A Phonetic Study of Plosives in the Hakodate Dialect

Michinori Suzuki
International Christian University

1 Introduction

This paper reports findings from experiments conducted in Hakodate, Hokkaido, the northernmost island of Japan. In particular, this study focuses on the voicing contrast in Hakodate dialect.

Kawahara et al. (2018) points out that it is oversimplifying to explain voicing contrast by the existence of "vocal folds vibration". The voicing contrast can be observed in varieties of acoustic levels (=multiplicity of acoustic cues) as (1) shows.

(1) a. vocal folds vibration during a consonant
b. the duration of consonants
c. the duration of a preceding vowel
d. F0 of the surrounding vowels
e. F1 of the surrounding vowel
f. F1 cut back phenomenon in a following vowel
g. the strength of the burst of consonants

Based on the observation of multiplicity of acoustic cues, Kawahara et al. (2018) mainly refers to the phonetic module, the phonological module, and the interface between them.

Shimizu (2018) investigated the voicing contrast of initial plosives in Japanese, English, Polish, Chinese, Thai, and Burmese. Shimizu primarily identified VOT and F0 as acoustic cues in the six languages, and also refers to burst spectrum, F0 contour, and the tongue contact area. The author concluded that VOT is a useful acoustic cue for distinguishing the voicing contrast except for Burmese. Although other five languages distinguish VOT between voiced and voiceless non-aspirated plosives, Burmese does not have a clear distinction between them. In particular, he argued that the vocal folds vibration is created by the complementary functioning of some factors such as airflow from our lung, the tension of laryngeal muscles, and the difference in the air pressure between subglottis and supraglottis, as mentioned in previous studies.

As investigated by Shimizu (2018), voicing contrast exists in many languages. For example, Japanese has words *toki* 'time' and *doki* 'cray pot.' English also has words 'bet' and 'bed.' However, at the same time, languages use different acoustic cues for distinguishing voicing categories (Abramson & Lisker 1970, Shimizu. 1989). For instances, the voice onset time (VOT) duration for [b] in English and Spanish is less than 30 ms and 15 ms respectively, and it must be negative in Thai. In addition to this, the ways of distinguishing voicing contrast could be different within the same language. Inui (1994) reported that in north-eastern Japanese dialect, [p] is realized as voiced and [b] is realized as pre-nasalized, even though [p] is usually voicelss and [b] is usually voiced.

Ogawa et al. (2017) examined voicing in plosives from Kanto and Hokkaido Japanese using Electroglottograph (EGG) data. They report preliminary findings from four speakers (two from each dialect) in their 20s. They found that vocal folds vibration during the production of voiced plosives discontinues with Kanto speakers, while it continues with Hokkaido speakers. The two participants from Hokkaido were born and raised in Sapporo, which is an inland city in Hokkaido.

According to Takada (2011), dialects in Hokkaido could be roughly divided into two areas: coastal areas around south-west peninsula (Oshima peninsula), and inland areas. Therefore, the ways of distinguishing voicing contrast might be different between the two areas.

Though acoustic signals demonstrate multiple cues, it cannot directly observe what is happening to our vocal folds. In order to observe the vocal folds movment, this study uses electroglottograph (EGG). There are few studies on Japanese dialects from a detailed viewpoint of phonetics. Especially, no phonetic study had been conducted regarding voicing in Hakodate dialect. This study fills the gap by examining the voicing in plosives in Hakodate dialect both acoustically and articulatorily.

The research question in this paper is to understand the ways of distinguishing voiced and voiceless plosives in Hakodate dialect. The findings indicate that there is almost no acoustic distinction between voiced plosives and voiceless plosives in word-initial positions. It means that those sounds are all acoustically voiceless. However,

Proceedings of *AJL 3*

the results also suggest that VOT might be an acoustic cue for distinguishing voiced and voiceless plosives in word-medial positions.

2 Experiment

2.1 *Participants* Target participants are speakers who were born and raised in five regions in Hokkaido (Hakodate, Otaru, Asahikawa, Kitami, and Kushiro). Since children's languages could be affected by the environment, parents and grandparents of the target participants were also born in the same region. Since Hokkaido is large, one single region cannot represent the whole picture of Hokkaido dialect. Therefore, participants were selected from five different regions in Hokkaido as in Figure 1.

