Neutralization of voicing distinction of stops in Tohoku dialects of Japanese: Field work and acoustic measurements

Ai Mizoguchi[1,2]*, Ayako Hashimoto*[3]*, Sanae Matsui*[4]*, Setsuko Imatomi*[5]*, Ryunosuke Kobayashi*[4]*, and Mafuyu Kitahara*[4]

[1]Maebashi Institute of Technology, Gunma, Japan
[2]National Institute for Japanese Language and Linguistics, Tokyo, Japan
[3]Tokyo Kasei-gakuin College, Tokyo, Japan
[4]Sophia University, Tokyo, Japan
[5]Mejiro University, Tokyo, Japan

aimizoguchi@maebashi-it.ac.jp, hassy@kasei-gakuin.ac.jp, s-matsui-9hf@eagle.sophia.ac.jp, imatomi@mejiro.ac.jp, q-wang-2v2@eagle.sophia.ac.jp, mafuyu@sophia.ac.jp

Abstract

Research on Tohoku dialects, which is a variety of Japanese, has found that the voiceless stops /k/ and /t/ in the intervocalic position are frequently realized as voiced stops. However, the phenomenon has mainly been judged aurally in the Japanese linguistics literature and has not been confirmed by acoustic measurements. We measured the VOT of data originally collected in the survey of Tohoku dialects by [1]. The data used in this study includes two age groups from eight sites. The results demonstrate that for word medial stops, the VOT distribution of voiced and voiceless stops largely overlapped, while, the laryngeal contrast was maintained for the word initial stops. Intervocalic voicing neutralization was confirmed by quantitative acoustic measurements. The effects of neighboring vowels were also investigated to show that height, but not duration, had a significant effect on voicing neutralization. Our results shed light on the phonetic nature of Tohoku dialects as well as on their phonological structure, such as the role of voicing contrast.

Index Terms: intervocalic voicing, neutralization, Tohoku dialects, VOT

1. Introduction

Tohoku dialects are spoken in the Tohoku district, the northern part of Honshu (the mainland) in Japan. They show some salient characteristics in pronunciation that distinguish them from other dialects in Japan [2–4]. Tohoku dialects can be further divided into two groups depending on their properties: northern Tohoku dialects and southern Tohoku dialects.

The most prominent characteristic of the consonants found in Tohoku dialects is that the voiceless stops /k/ and /t/ are voiced intervocalically in both northern and southern Tohoku dialects. For example, /atama/ 'head' is pronounced as [adama] and /kaki/ 'persimmon' as [kaɡi]. On the contrary, word-initial voiceless stops are not voiced. This fact suggests that voiceless stops are voiced intervocalically because they assimilate into the voicing feature of neighboring vowels, which are fundamentally voiced, and that they are not voiced word-initially because there is no vowel before them. As a result of the intervocalic voicing of voiceless stops, certain words are merged in the southern Tohoku region. That is, /ito/ 'string' is pronounced as [ido] and /ido/ 'well' as [ido]. On the contrary, such a merging is avoided in northern Tohoku dialects because the voiced obstruents /b/, /d/, and /z/ are pre-nasalized intervocalically as [^{m}b], [^{n}d], and [^{n}dz], respectively. As for /g/, a velar stop, it is fully nasalized as [ŋ]. Since /ito/ is pronounced as [ido], while /ido/ as [i^{n}do] in northern Tohoku dialects, the distinction is maintained.

[2] surveyed the sounds of Tohoku dialects and suggested the effects of adjacent vowels on intervocalic voicing. In terms of acoustic measurements, VOT, which is a widely used acoustic metric of voicing [5–8], was investigated for the word-initial stops in Tohoku dialects showing a bi-modal VOT distribution for voiced and voiceless stops [9]. Regarding the utterance position, the word-initial stops in an isolated utterance tend to have a longer VOT than the ones in a sentence in various languages [5, 6, 10]. Word-medial stops tend to have a greater voiced portion during the closure in American English than word-initial stops do [11]. However, few studies have investigated the VOT of word-medial stops in a language that features voicing neutralization.

In this study, we measured the VOT of the data collected in the recent survey by [1], which was conducted to investigate present-day Tohoku dialects, and confirmed intervocalic voicing neutralization using quantitative acoustic measurements.

2. Method

The speech data for the present study were originally collected for a project on the phonological descriptions of Tohoku dialects [1]. A brief overview of the recording in the project and the description of the portion of data in the present study along with the acoustic and statistical analyses are given in this section.

