THE ROLE OF EXTRA-LINGUISTIC FACTORS IN PITCH RANGE VARIATION: A CORPUS STUDY OF SPOKEN JAPANESE

Sano Shinichiro[1], Sano Céleste[2]

Keio University[1], Bunkyo Gakuin University[2]
s-sano@keio.jp, sanoceleste@gmail.com

ABSTRACT

The phonological organization of a language restricts the way in which pitch change can be realized lexically and/or intonationally. Previous studies on tonal languages (e.g., Chinese, Thai) have shown that a phonological grammar that demands a particular tonal classification affects the variability in pitch range; namely, in tonal languages smaller variations are observed compared to non-tonal languages (e.g., English). Japanese is a language with lexical pitch accent, but no intonational pitch accent reported, although it utilizes phrasing, for example in focus marking.

This study investigates the range of pitch accent in Japanese spontaneous speech. Based on the analysis of two corpora of spoken Japanese (Corpus of Spontaneous Japanese and Showa Spoken Japanese Corpus), we examined the effects of gender, age and style on pitch range as well as diachronic variations. Findings suggest that language-specific restrictions on pitch range can be overridden by the influence of extra-linguistic factors.

Keywords: Japanese, corpus, prosody, pitch range, spontaneous speech

1. INTRODUCTION

Languages have phonological organizations that restricts the organization of intonational patterns. Namely, the way pitch patterns are realized both lexically and in the intonation depends on language-specific properties [1, 2, 3]. In specific situations, speakers of a language are allowed to choose from an inventory of intonational tones which are limited by the phonological organization of the language. Previous studies on Infant Directed Speech, for example, indicate that tonal languages (e.g., Chinese, Thai) exhibit less variations in intonation than languages that do not have tonal contrasts (e.g., English) [4]. The same observation was also made for Japanese compared to Western languages [5].

Japanese is a language with lexical pitch accent but no intonational accent [6]. It has however been reported to make use of phrasing to mark focus for example. Pitch accent is involved in a variety of minimal pairs, as exemplified in (1) which suggests its crucial role for lexical distinction in Japanese.

(1)	*kaki-ga*	LHH	'persimmon- NOM'
	kaki-ga	HLL	'oyster- NOM'
	kaki-ga	LHL	'fence- NOM'

Contrary to English where accent placement is determined by an interplay of various linguistic and extra-linguistic factors, pitch accent in Japanese is a property of a given word. A logical consequence of this should be a restriction of the allowed variations in pitch [7], as they might interfere with the meaning, resulting in a neutralization of contrastive items.

Previous studies point out some other evidence for the nature of Japanese with regard to intonation, such as the use of particles and the interaction of culturally specified rules. [5] notes that contrary to English where questions are marked with intonational cues, Japanese can make use of the interrogative particle –*ka*. Without recourse to intonational cues, Japanese listeners can identify if the sentence in question is affirmative or interrogative, which is signalled by -*ka*. Additionally, a body of psycho-linguistic research on perceptual cues for emotions in speech shed light on the influence of social rules on Japanese speech. Indeed, they found that compared to other languages the changes in their pitch were attenuated possibly due to social rules governing the display of social affects in Japanese (e.g., [8]). That is, not displaying emotions is a more socially accepted behavior in Japanese culture.

However, previous studies focusing on infant-directed speech in Japanese shed light on the fact that in specific pragmatic situations speakers can make use of pitch range modulation [4, 9]). Indeed, significant pitch expansion was observed for interactions with infants when compared to adult.

Previous studies have investigated variations in pitch mainly from an experimental and qualitative point of view. Furthermore, many studies deal with this topic but by focusing on specific pragmatic contexts (e.g., for infant-directed speech [4, 9]). Consequently, there is a need for more studies on the interaction between pitch and intonation in a spontaneous (adult) speech context. This paper explores the variations of pitch range in spontaneous speech using corpus data. Specifically, we examined the effect of gender, age and style on the variations in pitch range. Diachronic variations were also explored, based on the comparison of two corpora composed of speech recordings from different eras. The goals of this paper are as follows:

- Explore the patterns of pitch range in spontaneous speech.
- Examine the effect of extra-linguistic factors that may affect the patterns of pitch range.
- Investigate diachronic variations in pitch range using recordings form different eras.