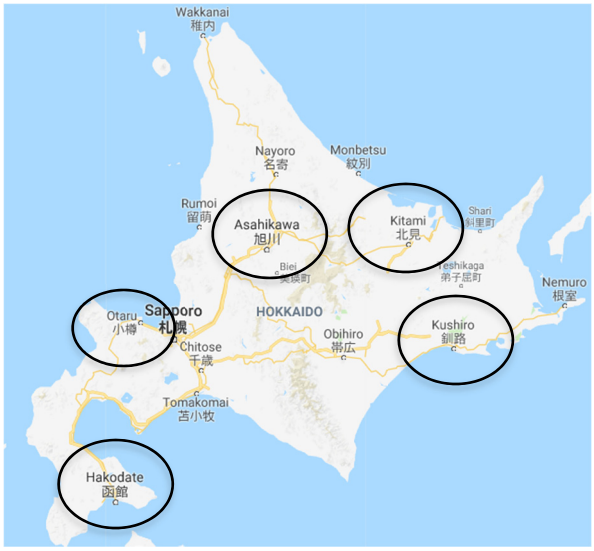

Figure 1: The regions where experiments were conducted (Hakodate, Otaru, Asahikawa, Kitami, and Kushiro (from lower-left to right)).

The regions in Figure 1 were chosen based on several criteria. First, an area must have a sizable population (more than 100 thousand people[1]). Second, regions are not adjacent to one another. Sapporo, a prefectural capital, was avoided since there was an organized migration around the end of 19th century. At that period, many people migrated to Sapporo from other regions. As a result, their dialect could possibly be affected by other dialects. In each region, one to six speakers were recorded. In total, twenty speakers: twelve females and eight males, were recorded. The target age group was from the 50s to 60s. In Hakodate, three speakers were recorded.

2.2 *Stimuli* The stimuli consist of the following six types.

Type	Target Consonant	Examples
a. C1 V C2 V	initial voiced plosives	[doki] 'clay pot'
b. C1 V C2 V	initial voiceless plosives	[kuma] 'bear'
c. C1 V C2 V	initial nasals	[mori] 'foresst'
d. C1 V C2 V	medial voiced plosives	[çigo] 'protection'
e. C1 V C2 V	medial voiceless plosives	[kata] 'shoulder'
f. C1 V C2 V	medial nasals	[kuni] 'country'

Table 1: Six types of the stimuli.

The stimuli consist of 96 words (18 words for each category for plosives and 12 words for each nasal category). These 96 words have variations in voicing (plosives [b, d, g] vs. [p, t, k] as well as nasals [m, n]), places of articulation (labial [p, b, m], coronal [t, d, n], dorsal [k, g]), pitch accents (LH, HL), positions (word-initial and word-medial), and heights of the surrounding vowels (low [a], mid [e, o] and high [i, u]). All the words consist of two morae except for two words, and the target sounds are placed in the word-initial consonants or the word-

[1] The population of each city is as follows. Hakodate: 271479, Otaru: 125028, Asahikawa: 347207, Kitami: 122198, Kushiro: 178394 (Statistic Bureau of Japan, 2015).

medial consonants. In this paper, C1 and C2 indicate initial and medial consonants respectively. For example, the words in category a. and b. have voiced and voiceless plosives in C1 position as in [doki] 'clay pot' or [kuma] 'bear.' Similarly, the words in category d. and e. have voiced and voiceless plosives in C2 positions such as [higo] 'protection,' or [kata] 'shoulder.' The target words in category c. and f. have nasals such as [mori] 'forest,' or [kuni] 'country.' Since this study focuses on plosives, results of nasals are not reported in this paper. 18 words were prepared for category one, two, four and five, and 12 words were created for category three and six. A complete list of the stimuli can be found in the appendix.

2.3 *Data Collection* Participants were asked to read out randomized stimuli words five times in a frame sentence *korega X desu* 'This is X'. The speech was recorded using a head-worn microphone (XLR Shure WH-30) and Electroglottograph (EG2-PCX2, Glottal Enterprise) with a Roland USB Audio Interface Rubix 24 that receives analog EGG signal from the microphone and EGG machine. Electroglottograph (EGG) is a non-invasive method to measure the openings and closures of vocal folds from electrical resistance between two electrodes on a participant's throat. During the experiment, all stimuli were presented in a PowerPoint slideshow, and an experimenter advanced the slide. Participants handed in a consent form and demographic questionnaire before they began the experiment.