2.1. Recording sites

The left panel in Figure 1 shows the map of the Tohoku district in Japan. The recording sites in the survey by [1] are shown in the right panel. To capture the comprehensive characteristics of Tohoku dialects, distinct sites with sufficient distance from the geographical, historical, and cultural points of view were chosen.

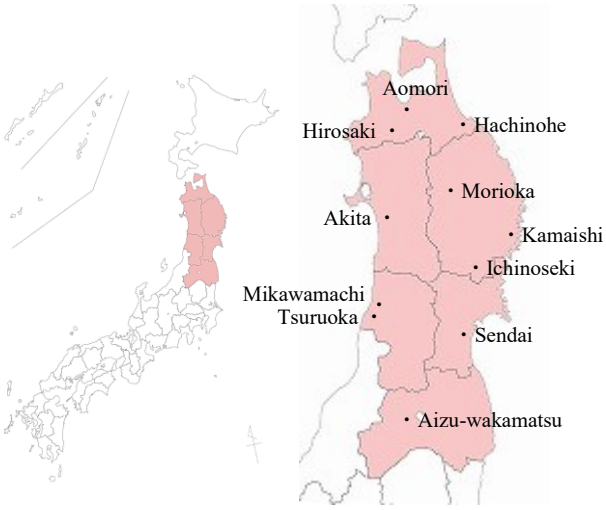

Figure 1 Left: *Tohoku district in Japan.* Right: *Eleven recording sites in the survey [1].*

2.2. Participants

Tohoku dialects were recorded from 2012 to 2016 in 11 sites covering all the six prefectures in the Tohoku district. The total number of recorded speakers was 61, whose ages ranged from 10 to 92. In the present study, the data on 24 speakers from eight sites in two age groups (69–85 years and 33–53 years) were analyzed. Table 1 summarizes the eight sites, age groups, and number of speakers.

Table 1: *Number of participants in each site.*

Site	69–85 years	33-53 years
Aomori	0	2
Hachinohe	2	0
Akita	2	2
Morioka	1	2
Kamaishi	2	0
Ichinoseki	2	1
Tsuruoka	2	2
Aizu-wakamatsu	2	2

2.3. Materials and procedures

The picture task, which involved 48 photos or illustrations presented to participants sequentially, was conducted by one of the authors. Participants were asked to pronounce the name of the depicted object in their usual spoken language. The recording was done at town halls or the participant's home using a SONY ECM-MS957 microphone on a SONY PCM-D50 recorder (16bit, 44kHz).

Words containing coronal or velar stops (/t/, /d/, /k/, /g/) were selected for acoustic analysis. Table 2 presents the set of expected words shown to participants by photos or illustrations. Even when participants produced a different word from the one expected, if it contained a coronal or velar stop, the words were included in the analysis. As a result, the total number of analyzed words was 84.

Table 2: *Word list.*

	Initial	Medial	Both
Coronal	tokei takigi daikon	gitaa kutsushita hata natto mado budoo	tomato
Velar	kutsu kutsushita kisha kusa kuchibashi kujira gitaa	tsuki, suki fuki, yuki takigi, azuki oke, tokei yakan, mikan okashi, suika shika chikarakobu daikon, neko baiku, omikuji tsukushi hoshigaki nagagutsu	kamakiri kaki kiku

2.4. Acoustic measurements

The words were transcribed and segmented using Praat [12] by trained phoneticians and checked by another phonetician in our group. The target stops were visually identified by the clear existence of a closure and the following burst. The start of the voicing was identified by the visible concentration of energy below the 1kHz range. Only the segments with a clear burst and voicing were used to measure the VOT. In other words, those without a visible burst or with a devoiced vowel were omitted from further analyses. Vowels were identified by a visible structure of F2 and F3, which did not necessarily coincide with the start or end of the voicing. The duration and pitch of the surrounding vowels were also measured. Figure 2 shows an example utterance in which the burst and voice onset for the word-initial stop were measured and those for the word-medial stop were not measured due to continuing voicing throughout the closure.

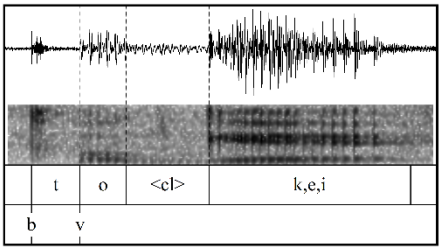

Figure 2: *Examples of the measured burst (b) and voice onset (v) for [t] in /tokei/. The burst and voice onset were not measured for [k] due to continuing voicing throughout the closure (<cl>).*

2.5. Statistical analyses

A linear mixed-effects model analysis was conducted in R [13], using the lme4 [14] and lmerTest [15] packages. The models were selected using a step-down model building process and the log-likelihood comparisons.

