SEGMENTAL AND AUTOSEGMENTAL MODIFICATION IN RAKUGO STYLE JAPANESE

Shigeto Kamano[1)], Chikau Sakamoto[2)], Rin Tsujita[3)], Seunghun J. Lee[4)]

[1,2,3,4)]International Christian University, [4)]IIT Guwahati
shig.kamano@gmail.com[1)], sakachi.jp@gmail.com[2)], c231421r@icu.ac.jp[3)], seunghun@icu.ac.jp[4)]

ABSTRACT

Rakugo is a Japanese performance art form where a solo performer delivers a story with comedic parts, often altering their voice when they produce a speech style unique to Rakugo. This paper compared the phonetic patterns of Rakugo-style speech and natural speech produced by 8 Rakugo performers (called *Hanashika*). An analysis of 32 bi- or tri-syllabic words varying in onset type, accent type, and vowel in word-initial syllables show differences in the Rakugo speech style. Results show that pitch range is larger and syllable duration is longer in the Rakugo style. Vowels were more dispersed but high vowel devoicing occurs with much less frequency in the Rakugo style. The expanded use of acoustic cues is suggested to be the phonetic features of Rakugo.

Keywords: Rakugo speech, Japanese, pitch range, syllable duration, vowel dispersion, vowel devoicing

1. INTRODUCTION

Rakugo is a storytelling form that originates from performances conducted since the 17th century. Rakugo stories are told by a *Hanashika* (the narrator), who personifies all the characters in a story. Both the personification part and the narration part result in the modification of voice by a *Hanashika*. Speech styles used in performance arts are often understudied. Some exceptions exist, however: [1], [2] on maid voice and [3] on anime voice.

Modification of speech in performance arts is well recognized by the audience, but identification of such phonetic aspects is less known. [1] reports that maid voice can be produced with raised F0, more dynamic F0 movement, or exaggerated vowels. Results in [2] show that anime voice is produced with stronger harmonics than normal speech.

In this paper, we focus on the speech production in non-Rakugo vs. Rakugo speech style and investigate the similarities and differences between the two, such as vowel quality, durational differences, vowel devoicing, and pitch range.

The rest of the paper is organized as follows. In section 2, methods for data collection are described. The results in section 3 demonstrate that Rakugo and non-Rakugo speech styles are not identical. We conclude the paper with a discussion section.

2. METHODS

2.1. Stimuli

32 randomized sentences containing target words in Table 1 were produced 3 times each by 8 Kanto Japanese speakers (in two different styles: Rakugo and non-Rakugo. The frame sentence was *kono ___-ttenaa, donna imidai* for the Rakugo style and *kono ___-to yuu tangono imiwa nandesuka* for the non-Rakugo style (both means "What is the meaning of this word _____ ?"). The final number of tokens is 1536 (32 target words x 8 participants x 3 repetitions x 2 styles).

target	accented	unaccented
na	*naka* 'middle' *nana* 'seven'	*naki* 'crying' *nami* 'moderate'
ni	*niki* '2 terms' *niji* '2 o'clock'	*nishi* 'west' *niji* 'rainbow'
nu	*nuka* 'nuka' *nugu* 'taking off'	*nuke* 'missing' *nuno* 'cloth'
ne	*negi* 'green onion' *neko* 'cat'	*neta* 'content' *nekki* 'heat'
no	*nodo* 'throat' *noshi* 'noshi'	*noki* 'eave' *nozomi* 'hope'
ta	*tane* 'seed' *take* 'length'	*tana* 'shelf' *take* 'bamboo'
te	*teko* 'lever' *teba* 'wings'	*teki* 'enemy' *tetsu* 'iron'
to	*toki* 'Japanese Ibis' *tomo* 'friend'	*toku* 'benefit' *tomoni* 'together'

Table 1: A stimuli table varying in onset nasality and vowel types.

