

The F0 fall delay of lexical pitch accent in Japanese Infant-directed speech

Yoko Saikachi¹, Mafuyu Kitahara², Ken'ya Nishikawa¹, Ai Kanato¹, Reiko Mazuka^{1,3}

¹Laboratory for Language Development, Brain Science Institute, RIKEN, Japan

²School of Law, Waseda University, Japan

³Department of Psychology and Neuroscience, Duke University, USA

ysaikachi@brain.riken.jp, kitahara@waseda.jp, nisi@brain.riken.jp,
53_like_pe-7-73@fuji.waseda.jp, mazuka@brain.riken.jp

Abstract

The current study examined the acoustic modifications of the lexical pitch accent in Tokyo Japanese infant-directed speech (IDS), with the focus on the F0 fall delay, where the alignment of the F0 turning points associated with pitch accents were delayed with respect to the accented mora. The RIKEN Mother-Infant Conversation Corpus (R-JMICC) [1] produced by 21 mothers from Tokyo area, was used to investigate the alignment of the F0 turning points. Two-piece linear regression was used to locate the turning points and the frequency of F0 fall delay was computed in IDS and in adult-directed speech (ADS). The results revealed that the frequency of F0 fall delay depended on the syllable structures of the accented syllable as well as the prosodic conditions (the presence of the boundary pitch movements and non-lexical lengthening) typically observed in Japanese IDS. We found significantly more frequent F0 fall delay in IDS compared to ADS, when the prosodic conditions were taken into account. The results indicate that the language-specific prosodic structure should be considered in order to characterize the F0 fall delay of lexical pitch accents in IDS.

Index Terms: F0 turning points of pitch accents, Infant-directed speech

1. Introduction

It has been well studied that people change the way of talk in front of infants: infant-directed speech (IDS) [2]. Mainly described characteristics of IDS are exaggerated intonation, slower speech rate, and extreme articulation when compared with adult-directed speech (ADS) [3][4][5]. The modifications in IDS has been claimed to facilitate infants' language development by eliciting attention and positive affect [6][7].

While extensive work has been conducted on the exaggerated intonation in IDS when compared with ADS, such as the higher pitch level and expanded pitch range at the level of the phrase or sentence [8], limited work has concerned with whether the acoustic characteristics at the lexical level are preserved or modified by the exaggerated intonation in IDS. The recent work on lexical tone in Mandarin IDS [9] found that tones in Mandarin IDS were not distorted by the exaggerated intonation, but were exaggerated as in the case of phonetic exaggeration of vowels in IDS [10].

In Tokyo Japanese, accent is principally manifested by F0 modulation, as a sharp F0 fall. The turning point (the start of the sharp F0 fall) associated with the pitch accent, however, is occasionally placed after the end of the associated mora [11] [12]

without listener detecting any change in accent placement [12]. This phenomenon is called "F0 fall delay (*ososagari*)" and tends to occur more frequently in female speech [13], and the paralinguistic function of the delay has been claimed to convey a sense of femininity in Japanese [14]. Figure 1 shows the F0 contours for the same word *ma'ma* 'mother' with and without the F0 fall delay (*ososagari*).

Previous work on Japanese IDS [15] showed that the F0 fall delay (*ososagari*) was observed in IDS as well as in ADS with the amount of delay being larger in IDS than in ADS. Although this previous study revealed the general tendency of turning points associated with pitch accents to be delayed in IDS, they did not examine factors which would affect the timing of F0 turning points such as the presence of the prosodic boundary [16] and the phonological structure of the accented syllable [17]. The F0 fall delay may not occur equally, and factors that would affect the alignment of tonal targets should be considered in order to fully characterize the acoustic modifications of pitch accents in Japanese IDS.