3. Results

3.1. VOT by segment position

We measured and analyzed the voice onset time of the word-initial and -medial stops /t, d, k, g/ in 84 varieties of stimulus words uttered by 24 native speakers of Tohoku dialects. This resulted in 830 target segments. For the word-medial position, the cases in which the voicing from the previous vowel was continued throughout the closure were excluded from the analysis due to the impossibility of identifying the voice onset for the target segment. Due to this measurement difficulty and the selection of the stimulus words, the numbers included in the analysis became imbalanced between the segments and segment positions (Table 3). Thus, the results, especially those from the word-medial /g/, should be considered as a reference.

Table 3: *The number of target segments included in the analysis, VOT and its SD, and the number of fully pre-voiced segments in the word-medial position excluded from the analysis.*

Segment	N	VOT mean (ms)	SD	Fully pre-voiced
Word-initial				
/t/	36	47.56	15.98	-
/d/	23	14.22	9.00	-
/k/	179	68.05	25.99	-
/g/	21	22.95	23.97	-
Word-medial				
/t/	117	19.27	9.34	3
/d/	24	15.67	8.24	37
/k/	426	37.47	19.10	59
/g/	4	26.00	4.08	8

Figure 3 shows the VOT of the word-initial and -medial /t, d, k, g/ uttered by the 24 native speakers of Tohoku dialects.

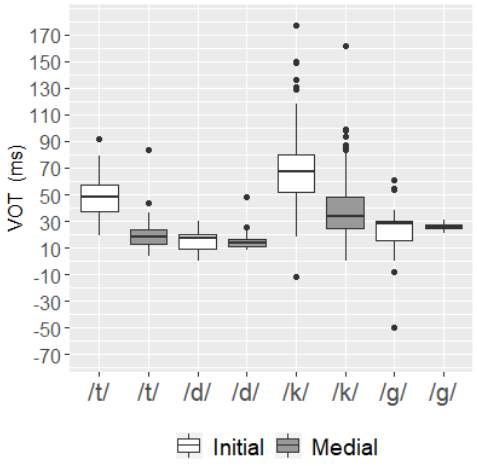

Figure 3: *Word-initial and -medial /t, d, k, g/ VOT.*

The selected mixed-effects linear regression model predicting the VOT from the fixed effects of the interactions of voicing contrast (voiced/voiceless), place of articulation (alveolar/velar), and segment position (initial/medial), and the segment duration and the following segment, with random intercepts of speakers and stimulus words, revealed that the interaction between voicing and place of articulation had a significant effect (β=-30.46, t=-3.35, p<.01). A pairwise post hoc test comparing the estimated means showed that the VOT of the word-initial /t/ was significantly longer than that of the word-medial /t/ (p<.001) and that the VOT of the word-medial /t/ was not significantly different than that of the word-medial /d/ (p=.98). In addition, the VOT of the word-initial /k/ was significantly longer than that of the word-medial /k/ (p<.001).

Figures 4 and 5 show the case number normalized histograms of the word-initial (Figure 4) and -medial (Figure 5) VOT of /t/, /d/, /k/, and /g/. For the word-initial stops, the VOT distribution looks bimodal with a slight overlap in the middle (Figure 4). By contrast, for the word-medial stops, there is no large variability of VOT across the voiced and voiceless stops (Figure 5).

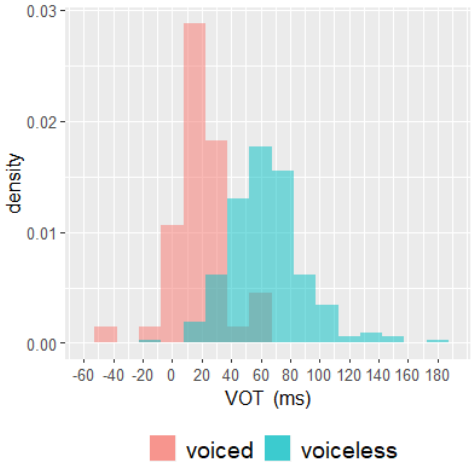

Figure 4: *Normalized word-initial /t, d, k, g/ VOT.*

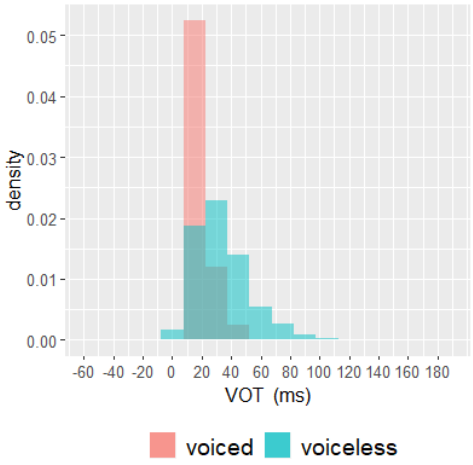

Figure 5: *Normalized word-medial /t, d, k, g/ VOT.*

3.2. Effects on the VOT of the word-medial /t/ and /k/

The effects on the VOT of the word-medial /t/ and /k/ were further explored and a mixed-effect model was chosen, which consisted of the fixed effects of the place of articulation, stimulus word duration, previous segment, and following segment, and a random intercept of speakers. Table 4 shows the