2.2. Recording sessions

We recruited four male and four female participants who have performed Rakugo storytelling as a member of a Rakugo club (mean years of experience = 3.5). The age of participants was 18 to 23 at the

time of recording (mean = 21.2). All participants utilize Edo Rakugo, a style that uses the Edo dialect (old Tokyo Japanese).

The participants produced the Rakugo style speech while seated in a *seiza* style, where the participant sits on a cushion following the mode of Rakugo performance. The non-Rakugo speech was elicited while sitting on a chair. After reading a snippet from a Rakugo story for approximately one minute to immerse themselves into Rakugo speech style, participants produced the target items in the Rakugo style. The non-Rakugo sentences were elicited after participants changed their sitting position. The recording was conducted in a quiet room using the audio recording software Audacity.

2.3. Annotation

Target words in the elicited data set were first processed using Praat [4]. Segment-level annotation was conducted manually by examining the periodicity of vowels as well as information from the spectrogram.

A single interval was marked for vowels and nasals. Voiceless plosives were annotated for their closure duration and voice onset time (vot), while voiced plosives were annotated for the closure duration (cvo) and vot. In this paper, boundaries were selected in a conservative manner. An example of an annotated target word is shown in Figure 1.

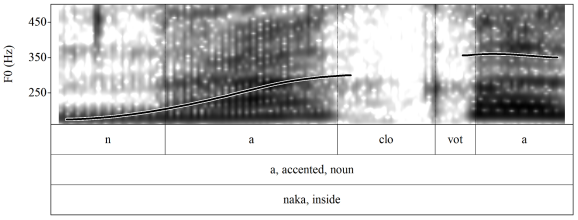

Figure 1: Example of the annotation.

Annotations were conducted by three Japanese-speaking assistants, two of whom were trained in linguistics. All items were double-checked for their annotation accuracy.

3. RESULTS

3.1 Vowel Quality

R [5] was used to create vowel charts to examine the distribution of vowels in the Rakugo and non-Rakugo styles, using the extracted values of F1 and F2 from the mid-point of an annotated vowel interval. Aberrant formant values (74 out of 1515 values) were excluded from analysis by calculating Cook's distance for each pair of gender and vowel. Gender differences discussed in a previous study [6] are also observed in our data, as shown in Figures 2 and 3.

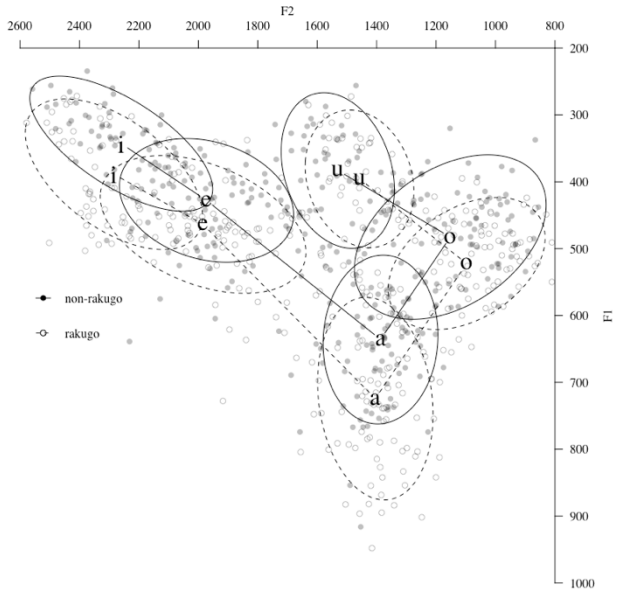

Figure 2: Vowel plot of male speakers.
Non-Rakugo in solid lines and Rakugo in dotted lines.

As Figure 2 shows, male speakers tend to produce vowels in a lower position in the Rakugo style than in the non-Rakugo style. Such lowering tendency has been reported for non-high vowels [o], [e], and [a] in the previous study where syllables without any context are used as stimuli [6]. However, in the current study in which speakers are asked to speak words in different styles, the high vowel [i] is also subject to the lowering tendency in the Rakugo style.