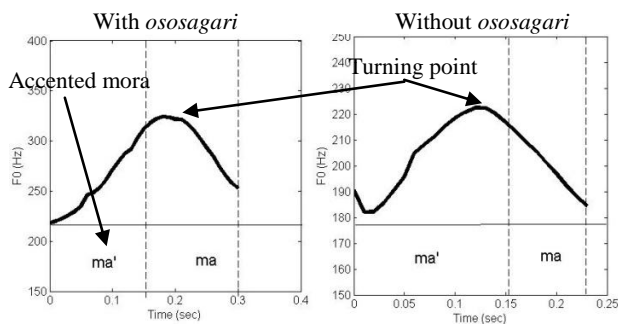


Figure 1: F0 contours of the word *ma'ma* 'mother' with and without *ososagari*.

The present study extended the previous investigation by examining how often F0 fall delay (*ososagari*) of pitch accents would occur in IDS and whether the frequency of F0 fall delay would depend on the phonological structures of the accented syllable as well as prosodic structures, including the presence of boundary pitch movements and non-lexical lengthening. We used the recorded IDS with detailed intonational annotations. Determining the characteristics of pitch accents in spontaneously produced IDS with the reference to phonological structure would be an important step towards determining the role of IDS in the acquisition of pitch accents by infants.

2. Method

2.1. Corpus description

In this study, we analyzed the RIKEN Japanese Mother-Infant Conversation Corpus (R-JMICC) [1], a speech database which contains 22 mothers' conversations (ranging in age from 25 to 43 and native Japanese speakers from the Tokyo area), with their 17-24 month old children (13 boys and 9 girls) (IDS), and with a female experimenter (ADS). The utterances produced by mothers were recorded to a DAT recorder (sampling rate=44,100 Hz) using a head-mounted microphone in a sound-attenuated booth in the Laboratory for Language Development at RIKEN.

The database consists of speech signals and various linguistic annotations such as transcriptional texts and segmental, morphological, and intonational labels. The intonational labeling is based on the X-JToBI scheme [18][19][20], which is based on a phonological model of Japanese intonation [21]. In ADS, the mothers talked with a female experimenter for about 10 minutes. In IDS, mothers talked to their infants with picture books for 15 minutes and with toys for 15 minutes. In this paper, only the book part was analyzed. The data from one mother was excluded from the analysis because it was difficult to reliably extract F0 contour due to her overly creaky phonation.

2.2. Data analysis

2.2.1. Selection of tokens in the database

For the current analysis, bisyllabic tokens with the accent on the first syllable ($\sigma'\sigma$) were selected from the database. $/'$ marks the accent position as the left side of this symbol. The accented syllables immediately followed by pauses were not included for the analysis. Only voiced tokens were used. The syllable structures of the accented syllables were either (C)V' (a one-mora syllable) or (C)V'N (a two-mora syllable ending with a moraic nasal). In this study, (C)V'H (a two-mora syllable ending with a long vowel) and (C)V'V (a two-mora syllable ending with a diphthong) were not employed due to the difficulty associated with the segmentation of the end of the accented mora. Filled pauses (like *hum*, *huh*), word-fragments, and utterances overlapped with laughter, coughing, and breathing were excluded from the measurements in order to prevent the use of incorrect F0. The total numbers of tokens that satisfied these criteria were 1061 tokens in ADS and 1459 tokens in IDS. Examples and number of tokens in the database depending on the syllable structure are listed in the table 1, including the types of tokens used in the analysis as well as those not used in the current analysis.

Table 1. Examples and number of tokens in the database.

Syllable structure of σ'	Example ($\sigma'\sigma$)	N (ADS)	N (IDS)	Analysis
(C)V'	ma' ma	893	1234	Included
(C)V'N	ze'N bu	168	225	
(C)V'H	ne'e ne	139	136	Not included
(C)V'V	ma'i ni	86	110	

2.2.2. Determination of the turning points

For the selected tokens, F0 contours were extracted from the audio waveform by using the autocorrelation algorithm on Praat [22] with the pitch floor being 75 Hz and the pitch ceiling being 1000 Hz. F0 tracking errors were corrected manually by editing Pitch Objects. After extracting the F0 contours, the location of the turning points associated with the pitch accents was determined by using the two-piece linear regression, fitted through the region from the onset of the accented syllable to the end of the following syllable as illustrated in Figure 2.