significant effects of the previous and following segments on the VOT of the word-medial /t/ and /k/.

Table 4: *Significant fixed effects in the mixed-effects model on the VOT of the word-medial /t/ and /k/.*

Predictor	Estimate β (ms)	*t*-value	*p*-value
previous [e]	16.00	3.48	<.001
previous [i]	8.68	2.99	<.01
previous [o]	9.17	1.99	<.05
previous [ɯ̟]	-7.17	-2.28	<.05
following [aː]	-10.58	-2.10	<.05
following [i]	20.74	6.55	<.001

To narrow the adjacent vowel effect, the vowel durations were used as the fixed effects (*vot ~ place + stimulus word duration + previous/following segment * previous/following segment duration + (1|speaker)*). The models showed no significant effects of adjacent vowel durations.

The f0 values at the 3/4 time points for the previous vowel and at the 1/4 time points for the following vowel were also used and the model (*vot ~ place of articulation + stimulus word duration + previous/following vowel * previous/following vowel f0 * speaker's sex + previous/following vowel * previous/following vowel f0 * speakers' sex + (1|speaker)*) showed that for male speakers, f0 had a significant effect on VOT when the previous vowel was [e] (β=-.90, t=-2.07, p<.05) and [i] (β=.70, t=2.83, p<.01). It also showed a significant effect on VOT when the following vowel was [e] (β=-.19, t=-2.49, p<.05) and [oː] (β=-.77, t=-2.28, p<.05).

4. Discussion

4.1. Intervocalic voicing neutralization

Figure 3 and the linear mixed-effects analysis showed that the VOTs of /t/ and /k/ were significantly shorter in the word-medial position than in the word-initial position. The VOT of the word-medial /t/ did not significantly differ than the word-medial /d/. The comparison between the word-medial /k/ and /g/ was not reported here due to the small number of measurable cases for the word-medial /g/, but 59 cases of the word-medial /k/ were produced with full pre-voicing (Table 3). In addition, Figure 5 shows the less variability across the voiced and voiceless stops and overlap of the VOT in the word-medial position. These results showed that the voicing contrast tended to disappear in the word-medial position in these dialects. This neutralization has long been acknowledged as characteristic of Tohoku dialects in the Japanese linguistics literature [2–4], and the current results confirmed this well-known phenomenon using quantitative measurements.

4.2. Effects of adjacent vowels

Table 4 shows the effects of adjacent vowels on the VOT of the word-medial voiceless stops. If the previous vowel was a high vowel [i], or a mid-vowel [e] or [o], the VOT was predicted to be longer than the baseline mean value. If the previous vowel was [ɯ̟] (devoiced /u/), the VOT was predicted to be shorter than the baseline.

As for the following vowel, if it was [aː] (long /a/), the VOT was shorter than the baseline and if it was [i], the VOT was longer than the baseline.

VOT is affected by the phonological context [2, 16]. According to [16], VOTs are longer before high vowels than before mid- or low vowels. Indeed, if the vocal fold tension is high as occurs with high vowels, voicing is difficult and thus, the VOT becomes longer [17]. This explanation is mostly compatible with our finding that when the following vowel was a high vowel [i], the VOT was longer, whereas if it was a low vowel [aː], the VOT was short. However, not all the high and low vowels used in the current experiment showed the same significant effects. It also needs to be examined further why the devoiced /u/ had a shorter VOT of the following stop. To narrow the adjacent vowel effects, vowel duration and f0 were added into the analyses. The results showed no effects of vowel duration on the VOT irrespective of the vowels. In American English, the previous vowel duration is longer before voiced consonants than voiceless consonants [18, 19]. The difference between previous vowel duration in American English plays a role in distinguishing the following voiced/voiceless contrast. However, for the Tohoku dialects in the current study, although the word-medial voicing contrast was often neutralized, the previous vowel duration was not relevant for the VOT. As described in the Introduction, word-medial voiced and voiceless stops are distinguished by the pre-nasalization of voiced stops in the northern variations of Tohoku dialects, but not in the southern variations.