2. METHODS

Data was retrieved from two different corpora of spoken Japanese. The first corpus is the *Corpus of Spontaneous Japanese* ([10], henceforth CSJ), a large-scale database of spontaneous speeches, from which a subset that comprises 12 speech samples (about 10,800 seconds, six male, six female, four speakers per speech style, age range: 25-69 yo.) was extracted. The CSJ has a variety of annotations that were used for analysis. Namely, pitch range measurements were conducted for each utterance following the annotations in the TextGrid files provided with the CSJ. We refer as utterance to what [10] calls transcription units, and which are defined as utterances separated by "longer-than-200 ms" pauses.

Gender and age are annotated in the corpus. The corpus also provides three different styles: (i) Academic Presentation Speech (A), a collection of recorded academic (conference) presentations; (ii) Simulated Public Speaking (S), recordings of solicited speeches about everyday life; (iii) Dialogue (D). A and S are monologues, while D is a conversation between two speakers. The three styles differ in terms of presence/absence of real audience and degree of formality. A and D involve a real interaction context between a speaker and an audience (one or several listeners), while in S there is no audience except for recording staffs. Regarding the formality, A is more formal than S and D.

In addition, for diachronic comparisons, we used data from the other speech corpus: the *Showa Spoken Japanese Corpus* ([12], henceforth SSJC), a database of speech samples recorded in the 1950's. This corpus does not come with speaker information but gender was identified from both listening and the personal information (e.g., name) given by the speakers in their speech. Although birth year is unknown, the contents of the recordings suggest that they were born from the end of the Meiji period (1900's) to the beginning of the Showa period (1920's). From the SSJC, we extracted and annotated the files whose recording quality was appropriate for acoustic analysis. As a result, we obtained about 2,000 seconds of speech samples from 19 speakers (six female, 12 male). Following the methodology for annotation in the CSJ, sound files were annotated automatically using a Praat [11] script detecting pauses of more than 200ms and boundaries were checked and added manually when necessary.

In total, we analyzed a total of 1,031 tokens for the SSJC and 3,785 for the CSJ. For each annotated clause, duration, minimum f0 and maximum f0 were retrieved using a Praat script, and pitch range was obtained based on this data.

In order to normalize the data that includes both male and female speakers, we converted the raw f0 measurements in semitones for pitch range analysis. Indeed, [4, 5, 14] show that proportional variations in pitch are more related to perception than absolute changes. The semitone scale is a logarithmic scale that divides changes in frequency on one octave in 12 equal intervals that allows to normalize the data, so that rather than equivalent differences in absolute frequency, it is equivalent proportional differences that are evaluated. Statistical skews were tested using R [15].

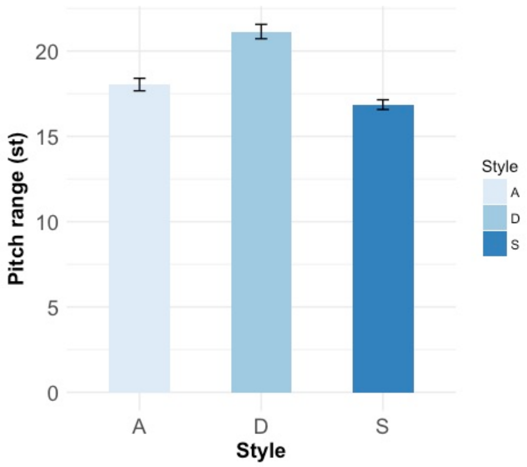
3. RESULTS

Analyses were conducted on the data in order to investigate the effect of gender, birth year and style on pitch range variations. The results are presented in terms of speech style, gender and their interaction in a first subsection. The following sections introduce respectively results for age and diachronic variations.