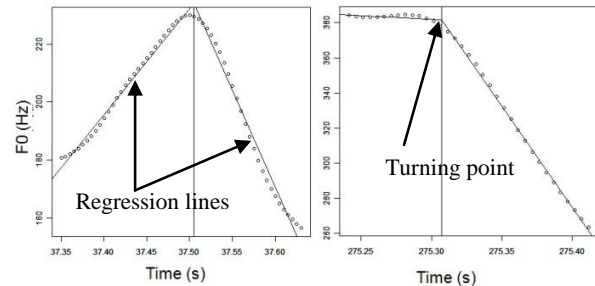


Figure 2: Examples of two-piece linear regression lines and the intersections (turning points) superimposed on the F0 contours from the beginning of the accented syllables to the end of the following syllable (*[ya'na]* in the left panel and *[ru'no]* in the right panel).

2.2.3. Examination of the labeled turning points

The labeling of the turning point was checked by the first and forth authors to remove the tokens with irregular F0 curves due to segmental perturbations and those with unclear boundaries of phonetic segments. The least-square fitting resulted in reasonable labeling of turning points for 57.8% of selected tokens in ADS and for 57.8% in IDS. There are tokens where we need to manually adjust the region to realize a better fitting of the F0 curve (17.1% in ADS, 17.2% in IDS). For some of the tokens, the F0 track (computed by Praat) was too irregular (due to strong segmental effects, heavy glottalization, or vocal fry) to permit the alignment of the turning points to be labeled (25.1% in ADS, 25.0% in IDS). The final number of utterances which could be labeled was 795 (ADS) and 1094 (IDS) making of 74.9% and 75.0% of the total number of selected tokens.

2.2.4. Calculation of the frequency of ososagari tokens

The tokens used for the analysis were those whose turning points were kept as originally labeled and manually adjusted when they were examined during the previous step (2.2.3.). Based on the alignment of the turning point with respect to the end of accented mora, we calculated the frequency of *ososagari* tokens in each mother's data, and took the average across mothers. The data were further analyzed depending on 1) the syllable structure of the accented syllable, 2) the presence of Boundary Pitch Movements on the syllable following accented syllable, and 3) the presence of non-lexical lengthening of accented vowel. The percentages of *ososagari* tokens were arc-since transformed, and paired-sample *t* tests were conducted in order to examine the significance of differences in frequency of F0 fall delay between ADS and IDS.

3. Results and Discussion

3.1.1. Overall results

The overall results showed that *ososagari* tokens were observed in both ADS and IDS with a slightly higher average frequency in IDS, as shown in the bar graph for ‘all data’ condition in Figure 6. Error bars indicate standard errors. Paired-sample *t* tests showed that the overall mean rates of *ososagari* tokens did not differ significantly between ADS and IDS, $t(20) = 0.418$, $p = 0.681$.

3.1.2. Effects of syllable structures

Figure 3 shows the occurrence rate of accented syllable depending on the syllable structure of accented syllable in ADS and IDS. ADS and IDS showed similar distribution of syllable structures of the accented syllables. CV’ tokens occurred significantly more frequently than CV’N tokens in both ADS, $t(20) = 13.179$, $p < 0.0001$ and IDS, $t(20) = 15.657$, $p < 0.0001$. There was no significant difference in the rate of CV’ syllable structure between ADS and IDS, $t(20) = 1.155$, $p = 0.262$. The *ososagari* tokens occurred significantly more frequently for the CV’ syllable structure compared to CV’N in ADS, $t(20) = 2.495$, $p = 0.021$. In IDS, the frequency of *ososagari* was also generally higher for CV’ compared to CV’N condition, but the difference between the different syllable structures was not statistically significant ($t(20) = 1.797$, $p = 0.087$).