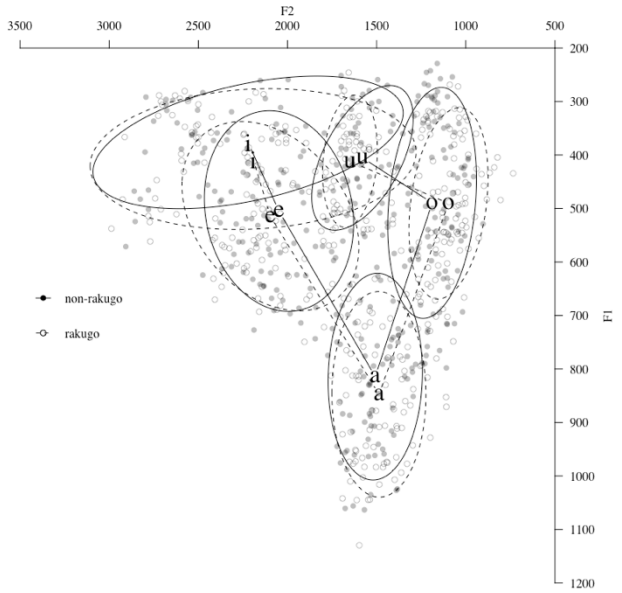

Figure 3: Vowel plot of female speakers.
Non-Rakugo in solid lines and Rakugo in dotted lines.

Although [6] describe that female speakers produce vowels in the Rakugo style in less varied styles, their results do not extend to the vowel pattern in Figure 3 because female speakers show variation to a lesser degree, instead of less variation.

3.2 Duration

An examination of the first syllable in the same target words read in both styles revealed that speakers tend to produce the target syllable longer in the Rakugo style than in the non-Rakugo style. To show this relationship, a correlation plot is created.

Figure 4 displays durational differences between Rakugo and non-Rakugo styles of the same target words. The duration of the first syllable in the non-Rakugo style is plotted on the x-axis, and the duration in the Rakugo style on the y-axis. As shown in the regression line plotted by R, a positive correlation was found (R^2 = 0.35). Round-shaped dots above the line represent cases when a speaker produces the first syllable of a word as longer in the Rakugo style than in the non-Rakugo style. On the other hand, tokens in which a speaker produces the first syllable as longer in the non-Rakugo style are triangle-shaped. The distribution of the dots indicates that Rakugo speakers tend to produce each segment longer in the Rakugo style.

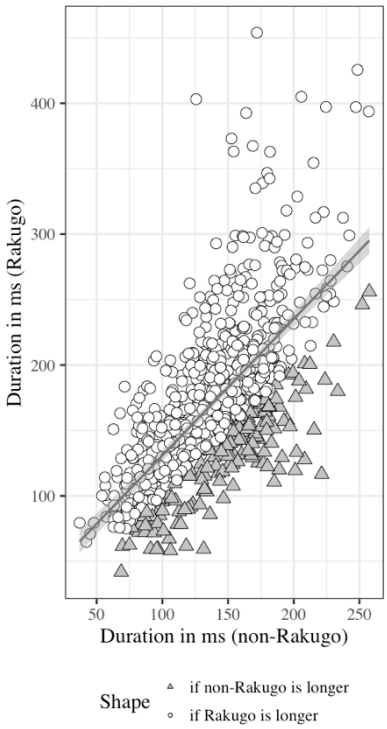

Figure 4: Correlation between non-Rakugo and Rakugo (Duration).

Durational differences between the two styles in individual tokens by speakers are shown in Figure 5.

Six out of eight speakers used durational cues by producing longer vowels (more than 70% of the tokens) in the Rakugo-style speech.