2.4 *Processing* The three medial repetitions of 96 words with target sounds were analyzed since speakers tended to make more errors in the first repetition. The last repetition was also excluded due to the possibility of priming. 288 tokens were obtained from each of the three participants with Hakodate dialect. In total, 648 tokens were processed for further analysis. These sound files were annotated using software Praat (Boersma and Weenink, 2018), and the graphs and statistical analyses were made using R (R Core Team, 2017). Based on the waveform of acoustic signals, the closure durations of plosives with pre-voicing were labeled as closure voicing and others were labeled as closure.

3 Results

3.1 *Acoustic Results* First, two histograms in Figure 2 indicate the number of voicing in word-initial (left) and word-medial (right) positions. Black bars demonstrate "closure" and grey bars indicate "closure voicing." Closure means that the voicing discontinues and closure voicing indicates that the voicing continues. The x-axis indicates the segment. The segments are grouped by voicing status (/p, t, k/ and /b, d, g/). The y-axis indicates the number of occurrences, though the actual count is not important.

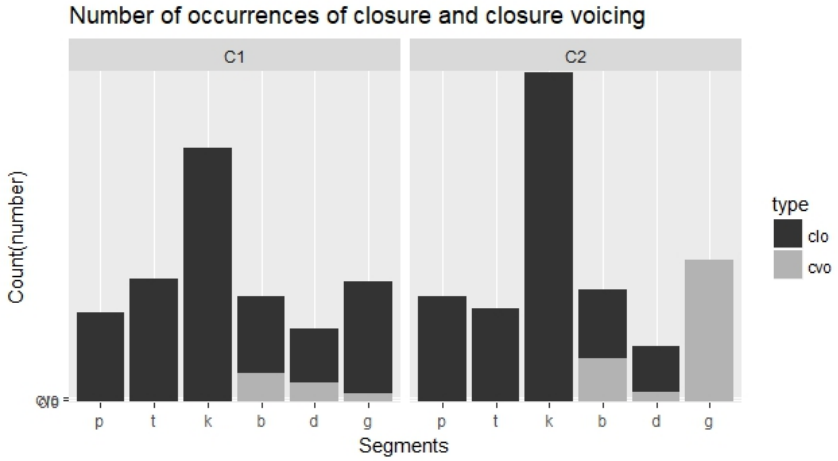

Figure 2: The number of closure (clo) and closure voicing (cvo) in word-initial and word-medial positions.

In general, voiceless plosives are produced without voicing in all segments for both C1 and C2 positions. Although voiced plosives are supposed to be produced with voicing, they were not always produced with voicing. In word-initial positions, voiced plosives were produced without voicing in most of the cases, though they were sometimes produced with voicing. The situation is similar in the word-medial positions except for voiced velar plosive /g/. There is a clear distinction between /k/ and /g/. It means that /k/ was pronounced without pre-voicing and /g/ was always pronounced with pre-voicing.

Moreover, it was also observed that /g/ is pre-nasalized. For other voiced plosives, they were also often not

voiced. These observations are summarized in Table 2.

Acoustics	C1(word-initial position)	C2(word-medial position)
/b/	Often not voiced	Often not voiced
/d/	Often not voiced	Often not voiced
/g/	Often not voiced	Voicing([ᵑg])
/p/	No voicing	No voicing
/t/	No voicing	No voicing
/k/	No voicing	No voicing

Table 2: The summary of all data in Figure 2.

An interesting point is that the voiced plosives are often not voiced in word-medial positions except for /g/. To continue vocal folds vibration, the air pressure of supraglottis have to be lower than that of subglottis. However, since oral occlusion is created when producing voiced plosives, the supraglottal air pressure soon goes up and vocal folds vibration discontinues since the airflow also stops. The results for /b/ and /d/ in Table 2 reflect this phenomenon. Since pre-nasalization is obserced during the production of word-medial /g/, the airflow leaks into nasal cavity to accommodate the aerodynamic restriction in voiced ploives.

According to the table, the existence of voicing does not play an important role in making a distinction between voiced and voiceless plosives. Then, how do Hakodate speakers distinguish voicing contrast? To answer this question, the relationship between closure duration and VOT were analyzed. The graphs in Figure 3 demonstrate the closure duration and the ones in Figure 4 indicate the VOT. The x-axis indicates the segments and y-axis indicates the length of the closure duration.