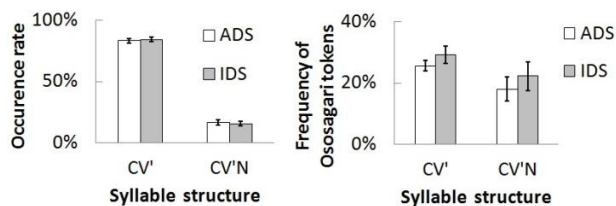


Figure 3: Occurrence rate (in the left panel) and frequency of *ososagari* tokens (in the right panel) depending on the syllable structure.

3.1.3. Effects of Boundary Pitch Movements (BPMs)

Boundary Pitch Movements (BPMs) are characteristic pitch changes that mark the end of a prosodic phrase and contribute to the pragmatic interpretation of the utterance, such as questioning, emphasis, and conjunction [20]. The main types of BPMs includes H% (rise), LH% (sustained low followed by a rise), and HL% (rise-fall).

The possible effects of the presence of BPM on the syllable following accented syllable are illustrated in Figure 4. The turning points tended not to cross over the end of an accented mora /na/ when they were immediately followed by BPM (*with BPM* condition in Figure 4), probably due to a constraint mandating the occurrence of low F0 region at the syllable bearing BPM before the F0 rise. On the other hand, when the accented mora was not immediately followed by BPM (*without BPM* condition in Figure 4), the turning point might cross over it. Therefore, we expected that the distribution of *ososagari* tokens would change depending on whether the accented syllable was followed by the syllable with BPM or not.

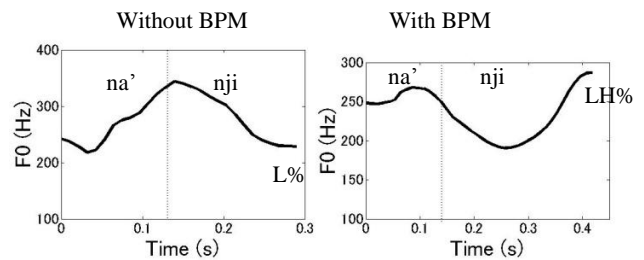


Figure 4: F0 contours for the token [na'ni] ‘what’ without BPM (in the left panel) and with BPM (LH%) on the following syllable (in the right panel).

Figure 5 shows the occurrence rate of tokens with BPM on the syllable following the accented syllable. The frequency of tokens with BPM was significantly higher in IDS than in ADS, $t(20) = 7.1$, $p < 0.0001$. Tokens with BPM were observed in 15 speakers in ADS and all of the speakers in IDS. Using the data from these speakers, further analysis was conducted to examine the effect of the BPM on the frequency of *ososagari* tokens. There was a significantly higher frequency of *ososagari* tokens when the accented syllable was not followed by the syllable with BPM compared to the case when the accented syllable was followed by BPM in ADS, $t(14) = 11.879$, $p < 0.0001$ as well as in IDS, $t(20) = 17.151$, $p < 0.0001$.

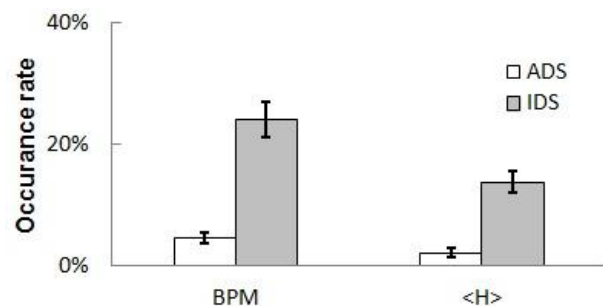


Figure 5: Occurrence rate of BPM and <H> tokens.

