlin, 171-201, 2002.