Shift of voice onset time and enhancement in Japanese infant-directed speech

Hyun Kyung Hwang[1], Reiko Mazuka[1,2]

[1]RIKEN Center for Brain Science, [2]Duke University
hyunkyung.hwang@riken.jp, reiko.mazuka@riken.jp

ABSTRACT

Infant-directed speech (IDS) and adult-directed speech (ADS) are compared to test whether laryngeal contrast is enhanced when mothers talk to their infants against the backdrop of sound change with respect to voice onset time in Japanese stops. The results of a production study demonstrate that voice onset time of initial stops is still a crucial acoustic parameter for laryngeal contrast in Japanese despite large overlap between the categories. Further, it is revealed that mothers spread the innovative pattern rather than enhance laryngeal contrast. Mean F0s exhibit greater differentiation for voicing distinction in IDS, offering implications for infants' acquisition of laryngeal contrast in Japanese.

Keywords: Laryngeal contrast, voice onset time, infant-directed speech, enhancement, Japanese.

1. INTRODUCTION

Human infants must learn the acoustic cues that differentiate contrasting sounds in a language. One of the challenges they encounter can be varying cues in the input across generations if there is an ongoing sound change. Recently, an apparent change-in-progress was reported with respect to voice onset time (VOT) of initial voiced stops in Japanese, that is, lack of pre-voicing among younger speakers ([6, 15]). This leads us to question whether the shift of VOT was reflected in infant-directed speech (IDS) as well, which is the primary speech input for infants' language acquisition.

IDS is often characterized by "hyper-articulation" because this particular speech style exhibits higher pitch [5], larger pitch range [4, 5], or expanded vowel space [2]. More important, it has often been believed that IDS is clearer and enhances certain features to improve the learnability of contrastive sounds when they talk to their infants. Given the general characteristics of IDS, Japanese provides an interesting testing ground for the possibility that mothers still exhibit pre-voicing when they talk to their infants to facilitate learning of the laryngeal contrast by making acoustic differences more salient. With respect to this particular acoustic dimension, however, conflicting results have been reported even in the same language. For instance, in comparing VOT in IDS and adult-directed speech (ADS) in English, Burnham et al. [3] and McMurray et al. [10] find that voicing contrast is enhanced, whereas Baran et al. [1] and Synnestvedt [14] fail to find any enhanced contrast in IDS.

In Japanese, acoustic characteristics of laryngeal contrast in IDS have rarely been discussed. Recently, Hwang et al. [7] examined VOTs of utterance-initial stops using an IDS corpus of spontaneous speech. They demonstrated that VOTs are less differentiated in IDS by producing significantly shorter VOTs for voiceless stops, implying that IDS does *not* necessarily facilitate the perceptual development of infants. However, spontaneous speech is not controlled by nature; the frequencies of stops between voiced and voiceless categories, or the occurrence of different places of articulation, are imbalanced [16]. Further, other potentially important cues are not investigated, such as the F0 or voice quality of following vowels. Thus, this study explores initial stops of words in citation, which is a typical situation when a mother teaches her infant novel words. This study also aims to provide more comprehensive data by considering various acoustic dimensions involving vocalic cues.

Specifically, the following questions are addressed regarding laryngeal contrast in Japanese; 1) Have stop VOTs shifted in maternal input as well? Or do mothers still yield pre-voiced stops prevalently in IDS to facilitate learning of voicing contrast in citation speech? 2) If IDS indeed involves facilitative function, what acoustic dimensions are enhanced in Japanese? 3) Given the VOT shift, what are the acoustic correlates of voicing contrast for initial stops?

2. METHODS

2.1. Materials

Three minimal pairs, which began with bilabial stops, were recorded. [1] In constructing the test materials, accentedness, the location of the accent, and word length were controlled; the target consonants were initial stops of accented morae in bi- or tri-moraic words. Tested minimal pairs are given in Table 1 below.

Table 1: Target minimal pairs.

Voiced		Voiceless	
baɾi	'Bali'	paɾi	'Paris'
bentɕi	'bench'	pentɕi	'pliers'
bo:ɾɯ	'ball'	po:ɾɯ	'poll'

2.2. Participants and recording

Participants were 21 mother-infant dyads. Mothers ranged in age from 29 to 42 (mean: 35.5 years), and their infants were either 5 months (13 infants, 9 girls) or 9 months (8 infants, 4 girls) old. All speakers were born and grew up in Tokyo or its surrounding areas and had no history of speech impairment.

All recordings were made in a sound-attenuated room. A mother wearing a head-mounted microphone interacted with her infant (IDS) or with an experimenter (ADS) using a set of pictures of the target words. Participants were instructed to give natural renditions at a comfortable speed. Each target word was repeated five times.