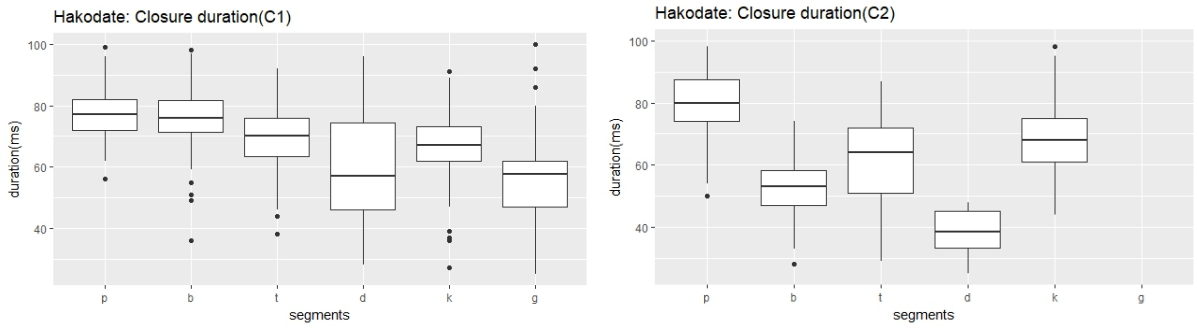

Figure 3: Closure durations for segments in word-initial and word-medial positions (there is no closure duration and VOT for word-medial /g/).

In Figure 3, boxplots show a comparison between voiceless and voiced plosives in terms of closure duration. Each pair of boxplot indicates a voicing contrast (/p-b/, /t-d/, and/k-g/). In word-initial positions, closure durations of each pairs are close. It means that closure durations do not distinguish voiceless and voiced plosives in word-initial positions. In contrast, in word-medial positions, the Inter Quarter Range (IQR)s do not overlap. This implies that the closure duration distinguishes voicing contrast in word-medial positions.

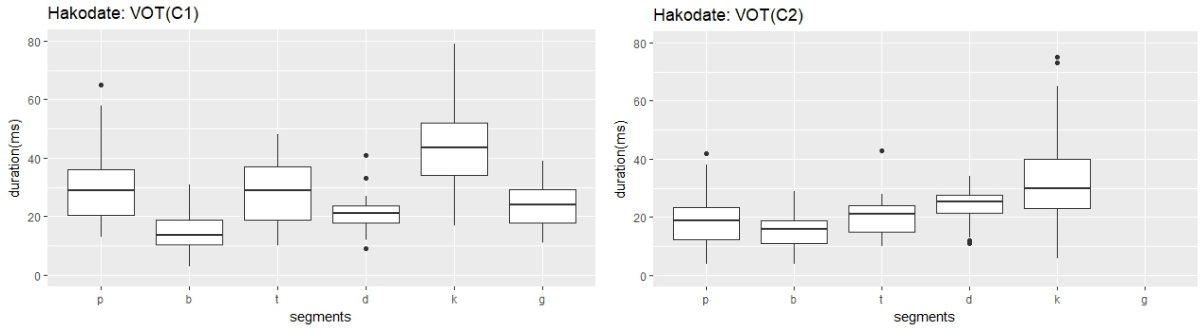

Figure 4: Voice Onset Time (VOT) for word-initial and word-medial positions.

On the other hand, Figure 4 illustrates that VOT might play a role in distinguishing voiceless and voiced plosives in word-initial positions. In word-medial positions, IQRs overlap between each pair of voicing contrast. Boxplots showing the VOT during word-initial consonants do not overlap except for alveolars. Still, however, the average VOTs of alveolars differ greatly in word-initial positions than word-medial positions. Therefore, voiced and voiceless plosives might be distinguished using VOT as an acoustic cue in word-initial positions.

To summarize these results, the acoustic cue for distinguishing the voicing contrast is different depending on the position in a word. Hakodate speakers use VOT for word-initial positions, and closure duration for word-medial positions.

3.2 *Articulatory Results* In the previous section, we have seen that voicing contrast is distinguished with closure duration in word-medial positions and with VOT in word-initial positions. However, closure duration is usually not an acoustic cue for the voicing contrast. In this section, results from Electroglottograph (EGG) with /b/, /d/ and /g/ in word-medial positions from two speakers are examined. In the following figures, the portions surrounded by rectangles indicate the target sound. In Figures 5, 6 and 7, the target sounds are surrounded by low vowels. First, examples of /g/ in the word-medial positions from three speakers are shown in Figure 5.