3.1. Speech Style and Gender

The effect of style on pitch range tested by two-way ANOVA was shown to be significant (p<.001). The results are presented in Figure 1. We observed the highest pitch range in D(ialogue) (m=21.14st), followed by A(cademic Presentation Speech) (m=18.03st), and the smallest pitch range was observed for the S(imulated Public Speaking) (m=16.85st). Pairwise comparisons by Tukey HSD test confirmed the significant difference between all three contexts.

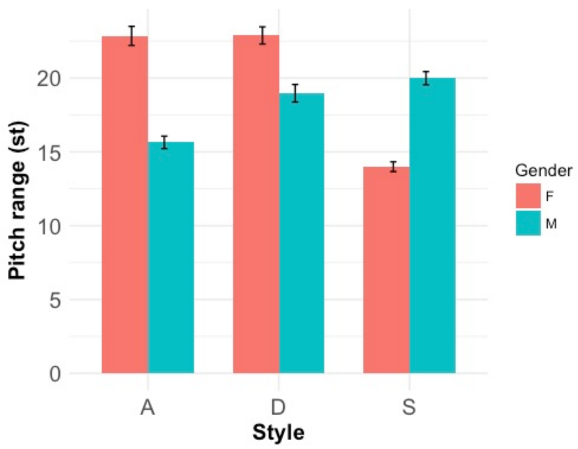
Figure 1: Pitch range by speech style (CSJ)



The results are consistent with the properties of each speech style and suggest the following. Higher pitch range is observed in D when compared to A and S, which corresponds to the affective dimension of this speech style where two participants are interacting. The absence of interaction with the audience, and the information-oriented nature of A justifies a lower pitch range. Lastly, for S, it is not the absence of interaction with the audience, but the absence of any real audience itself, that causes narrower pitch variations.

When looking at the influence of gender on pitch range, no significant variation (p=.1) was observed between female (m=18.80st) and male (m=18.06) speakers. However, a strong interaction between gender and style (p<.001) was observed, as illustrated in Figure 2.

Figure 2: Pitch range by speech style and gender (CSJ)



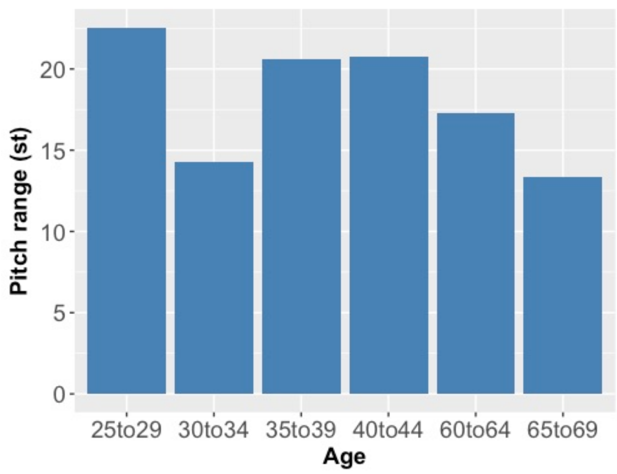
The results in Figure 2 suggest that pitch range in female and male speakers follow opposite patterns with regard to style. Female speakers exhibit a high pitch range in speech styles involving an audience (A, D) than in S without an audience. Statistical testing by multiple comparisons confirmed the absence of difference in pitch range between A and D in female speakers (p=1) while S was significantly different (p<.001). For male speakers, similar pitch ranges were observed for D and S (p=.7), which were higher than A (p<.001). This suggests that male speakers adapt their pitch variation patterns to the degree of formality of the context.

In sum, the interaction between gender and style shows a different behavior for men and women when it comes to the use of pitch variations in different stylistic contexts. Women are more sensitive to affective dimensions (presence or absence of audience, of interaction); on the other hand, men are more likely to be affected by referential ones (stylistic level, formal or casual context).