2.3. Measurements

VOT and segment boundaries were manually marked on each token. Two duration measures, VOT and duration of vocalic portions; one pitch measure, mean F0; and two phonation measures, H1-H2 (relative amplitude of first two harmonics) and H1-A1 (relative amplitude of first harmonic to first formant) were made by implementing a Praat script [17]. The pitch and phonation measures were calculated at the onset 25 ms of following vowels. A total of 1260 tokens were obtained (6 words * 2 registers * 5 repetitions * 21 mothers), and 1203 tokens were further analyzed. Fifty-seven tokens were excluded before the analysis due to invisible stop burst, noise, or disfluency.

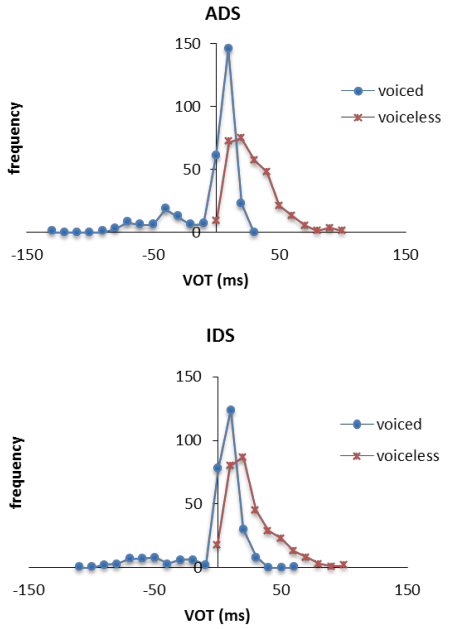
3. RESULTS

Acoustic characteristics of different speech registers were compared by conducting linear mixed-effects analyses. Register (ADS, IDS), voicing (voiced, voiceless), and infant age (5 months, 9 months) were entered into the model as fixed effects, and the speaker and item were entered as random effects. Because infant age was not significant in all the parameters, it will not be further discussed in the results.

3.1. VOT

The majority of voiced stops were realized with short-lag VOTs. Approximately 23% and 16% of the voiced stops exhibited pre-voicing in ADS and IDS, respectively. While only one speaker produced pre-voicing more than half of the time regardless of speech register, three more speakers yielded pre-voiced stops more than half of the time only in ADS. Voiceless stops, on the other hand, showed intermediate values between short- and long lag. As shown in Figure 1, a large overlap between the categories was observed in both ADS and IDS. Overall, the VOT distribution of initial stops was extremely similar regardless of speech register.

Figure 1: VOT distribution of voiced (circles) and voiceless (crosses) in ADS (top) and IDS (bottom).



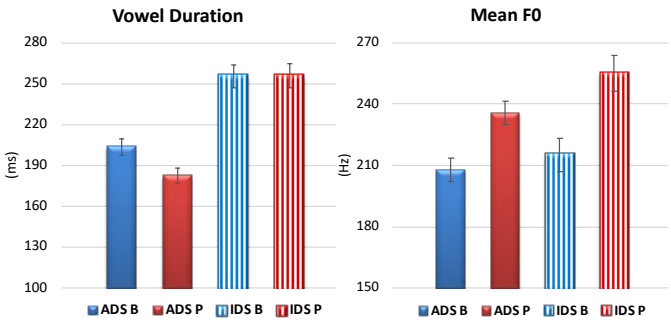
Statistical analyses revealed that voicing was significant (β = −14.88, t = −25.52, p < .0001*), whereas register did not reach significance (β = −0.56, t = −0.96, p=.33). Interestingly, the interaction between voicing * register was marginally significant (β = −1.16, t = −2.00, p = .0462*). While no significant difference was found for voiceless stops (32.62 ms in ADS vs. 31.42 ms in IDS), VOTs were slightly longer in IDS for voiced stops (0.54 ms in ADS vs. 3.98 ms in IDS), contrary to the hypothesis of enhancement.

3.2. Parameters in vocalic portion

Figure 2 demonstrates average vowel duration (ms) and mean F0 (Hz) of the two stop categories in both speech registers. Vowels after voiced stops were fairly longer than those after voiceless stops in ADS (204 ms after B vs. 182 ms in after P). On the other hand, vowels in IDS were considerably longer than those in ADS regardless of the voicing categories (193 ms in ADS vs. 256 ms in IDS). Statistically,

register (β = −32.48, t = −17.60, p < .0001*) and voicing (β = 6.27, t = 3.40, p = .0007*) as well as their interaction (β = 4.27, t = 2.31, p = .0209*) were significant. While IDS vowels were significantly longer, the difference between voicing categories did not reach significance, showing that voicing contrast was less differentiated in IDS in this temporal dimension.