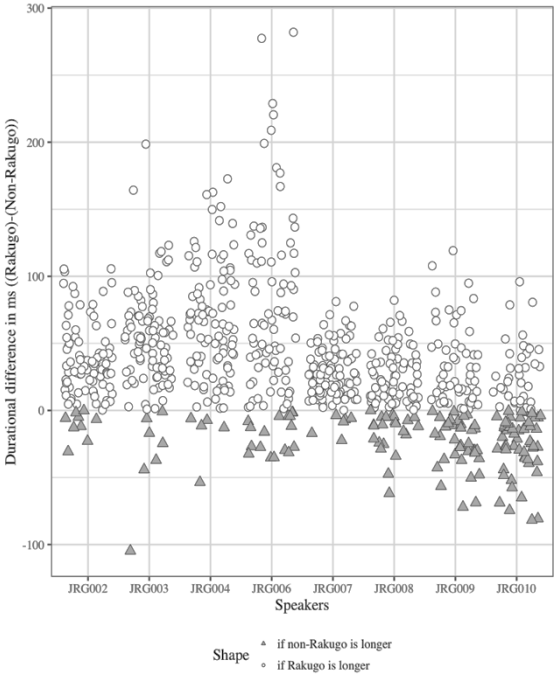

Figure 5: Durational difference between the non-Rakugo and Rakugo styles by Speakers.

		non-Rakugo	Rakugo
naki	unaccented	16 / 24	1 / 24
niki	unaccented	16 / 24	0 / 24
nishi	unaccented	17 / 24	1 / 24
teki	unaccented	18 / 24	0 / 24
tetsu	unaccented	22 / 24	0 / 24
toki	accented	16 / 24	0 / 24
toku	unaccented	18 / 24	0 / 24
	Total	123 / 168	2 / 168
		73.2%	1.1%

Table 2: A table showing tokens with the devoiced vowel in the final syllable.

3.3 High vowel devoicing

High vowel devoicing is a feature of Japanese that is well-reported (cf. [7] and reference therein). The number of words produced with a devoiced vowel in the final syllable is summarized in Table 2. While

target words in non-Rakugo speech undergo high vowel devoicing as expected, the same vowel produced in words with the Rakugo style is almost never produced with a devoiced vowel.

3.4 Pitch range

Speech in performance arts often exhibits widened speech range to increase the transmission of information. To check whether the Rakugo style displays this feature, we calculated the pitch difference between the maximum F0 and minimum F0 of the first and second vowels of each word, considering differing accent patterns. The results show that a larger pitch range is used in the Rakugo style than in the non-Rakugo style.

The data we analyzed consists of words in two accent types: falling pitch in the penultimate syllable (the -2 accent condition) and no falling pitch (the unaccented condition). Words with the -1 accent condition were present in the stimuli but excluded from the analysis due to overlapping effects in pitch accent patterns with unaccented words when the -1 accent words are produced in isolation.

As the words in the -2 accent condition contain a pitch drop between the first and second syllables, we subtracted the value of the minimum pitch of the second syllable from the value of the maximum pitch of the first syllable. Tokens were excluded if the first mora was produced with a lower pitch than the second mora. The distribution of the dots indicates that speakers use a larger pitch range in the Rakugo style than in the non-Rakugo style, as shown below in Figure 6.

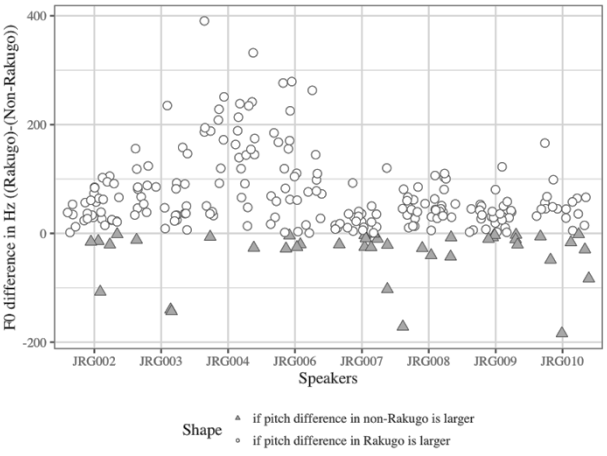

Figure 6: F0 difference in the -2 accent condition between the non-Rakugo and Rakugo styles by Speakers

As unaccented words in Japanese have a pitch rise from a low initial F0, the difference between the minimum pitch of the word-initial mora and the value of the maximum pitch of the second mora was calculated. Most words are plotted above the y = 0 line in Figure 7, which means that the pitch range in the first two morae of unaccented words also becomes larger in the Rakugo style than in the non-Rakugo style.