3.2. Age

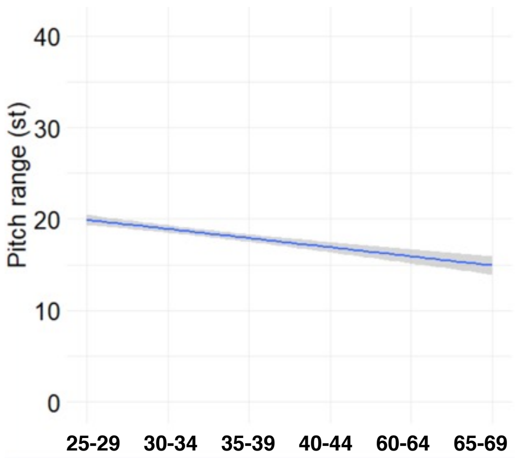
Next, we turn our attention to how the pitch range differs depending on the speakers' age. The influence of age on pitch range is illustrated in Figure 3 below. The number of speakers in each age span is as follows: 25 to 29 (4), 30 to 34 (2), 35 to 39 (3), 40 to 44 (1), 60 to 64 (1), 65 to 69 (1).

Figure 3: Pitch range by age group (CSJ)



Statistical testing by two-way ANOVA shows an inverse correlation between age and pitch range patterns (p<.001). Specifically, we can observe that the pitch range decreases as speaker age increases.
The maximum pitch range was observed for the youngest speakers (m=22.55st) and the lowest for the older speakers (m=13.33st). The decrease of pitch range with age is schematically presented in the fitted line with confidence interval in Figure 4.

Figure 4: Pitch range by speaker age (CSJ)

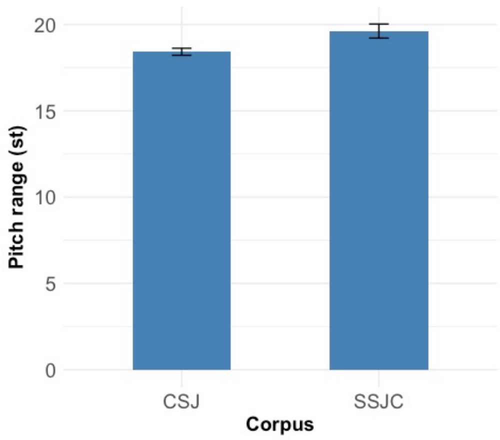


The generalization obtained from Figure 3 and 4 is that as people get older, pitch range decreases. This suggests the within-speaker development of the use of pitch variation (age-grading).

3.3. Diachronic variation

This section presents the results of the investigation on the chronological transition in pitch range using two corpora. The results are illustrated in Figure 5.

Figure 5: Pitch range by corpus (CSJ & SSJC)



Global pitch range in the CSJ was significantly (p<.01) lower (m=18.41) than in the SSJC (m=19.61). Given the fact that the recordings from the SSJC were made in the 1950's, results suggest a wider pitch range in earlier Japanese. Furthermore, we can also argue that the pitch range in Japanese has been chronologically decreasing.

6. CONCLUSION

Using spontaneous speech data from two different corpora, this study investigated the variation in pitch and how it is influenced by gender, birth year and speech style. In addition, we also compared the data with speech samples from the Showa era recorded in the 1950's.

Although the phonological properties of Japanese suggest that pitch variations should be minimal, our results shed light on the crucial role of extra-lingusitic factors: they can override phonological restrictions on tonal variation. Specifically, we observed a strong effect of the speech style. Japanese speakers modify their use of pitch according to the social context. Furthermore, the factors that affect pitch range adjustment differ depending on gender. The propensity of women to solidarity [16], [17], [18] makes them more sensitive to affective dimensions, such as the presence of an audience or the chances to interact or not with them. As such, they exhibit a wider pitch range when they are in an affective context. On the other hand, the fact that men are more affected by social factors such as formality of the situation is reflected in their wider pitch range in casual contexts. Age was also shown to affect pitch range. Results suggest that pitch range might be subject to variations across a speakers' life span. Lastly, we also observed a diachronic difference between data recorded in the 1950's and in the 2000's: a wider pitch range was observed for older recordings.