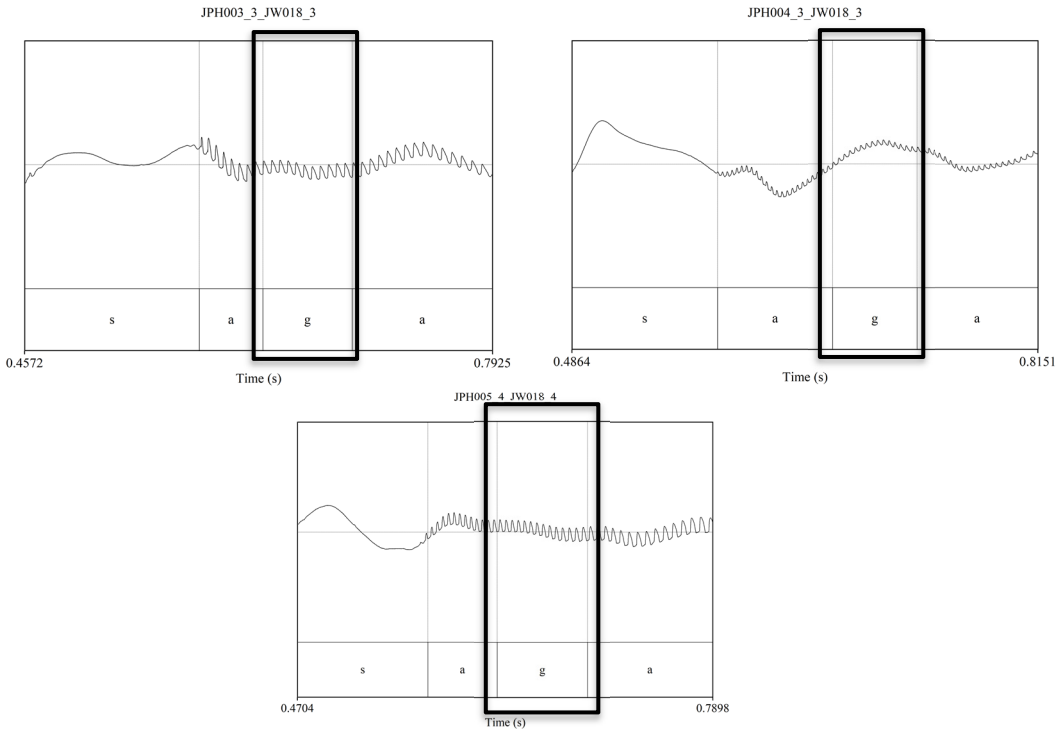

Figure 5: /g/ in the word-medial positions of *saga* 'nature' (left: JPH003, right: JPH004, bottom:JPH005).

Figure 5 illustrates vocal folds vibration during the production of word-medial /g/. Since there is a continuous vibration in producing a voiced plosive, some strategies were used to adopt the vocal folds aerodynamics. Next, EGG results of /b/ in the word-medial positions are demonstrated in Figure 6.