3.1.4. Effects of non-lexical lengthening (<H>)

In order to further evaluate the *ososagari* phenomenon in relation to temporal characteristics, we examined the effects of non-lexical lengthening of the accented vowel, marked as <H> in the corpus. Non-lexical lengthening (<H>) means that the duration of the vowel is lengthened without making any change in the meaning of the word. In Japanese, duration can be lengthened either phonemically or not phonemically. For example, the vowel /a/ in the second syllable in the word *obachan* ‘aunt’ can be lengthened to *obaHchan* meaning ‘grandmother’. On the other hand, lengthening the vowel /a/ in the first syllable in the word *na'nji* ‘what’ does not result in any change of the meaning of the word (*na'<H>nji* ‘what’), and corresponds to the case of non-lexical lengthening (<H>), examined in this study.

As Figure 5 shows, the frequency of tokens with non-lexical lengthening (<H>) was significantly higher in IDS than in ADS, $t(20) = 7.588$, $p < 0.0001$. Tokens with non-lexical lengthening (<H>) were observed in 9 speakers in ADS and 20 speakers in IDS. Using the data from these speakers, further analysis was conducted to examine the effect of the non-lexical lengthening (<H>) on the frequency of *ososagari* tokens. There was a

significantly higher frequency of *ososagari* tokens in non <H> tokens compared to <H> tokens in ADS, $t(8) = 4.752, p = 0.001$ as well as in IDS, $t(19) = 7.366, p < 0.0001$.

3.1.5. Effects of excluding BPM and <H> tokens

The effects of excluding BPM and <H> tokens were examined 1) by removing the tokens with <H>, 2) by removing tokens with BPM, and 3) removing the tokens with BPM and tokens with <H> at the same time. As Figure 6 shows, the difference in the average frequency of *ososagari* tokens between ADS and IDS became significant 1) by excluding <H> tokens, $t(20) = 2.622, p = 0.016$, 2) by excluding BPM tokens, $t(20) = 3.815, p = 0.001$, and 3) by excluding BPM tokens and <H> tokens at the same time, $t(20) = 4.119, p = 0.0005$.

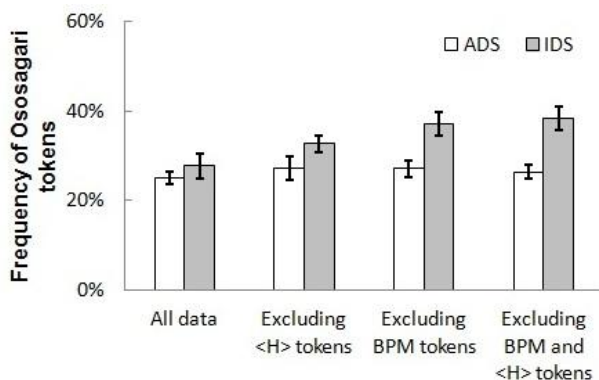


Figure 6: Frequency of *ososagari* tokens.

4. Conclusion

The present study examined F0 fall delay of Japanese lexical pitch accents in IDS, with reference to the language-specific prosodic structures. In Japanese IDS and ADS, the phenomenon of delayed F0 fall was observed, and this phenomenon turned out to be more pronounced in IDS compared to ADS when the accented mora was not followed by the BPMs, and/or when the accented mora was not non-lexically lengthened. These results suggest that Japanese IDS may exaggerate features of lexical pitch accents that potentially make them sound more feminine for infants, but only when it was less constrained by the prosodic contexts. Considering that the mora bearing BPM was the place where a prosodic exaggeration in terms of a pitch range expansion occurred in Japanese IDS [8], the current results may be viewed as a kind of trade-off between different ways of prosodic exaggeration (the F0 fall delay of pitch accents vs. pitch range expansion for the BPM and/or non-lexical lengthening) in Japanese IDS, and may point to the importance of exploiting a database with linguistic annotations to investigate the characteristics of lexical pitch accents in IDS.

5. Acknowledgements

We would like to thank Akira Utsugi, Yosuke Igarashi, Koki Miyazawa, and Andrew Martin for their invaluable discussions and suggestions. This work was supported by KAKENHI (No. 23720223).