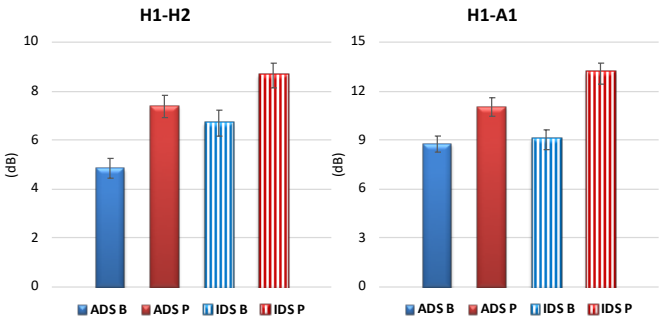
Figure 2: Average vowel duration (left) and mean F0 (right) in ADS (solid bars) and in IDS (striped bars). Error bars indicate standard errors.



As for mean F0, voiceless stops yielded higher mean F0 than their voiced counterparts across speech registers (211 Hz after B vs. 245 Hz after P). The effect of voicing on mean F0 was significant (β = −17.06, t = −20.91, p < .0001*). Moreover, IDS was produced with higher mean F0 than ADS (222 Hz in ADS vs. 236 Hz in IDS), revealing a robust effect of register on F0 (β = −6.51, t = −7.99, p < .0001*). Crucially, the interaction between voicing and register was significant (β = 3.31, t = 4.06, p < .0001*). The difference between the voicing categories was significantly greater in IDS (28 Hz in ADS vs. 40 Hz in IDS). This result showed that voicing contrast was indeed more differentiated in IDS, supporting the enhancement hypothesis.

Regarding the phonation measures, average H1-H2 (dB) and H1-A1 (dB) are demonstrated in Figure 3. As for H1-H2, there was a significant main effect of voicing (β = −1.09, t = −4.94, p < .0001*) and register (β = −0.77, t = −3.47, p = .0005*).

Figure 3: Average H1-H2 (left) and H1-A1 (right) in ADS (solid bars) and IDS (striped bars).



H1-H2 was greater after voiceless stops than after voiced ones (5.7 dB in B vs. 8.0 in P) and in IDS than in ADS (6.1 dB in ADS vs. 7.7 dB in IDS), indicating that voiceless stops and IDS were breathier than their counterparts. However, their interaction was not significant (β = -0.16, t = -0.72, p=.4691).

Concerning H1-A1, there was a significant effect of register (β = −0.60, t = −2.16, p < = .0306*) and voicing (β = −1.57, t = −5.66, p < .0001*). Voiceless stops yielded greater values than voiced ones (8.9 dB after B vs. 12.1 dB after P), and IDS than ADS (9.9 dB in ADS vs. 11.1 dB in IDS). However, the interaction of voicing and register was not significant (β = 0.44, t = 1.60, p = .1088). Both phonation parameters exhibited similar results not only with respect to overall patterning but also with respect to the lack of enhancement in IDS.

4. DISCUSSION

4.1. Shift of VOT

The results revealed that even in citation speech, mothers do not enhance contrast by producing pre-voiced stops prevalently when they talk to their infants. This indicates that the VOT target of voiced stops in Japanese has already shifted in maternal input as well. This finding is consistent with Hwang et al.'s [7] report that the laryngeal distinction in Japanese is not enhanced but less differentiated in this acoustic dimension in spontaneous speech. Thus, it appears evident that the VOT target of voiced stops has changed to a positive value across different speech styles and registers.

4.2. Enhancement of contrast in IDS

Do Japanese mothers not facilitate learning of voicing contrast in citation speech at all? The results revealed that the laryngeal contrast in Japanese is indeed enhanced in one dimension: mean F0. It may be the case that the tonal difference is salient and readily manipulatable by mothers.

On the other hand, no difference or less differentiation between the voicing categories in IDS was observed both in temporal dimensions, that is, VOT and vowel duration, and phonation measures, that is, H1-H2 & H1-A1. While vowel duration was remarkably longer in IDS, the differences between the voicing categories were lost because both categories were associated with longer vowels. Similarly, both phonation measures did not yield enhanced distinction in IDS, whereas both revealed IDS was breathier than ADS, corroborating a previous finding in spontaneous speech of Japanese ([12]). Perhaps these parameters are modified in IDS

mainly for other functions such as expressing affect or regulating infants' attention at the cost of enhancing laryngeal contrast. It has been widely acknowledged that IDS is used not only for facilitating infants' language development but also for encouraging, rewarding, or regulating the arousal level of infants ([5, 9]), or expressing positive emotion ([13]).

4.3. Cues to laryngeal contrast in Japanese

Given the large VOT overlap between the two categories, a question arises as to the acoustic correlates of the laryngeal contrast in Japanese. The data showed that utterance-initial voicing has a large effect on F0 and voice quality in the following vowel. In considering the results of the production experiment, it is conceivable that the VOT change is nearly completed, and additional cues such as F0 and phonation cues play an important role. This finding confirms the observation reported in previous studies that voiceless stops yield higher F0 ([6, 10]) and greater H1-H2 values ([10]) in Japanese. It is worth noting that cross-linguistically, the emergence of phonation or particularly tonal cues to voicing contrast is not uncommon [8].