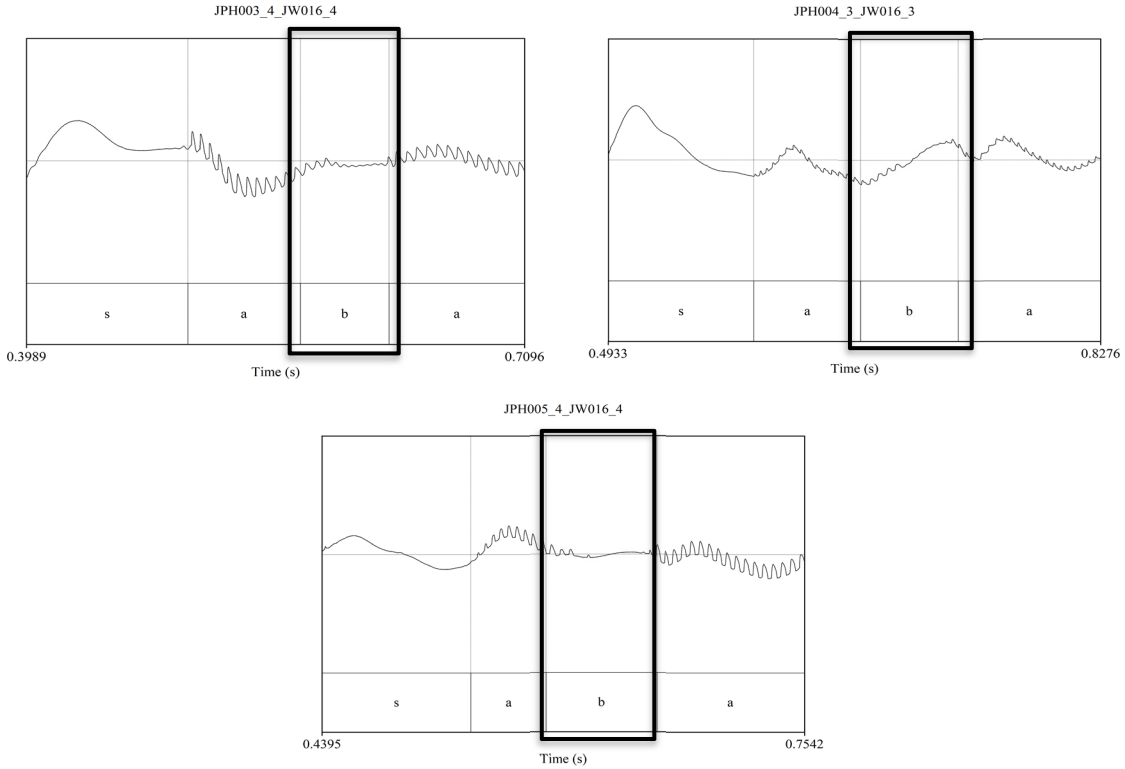

Figure 6: /b/ in the word-medial positions of *saba* 'mackerel' (left: JPH003, right: JPH004, bottom:JPH005).

In Figure 6, speakers demonstrate vocal folds vibration at the beginning of /b/ sound, signifying residual vibration from the preceding vowel (=the final vowel of the frame sentence). However, the vocal folds vibration diminishes due to the aerodynamic restriction. Next, Figure 7 illustrates three examples of /d/ in the word-medial positions.

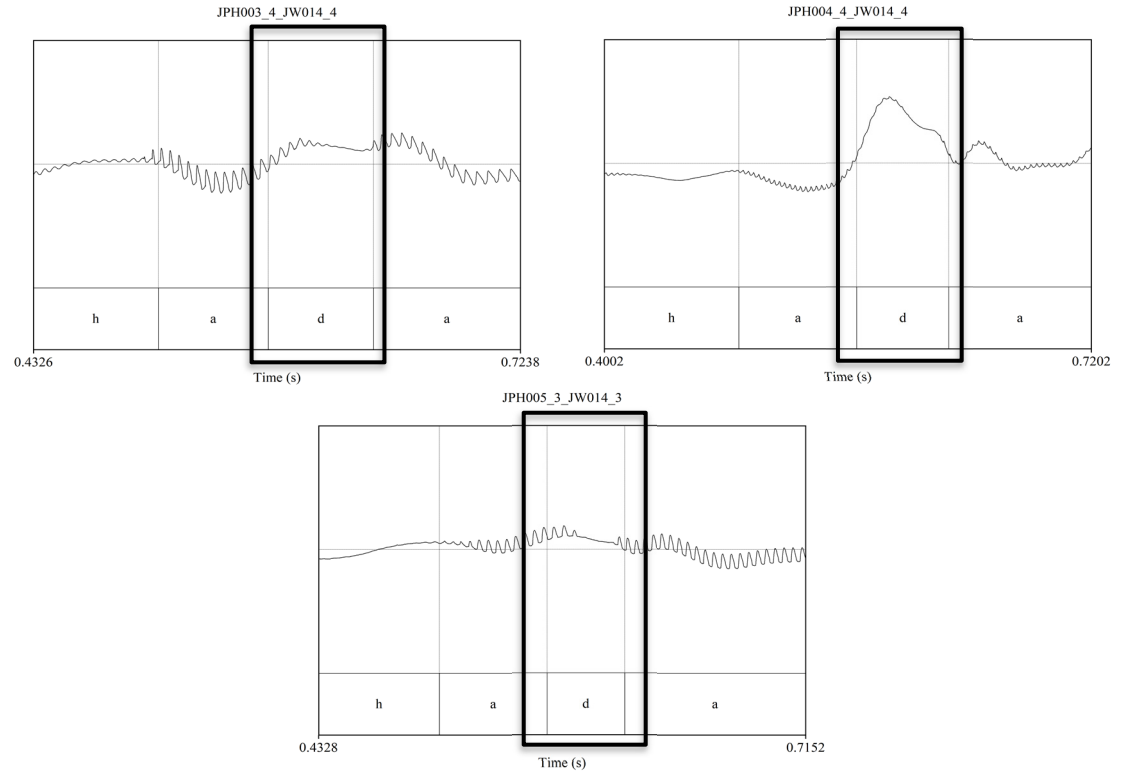

Figure 7: /d/ in the word-medial positions of *hada* 'skin' (left: JPH003, right: JPH004, bottom:JPH005).
Figure 7 indicates similar vocal folds vibration to /b/ sound. The vibrations during the production of /d/ were not maintained in the latter part of the sound, though the vibration of the preceding vowel initially continues. Based

on the observations above, continuous vibrations were discovered for /g/, but not for /b/ and /d/.