6. References

- [1] Mazuka R., Igarashi Y. and Nishikawa K., "Input for learning Japanese; RIKEN Japanese Mother-Infant Conversation Corpus," Technical report of IEICE, TL2006-16, 106 (165): 11-15, 2006.
- [2] Snow C. E. and Ferguson C. A., "Talking to Children: Language Input and Acquisition," Cambridge Univ. Press: London, 1977.
- [3] Fernald A. and Simon T., "Expanded intonation contours in mothers' speech to newborns", *Developmental Psychology*, 20, 104-113 (1984).
- [4] Fernald A. et al., "A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants", *Journal of Child Language*, 16, 477-501, 1989.
- [5] Grieser D. L. and Kuhl P. K., "Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese", *Developmental Psychology*, 24, 14-20, 1988.
- [6] Cooper R. P. and Aslin R. N., "Preference for infant-directed speech in the first month after birth," *Child Development*, 61, 1584-1595, 1990.
- [7] Fernald A. and Kuhl P. K., "Acoustic determinants of infant preference for motherese speech", *Infant Behavior and Development*, 10, 3, 279-293, 1987.
- [8] Igarashi Y. and Mazuka, R., "Exaggerated prosody in infant-directed speech? Intonational phonological analysis of Japanese-infant-directed speech," BUCLD 32: Proceedings of the 32nd annual Boston University Conference on Language Development: Cascadia Press, 177-188, 2008.
- [9] Liu H.-M., Tsao F.-M. and Kuhl P. K., "Acoustic analysis of lexical tone in mandarin infant-directed speech", *Developmental Psychology*, 43, 4, 912-917, 2007.
- [10] Burnham D., Kitamura C. and Vollmer-Conna, U., "What's new, pussycat? On talking to babies and animals", *Science*, 296, 1435, 2002.
- [11] Neustupný J. V., "Is the Japanese accent a pitch accent?" *Onsei Gakkai Kaihoo*, 121, 230-39, 1966.
- [12] Sugito M., "Nihongo akusento no kenkyuu," *Sanseido*, 49-75, 1982.
- [13] Hata K. and Hasegawa Y., "A study of F0 reset in naturally-read utterances in Japanese", *Proceedings of the International Conference of Spoken Language Processing*, 1239-42, 1992.
- [14] Hasegawa Y. and Hata K., "The function of F0-peak delay in Japanese", *BLS*, 21, 141-151, 1995.
- [15] Kitahara M. et al., "Characteristics of pitch accents in infant-directed speech," Technical report of IEICE, SP2008-101, (108-338), 133-136, 2008.
- [16] Shinya T. and Takasawa M., "Nihongo no hatuwa matu ni okeru ehuzero kakoo no arainmento ni tuite", in *Proceedings of the 14th general meeting of the Phonetic Society of Japan*, 149-154, 1999.
- [17] Ishihara T., "Tonal Alignment in Tokyo Japanese," Ph. D. thesis, University of Edinburgh, 2006.
- [18] Maekawa K., Kikuchi H., Igarashi Y. and Venditti J., "X-JToBI: An extended J_ToBI for spontaneous speech", *Proceedings of The 7th International Conference on Spoken Language Processing*, 1545-1548, 2002.
- [19] Igarashi Y., Kikuchi H. and Maekawa K., "Prosodic information" (in Japanese), NIJL Technical Report 124, National Institute for Japanese Language: Tokyo, 347-453, 2006.
- [20] Venditti J., "The J_ToBI model of Japanese intonation," in Jun S.-A. [Ed], *Prosodic Typology: The Phonology of Intonation and Phrasing*, 172-200, Oxford University Press, 2005.
- [21] Pierrehumbert J. and Beckman M., "Japanese Tone Structure," The MIT Press, 1988.
- [22] Boersma P. and Weenink D., "Praat: doing phonetics by computer (version 5.2.18)", Retrieved from <http://www.praat.org/>.