With respect to language acquisition, similar VOT targets between the categories and presumably varying cues across generations exacerbated by the lack of enhancement in most of the acoustic dimensions require that infants develop a finer attunement to learn voicing distinction. This implies that acquisition of voicing contrast may be delayed in Japanese, which is supported by a behavioral experiment in previous research [7].

5. CONCLUSIONS

This paper discussed the Japanese laryngeal contrast in different speech registers with special attention to the shift of VOT observed in voiced stops. The careful comparison between ADS and IDS attempted to better understand the nature of maternal input and to explore acoustic correlates of voicing contrast in Japanese.

The results of the present study demonstrate that the VOT shift is nearly completed at least among female speakers of this age group, and mothers spread this innovative pattern even when they say words in isolation to their infants. It is also revealed that the only dimension that is enhanced in IDS is mean F0 among the five acoustic parameters. The lack of enhancement in certain dimensions can be attributed to nonfacilitative function of IDS, including communication of affect or regulation of arousal level of infants. Further, the findings provide empirical evidence to support the emergence of F0 and phonation cues together with VOT to voicing distinction regardless of speech register.

6. ACKNOWLEDGEMENTS

The authors thank Yuri Hatano and Mihoko Hasegawa for their help with data collection and speaker recruitment. Special thanks go to Kengo Takeda for his invaluable help in data analysis. This work was supported in part by JSPS KEKENHI Grant Number 17K13458 to the first author and 16H06319 & #4903-17H06382/17H06383 to the second author.

7. REFERENCES

[1] Baran, J. A., Laufer, M. Z. and Daniloff, R. 1977. Phonological contrastivity in conversation: A comparative study of voice onset time. *J. of Phonetics* 5, 339–50.

[2] Burnham, D., Kitamura, C. and Vollmer-Conna, U. 2002. What's new, pussycat? On talking to babies and animals. *Science* 296, 1435.

[3] Burnham, E., Gamache J., Dilley, L., and Bergeson, T. 2013. Voice-onset time in infant-directed speech over the first year and a half. *Proceedings of Meetings on Acoustics* 19, 1–7.

[4] Ferguson, C. A. 1964. Baby talk in six languages. *American Anthropologist* 66, 103–114.

[5] Fernald, A. and Simon, T. 1984. Expanded intonantion contours in mother's speech to newborns. *Developmental Psychology* 20, 104–113.

[6] Gao, J. and Arai, T. 2018. F0 perturbation in a "pitch-accent" language. *Proceedings of the 6th Tonal Aspects of Languages*.

[7] Hwang, H. Mazuka, R. Takada, M. 2018. Enhancement of stop contrast or emergence of new targets?: Implications on language development in Japanese. Poster presentation at BUCLD 43.

[8] Kirby, J. 2018. Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages. *J. of Phonetics* 71, 326–354.

[9] Kitamura, C., Burnham, D. 1998. The infant's response to maternal vocal affect. In: Rovee-Collier, C., Lipsitt, L. Hayne, H. (eds.), *Advances in infancy research* 12. Stamford, CT: Ablex, 221–236.

[10] Kong, E., Beckman, M.E., Edwards, J. 2012. Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *J. of Phonetics* 40(6), 725–744.

[11] McMurray, B., Kovack-Lesh, K. A., Goodwin, D., McEchron, W. 2013. Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition* 129(2), 362–378.

[12] Miyazawa, K. Shinya, T., Martin, A., Kikuchi, H., Mazuka, R. 2017. Vowels in infant-directed speech: More breathy and more variable, but not clearer. *Cognition* 166, 84–93.

[13] Scherer, K. 1986. Vocal affect expression: A review and a model for future research. *Psychological Bulletin* 9, 143–165.

[14] Synnestvedt, A. 2010. *Voice onset time in infant-directed speech at two ages.* Doctoral dissertation, University of Maryland.

[15] Takada, M. 2011. *Nihongo no goto heisaon no kenkyu: VOT no kyojiteki bunpu totsujiteki henka* [Research on the word-initial stops of Japanese: Synchronic distribution and diachronic change in VOT]. Tokyo: Kurosio.

[16] Tsuji, S., Nishikawa, K., Mazuka, R. 2014. Segmental distributions and consonant-vowel association patterns in Japanese infant- and adult-directed speech. *J. of Child Lang* 41, 1276–1304.

[17] Xu, Y. 2013. ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France. 7-10.

[1] Although bilabial stops may not be ideal to test in Japanese, this place of articulation was recorded for direct comparison with other languages, as this study is a part of a larger cross-linguistic project.