We close our discussion by mentioning one of the possible reasons for wider pitch range in SSJC than in CSJ. The SSJC recordings we focused on in this study were extracted from a seminar on speech and accent accuracy in the 1950's. More specifically, the speaker reported that wrong and/or dialectal accent patterns were frequently observed on the radio and in everyday life. If this story reflects the language use at that time, it provides interesting insights on the observed results. When put in perspective with previous claims that phonological grammar governs intonational patterns, it suggests that the restrictions induced by the phonological grammar might have been less active in Japanese in the past, allowing more variation in pitch range.

7. REFERENCES

[1] Ladd, D.R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

[2] Ward, G. Hirschberg, J. 1985. Implicating uncertainty: The pragmatics of fall-rise. *Language* 61, 747–776.

[3] Pierrehumbert, J., Hirschberg, J. 1990. The meaning of intonational contours in the interpretation of discourse. In: Cohen, P. Morgan, J. Pollack, M. (eds.) *Intentions in Communication*. Cambridge, MA: MIT Press, 271–311.

[4] Grieser, D.L., Khul, P.K. 1988. Maternal Speech to Infants in a Tonal Language: Support for Universal

Prosodic Features in Motherese. *Developmental Psychology*, 24(1), 14–20.

[5] Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., Fukui, I. 1989. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language* 16, 477–501.

[6] Pierrehumbert, J., Beckman, M. 1988. *Japanese Tone Structure*. Cambridge, Mass: MIT Press.

[7] Venditti, J.J. 2005. The J-Tobi Model of Japanese Intonation. In: Jun, S. (ed.) Prosodic Typology: *The Phonology of Intonation and Phrasing*. Oxford Scholarship Online.

[8] Magno-Caldognetto, E. and Kori, S. 1983. Intercultural judgement of emotions expressed through voice: the Italians and the Japanese. 10th International Congress of Phonetic Sciences, Utrecht, Netherlands.

[9] Igarashi, Y., Nishikawa, K., Tanaka, K., Mazuka, R. 2013. Phonological theory informs the analysis of intonational exaggeration in Japanese infant-directed speech. *Journal of the Acoustical Society of America* 134, 1283–1294.

[10] Maekawa, K. 2003. Corpus of Spontaneous Japanese: Its design and evaluation. *Proc. Of SSPR*.

[11] Takanashi, K., Maruyama, T., Uchimoto, K., Isahara, H. 2003. Identification of "sentences" in spontaneous Japanese –detection and modification of clause boundaries-. *Proc. Of ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition*, 183–186.

[12] Maruyama, T. 2016. Showa hanashi kotoba Corpus no kōchiku [Construction of the Showa Spoken Japanese Corpus]. Nichijō kaiwa Corpus Symposium. National Institute for Japanese Language and Linguistics.

[13] Boersma, P. Weenink, D. 1992-2918. *Praat: doing phonetics by computer*. www.praat.org.

[14] Ward, W.D. 1970. Musical perception. In: Tobias, J. (ed.) *Foundation of auditory theory*. New York: Academic Press, 405–447.

[15] R Core Team. 2017. *R: A language and environement for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

[16] Holmes, J. 1984. Women's language: A Functional Approach. *General Linguistics* 24(3), 149–178.

[17] Holmes, J. 1998. Women's talk, The Question of Social-linguistic Universals. In: Coates, J. and Pichler, P. (eds.) *Language and Gender: A Reader* (2nd ed.). Oxford: Wiley Blackwell, 461–483.

[18] Coates, J. and Pichler, P.1998. *Language and Gender: A Reader* (2nd ed.). Oxford: Wiley Blackwell.