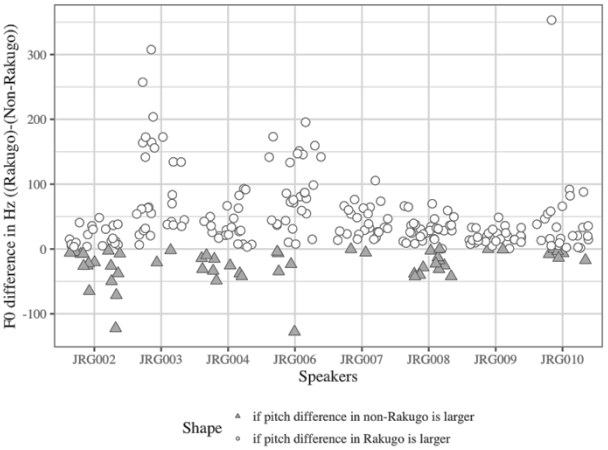

Figure 7: F0 difference in the unaccented accent condition between the non-Rakugo and Rakugo styles by Speakers

4. DISCUSSION AND CONCLUSION

Our research reveals that Rakugo-style speech exhibits dispersed vowels, longer syllable duration, lower rate of high vowel devoicing, and larger pitch range. Although in their normal voice participants devoiced their high vowels, due to the nature of rakugo, this common phenomenon of Tokyo Japanese is not observed in Rakugo style. However, since Rakugo speakers do not acquire but learn Rakugo-style speech, these segmental or suprasegmental modifications in the Rakugo style can be subsumed in an optional effort to convey Rakugo stories correctly and dramatically. The inter-speaker variations support the possibility that the phonetic characteristics in the Rakugo style are not obligatory.

While verbal performance or acting is well-studied in terms of its aspects of history and culture, the speech patterns in the performance often remain understudied. In this study, we focused on the speech patterns displayed in Rakugo, one of such verbal performances. More research on the phonetic characteristics of Rakugo-style speech such as the use of non-modal phonation or the implementation of pitch excursion will further elucidate what makes a Rakugo-style Rakugo.

5. ACKNOWLEDGEMENTS

We thank the Rakugo performers who participated in the data collection. This work was partially supported by JSPS KAKENHI Grant #20K00578

(POP), and the ILCAA joint research project "Phonetic typology from cross-linguistic perspectives (PhonTyp)" (2021-2023).

6. REFERENCES

[1] Kawahara, S. 2013. The phonetics of Japanese maid voice I: A preliminary study. *On-in Kenkyuu [Phonological Studies]* 16, 19–28.

[2] Kawahara, S. 2016. The prosodic features of the 'moe' and 'tsun' voices. *Journal of the Phonetic Society of Japan* 20(2), 102-110.

[3] Utsugi, A., Wang, H., Ota I. 2019. A voice quality analysis of Japanese anime. Proceedings of the *International Congress of Phonetic Sciences* 19, 1853-1857.

[4] Boersma P., Weenink, D., 2019. Praat: doing phonetics by computer. Accessed: Dec. 17, 2019. [Online].

[5] R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

[6] Kamano, S., Tsujita, R., Sakamoto, C., Lee, S. J. 2022. Vowel production in Rakugo Speech. *Proceedings of the 36th General Meeting of the PSJ*, Kobe, Japan. 151–155.

[7] Fujimoto, M. 2015. Vowel devoicing. In: Kubozono, H. (ed) *Handbook of Japanese Phonetics and Phonology*, De Gruyter. 167–214.