4 Discussion

So far, acoustic and articulatory results from three Hakodate speakers have been demonstrated. From acoustic results, closure duration is used for distinguishing the voicing contrast in word-medial positions. In word-initial positions, VOT is an acoustic cue for distinguishing voicing categories. Based on the articulatory results, vocal folds vibration continues for word-medial /g/, while it does not for word-medial /b/ and /d/. In Table 2, it was observed that only /g/ is pre-nasalized while /b/ and /d/ are not pronounced with pre-nasalization in Hakodate dialect. Pre-nasalization is also reported by Inui (1994) in Tohoku dialect, which is spoken in an area not distant from Hakodate. Based on these results, two cases could be considered from both phonetic and phonological perspectives.

First, from Table 2, only /g/ was produced with pre-nasalization while /b/ and /d/ were not allowed to be pre-nasalized. This might be a phonological restriction. From this observation, it could be hypothesized that [nasal] [+dorsal] sequence is allowed but [nasal] [+labial] and [nasal] [+coronal] are not in phonological module of Hakodate speakers.

Second, it might be the case that pre-nasalized /g/ is different from /g/ phonemically, unlike voicing cases of other plosives. This is because /g/ in word-medial positions always have closure voicing as Figure 1 indicates, while /b/, /d/, and /g/ in word-initial positions are often not voiced.

From these points, voicing phenomenon has both phonetic and phonological aspects. Since /b/ and /d/ did not show voicing in acoustic results and vocal folds vibration in articulatory results, they are similar to /p/ and /t/ respectively, and hence it is too simplifying to explain voicing contrast only by vocal folds vibration as mentioned in the previous study (Kawahara et al. 2018). Therefore, detailed analysis is needed from both phonetic and phonological points of view.

Although Ogawa et al. (2017) reported that vocal folds vibration in voiced plosives continue for /b/ , /d/ and /g/ in word-medial positions in Hokkaido dialect, this study observed that the vibration does not continue for /b/ and /d/. As Takada (2011) pointed out, this results might reflect the difference between coastal and inland areas. Participants were born and raised in inland areas in Ogawa et al. (2017), while Hakodate is a coastal area of Hokkaido.

5 Conclusion

This paper has shown how voicing contrast in plosives is realized in Hakodate dialect. Voiced and voiceless plosives cannot be distinguished in word-initial position in terms of acoustic signals. The situation with word-medial positions is similar to word-initial positions. Voiced plosives were produced without pre-voicing except for word-medial /g/, which is produced with pre-nasalization. Since the airflow leaks into the nasal cavity, this is a strategy of lowering supraglottal air pressure and maintaning vocal folds vibration. An acoustic cue for distinguishing voicing contrast depends on the positions of the plosive in a word. VOT makes a distinction in word-initial positions, and closure duration distinguishes voicing contrast in word-medial positions. Articulatory results indicated that vocal folds vibration is maintained for word-medial /g/, while it is discontinued for word-medial /b/ and /d/.

Acknowledgement

Earlier part of the research was funded by Kakenhi (C) 16K02641 to Dr. Seunghun Lee. The field work was supported by JICUF's "STUDENT TRAVEL FUND". I would like to thank Dr. Seunghun Lee, Dr. Tomoyuki Yoshida, the audience at AJL3, and other members of ICU linguistics for comments.

Appendix

The set of stimuli are as follows. All the words are written in roman letters and the meanings are given in the right. The blank in the meaning column means that the word is an onomatopoeia. Labels "initial" and "medial" indicate the positions of the target sounds. HL and LH mean pitch accents.