In terms of f0, although a few significant effects were seen, the pitch accent, which is lexically assigned to each word in Japanese, was not controlled for the stimulus words and this had a huge effect on vowels. Thus, it is not ideal to include this in the interpretation at this point. However, in discussing the vocal fold tension, it will be necessary to see the f0 effect on the VOT.

A closer look at vowel quality including vowel duration and f0 with more controlled stimuli may lead to a discussion on whether voicing neutralization occurs due to phonological and/or physiological factors.

5. Conclusion

The intervocalic voicing neutralization in Tohoku dialects is a widely known phenomenon. However, quantitative measurements and statistical analyses have long been lacking. In the future, it will be necessary to analyze more data from all sites across the Tohoku district to investigate its sociolinguistic aspects. Analyzing more data will also provide better statistical power by increasing the measurable target segments, especially voiced ones. In addition, it will be necessary to closely examine fully pre-voiced cases with a negative VOT, which could not be incorporated into our analysis. Finally, because word-medial voicing is diminishing among the younger generation [2], it is desirable to explore different age groups and the socio-demographic aspects of these dialects.

6. Acknowledgements

We are truly grateful to the participants of the survey of Tohoku dialects. This research was supported by JSPS Kakenhi 24520438.

7. References

[1] A. Hashimoto, "Tohoku hougen ni okeru kagyou tagyou shi 'in no yuuseika ni tsuite: Tohoku hougen onsei chousa kara II [On the voicing of /k/ and /t/ in Tohoku dialects: A survey on sounds of Tohoku dialects II]," *Tsuda Journal of Language and Culture*, 34, pp. 88–102, 2019.

[2] J. Ohashi, *Tohoku hougen onsei no kenkyuu [Research on sounds of Tohoku dialects]*, Tokyo: Oufuu, 2002.

[3] M. Shibatani, *The Languages of Japan*. Cambridge: Cambridge University Press, 1990.

[4] N. Tsujimura, *An Introduction to Japanese Linguistics*. Oxford: Blackwell Publishers inc, 1996.

[5] L. Lisker, and A. S. Abramson, "A cross-language study of voicing in initial stops: Acoustical measurements," *Word*, 20, pp. 384–422, 1964.

[6] L. Lisker, and A. S. Abramson, "Some effects of context on voice onset time in English stops," *Language and Speech*, 10, pp. 1–28, 1967

[7] T. Cho, and L. Ladefoged, "Variation and universals in VOT: Evidence from 18 languages," *Journal of Phonetics*, 27, pp. 207–229. 1999.

[8] A. S. Abramson, and D. Whalen, "Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions," *Journal of Phonetics*, 63, pp. 75–86, 2017.

[9] M. Takada, *Nihongo no gotou heisa'on no kenkyuu: VOT no kyoujiteki bunpu to tsuujiteki henka [Research on the word-initial stops of Japanese: Synchronic distribution and diachronic change in VOT]*, Tokyo: Kurosio, 2011.

[10] G. J. Docherty, *The timing of voicing in British English obstruents*, Berlin: Foris Publications, 1992.

[11] L. Davidson, "Variability in the implementation of voicing in American English obstruents," *Journal of Phonetics*, 54, pp. 35–50, 2016

[12] P. Boersma, and D. Weenink, Praat: doing phonetics by computer. Version 6.1.08, retrieved 21 December 2019 from http://www.praat.org/

[13] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Austria: Vienna, 2019. URL: https://www.R-project.org/.

[14] D. Bates, M. Maechler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, 67(1), pp. 1–48, 2015.

[15] A. Kuznetsova, P.B. Brockhoff, R.H.B. Christensen, "lmerTest package: tests in linear mixed effects models," *Journal of Statistical Software*, 82(13), 2017.

[16] D. H. Klatt, "Voice Onset Time, frication and aspiration in word-initial consonant clusters," *Journal of Speech and Hearing Research*, 18, pp. 686–706, 1975.

[17] P. Auzou, C. Ozsancak, R. J. Morris, M. Jan, F. Eustache, and D. Hannequin, "Voice onset time in aphasia, apraxia of speech and dysarthria: a review," *Clinical Linguistics & Phonetics*, vol.14, no. 2, pp. 131–150, 2000.

[18] A. S. House, "On Vowel Duration in English," *Journal of the Acoustical Society of America*, 33, pp.1174–1178, 1961.

[19] L. J. Raphael, "Preceding vowel duration as a cue to the perception of the voicing characteristic of word - final consonants in American English," *Journal of the Acoustical Society of America,* 51, pp. 1296–1303, 1972.