A. Stimuli in the high vowel context:

target	initial		initial		medial		medial	
	HL		LH		HL		LH	
[b]	buna	‘beech’	bira	‘handbill’	tobi	‘kite’	kabi	‘mold’
[d]→[ʒ]	jiga	‘ego’	jiku	‘axis’	kaji	‘fire’	niji	‘rainbow’
[g]	gumi	‘gummi’	guchi	‘complaint’	hugu	‘globefish’	kagi	‘key’
[p]	puchi	‘petit’	puka		kapi		gepu	
[t] →[ts]	tsuna	‘tuna’	tsuna	‘rope’	matsu	‘pine tree’	hatsu	‘first’
[k]	kuma	‘bear’	kiku	‘chrysanthemum’	toki	‘time’	saku	‘fence’
[m]	miso	‘miso’	mizu	‘water’	kami	‘god’	gomi	‘garbage’
[n]	nushi	‘master’	numa	‘marsh’	wani	‘crocodile’	kuni	‘country’

(Since [di] and [tu] sequences are not allowed in Japanese phonological system, those consonants change into [ʒ] and [ts].)

B. Stimuli in the mid vowel context:

target	initial		initial		medial		medial	
	HL		LH		HL		LH	
[b]	beru	‘bell’	boku	‘I’	hebi	‘snake’	kabe	‘wall’
[d]	doki	‘cray pot’	doku	‘poison’	mado	‘window’	mago	‘grandchild’
[g]	geki	‘play’	gomi	‘garbage’	higo	‘protection’	hige	‘beard’
[p]	pota		pero		kapo		supo	
[t]	teko	‘lever’	toshi	‘age’	koto	‘Japanese harp’	hito	‘human’
[k]	koki	‘extension’	kome	‘rice’	neko	‘cat’	hake	‘brush’
[m]	mozu	‘shrike’	mori	‘forest’	kumo	‘cloud’	himo	‘cord’
[n]	neji	‘screw’	nori	‘seaweed’	tane	‘seed’	mono	‘thing’

C. Stimuli in the low vowel context:

target	initial		initial		medial		medial	
	HL		LH		HL		LH	
[b]	bane	‘spring’	bara	‘rose’	soba	‘soba’	saba	‘mackerel’
[d]	dare	‘who’	dame	‘useless’	hada	‘skin’	muda	‘waste’
[g]	gasu	‘gas’	gaki	‘kid’	shiga	‘Shiga(place)’	saga	‘nature’
[p]	pasu	‘pass’	paneru	‘panel’	supa	‘spa’	toppa	‘breakthrough’
[t]	tako	‘octopus’	take	‘banboo’	kata	‘shoulder’	shita	‘tongue’
[k]	kame	‘turtle’	kane	‘money’	naka	‘inside’	shika	‘deer’
[m]	mada	‘not yet’	maki	‘firewood’	nama	‘raw’	tama	‘ball’
[n]	nasu	‘eggplant’	nami	‘wave’	wana	‘trap’	suna	‘sand’

References

Abramson, Arthur. Seymour, & Lisker, Leigh. (1970). Discriminability along the voicing continuum: Cross-language tests. Proceedings of the 6th International Congress of Phonetic Sciences, Prague (pp. 569-573).

Boersma, Paul & Weenink, David (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved 8 September 2018 from http://www.praat.org/

Gick Bryan, Ian Wilson & Donald Derrick (2013) Articulatory Phonetics. Willey-Blackwell.

Inui, Hideyuki. (1994) 東北方言にみられる語中有声閉鎖音の前鼻音化現象についての聴覚音声学的研究(A perception study of the prenasalization of word-medial voiced plosives in the Tohoku dialect. (https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-06851068/)

Kawahara et al. (2018). Why Study Voicing. Journal of the Phonetic Society of Japan. Vol 22, No. 2, 56-68

Mieko, Takada. (2011) 日本語の語頭閉鎖音の研究−VOTの共時的研究と通時的変化 (A Study of Japanese Word Initial Plosive -Synchronic Study and Diachronic Change of VOT). くろしお出版(Kuroshio Publishers)

Ogawa et al. (2017). An Electroglottograph Study of Voiced Plosives in Two Japanese Dialects. Journal of the Phonetic Society of Japan. Vol. 21. No. 3. 2017. 84.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Shimizu, Katsumasa. (1989), A Cross-Language Study of Voicing Contrasts of Stops, Studia phonologica (1989), 23: 1-12
Shimizu, Katsumasa. (2018). A Study on Phonetic Characteristics of Voicing Contrast of Stop Consonants. Journal of the Phonetic Society of Japan. Vol 22, No. 2, 69-80
Statistics Bureau of Japan (2015). 都市別人口 (Population by city), https://www.stat.go.jp/data/nenkan/65nenkan/02.html