PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

# Active control and passive consequence of vowel devoicing in Japanese: Evidence of highspeed movies and PGG

Masako FUJIMOTO[1]; Seiya FUNATSU[2]; Ichiro FUJIMOTO[3]

[1] Waseda University, Japan

2 Prefectural University of Hiroshima, Japan

3 Takushoku University, Japan

## ABSTRACT

Vowel devoicing is a phenomenon that is found in many languages including Japanese, Korean and Montreal French. The typical case of VD in Japanese is that high vowels /i/ and /u/ are surrounded by voiceless obstruents. However, the frequency of occurrence of VD differs depending on consonant types, accent, dialects, and so on. Previous physiological studies have shown that VD in Japanese involves both active re-organization of the glottal opening pattern during /C̥VC̥/ and passive overlapping of the glottal opening gesture for each voiceless consonant. The former is regarded as phonological and the latter, phonetic. This paper demonstrates the rich variety of the glottal opening pattern during VD using a high-speed digital movie of the laryngeal area and Photoglottography (PGG). This report covers the difference between typical and atypical devoicing environments and that between Tokyo and Osaka dialects, as well as the pattern of VD before geminates and consecutive devoicing.

Keywords: Vowel devoicing, high-speed digital movie, Photoglottography (PGG), before geminates, consecutive devoicing

## 1. INTRODUCTION

Vowel devoicing (VD) is a phenomenon that is found in many languages including Japanese, Korean and Montreal French. The general condition of VD in Japanese is that high vowels /i/ and /u/ are surrounded by voiceless obstruents. However, the frequency of occurrence of VD differs depending on consonant types, accent, dialects, and so on. VD is less frequent 1) when surrounding consonants are both fricatives such as /sis/ and /huh/ compared to /sik/ and /hut/; 2) when target vowels are accented as /ki’t/ compared to /kit/ (/’/ denotes accent nucleus); 3) in Kyoto or Osaka dialects compared to Tokyo (standard) Japanese (1-3).

Previous physiological studies have shown that VD in Japanese involves both active re-organization of the glottal opening pattern during /C̥VC̥/ (4,5) and passive overlapping of the glottal opening gesture for each voiceless consonant (6). The former is often regarded as phonological and the latter, phonetic. This paper demonstrates a rich variety of the glottal opening pattern during VD using a high-speed digital movie of larynx and Photoglottography (PGG). This report covers the difference between stop and fricative consonants and that between Tokyo and Osaka dialects, as well as the pattern of consecutive devoicing, VD before geminate consonants.

In many of the studies of VD, speech waveforms are used to speculate the states of the glottal settings, open or closed, as well as the characteristics of vocal fold vibration. However, seemingly similar acoustic outcomes can often be produced by different mechanisms in terms of active control at the higher level of speech production as well as passive consequences of glottal gesture overlap at the lower level of speech production (6). This paper gives evidence about the correspondence between acoustic signals and actual states of the glottis, with special reference to the vocal fold’s vibration during vowels.

[1] m.fujimoto5@kurenai.waseda.jp
[2] funatsu@pu-hiroshima.ac.jp
[3] ifujimot@ms.takushoku-u.ac.jp

## 2. Method

### 2.1 Speakers

Four speakers of Japanese (three males, one female) between 32 to 53 (av. 39) years old served as subjects. In the present paper, they are referred to TM, TF, OM and OF, where M and F stand for male and female, with T and O, for Tokyo (Standard Japanese) and Osaka dialect, respectively.

### 2.2 Test Words

Test words were selected to fit the necessity of several studies. The test word sets and break down lists for recordings differ among subjects. Ten words were selected for the analysis of the present study, as shown in Table 1. Two tokens were recorded for each word. All words are non-words: either three-mora without geminates or four-mora with geminates. Vowels /i/ and /e/ were used in this study, since the epiglottis does not tilt backwards for their production, which disturbs the glottal view.

Table 1 – Test words in the present study. 'St,' 'Fr,' 'C̥,' 'QC̥,' and 'Vnh' stand for stops, fricatives, voiceless consonants, voiceless geminate consonants and non-high vowels, respectively.

| devoicing condition | consonantal condition | words |
|---|---|---|
| Typical | St-St, Af-Af, Af-St | /ekiki/ [ekiki], /etiti/ [etɕitɕi], /etite/ [etɕite] |
| Atypical | Fr-Fr | /esisi/ [eɕiɕi], /esise/ [eɕise] |
| Before geminate | C-QC | /ekkiki/ [ekkiki], /ettiti/ [ettɕitɕi], /essisi/ [eɕɕiɕi] |
| Consecutive devoicing | St-St-St | /tititi/ [tɕitɕitɕi] |
| Non-high vowel | C-$V_{nh}$-C | /ekeke/ [ekeke] |

### 2.3 Data Recordings

Simultaneously with the speech sound and Photoglottogram (PGG), laryngeal images were digitally recorded using a fiberscope though a nostril at the rate of 4500 frames/sec. Speakers read five to ten words in a list in isolation during a recording session at their comfortable speech tempo and (pitch) accent pattern. Duration of each session was 6 seconds for TM, and 14 seconds for TF, OF and OM, depending on the recording equipment used for the experiments. Figure 1 schematically shows the setting of the recordings.

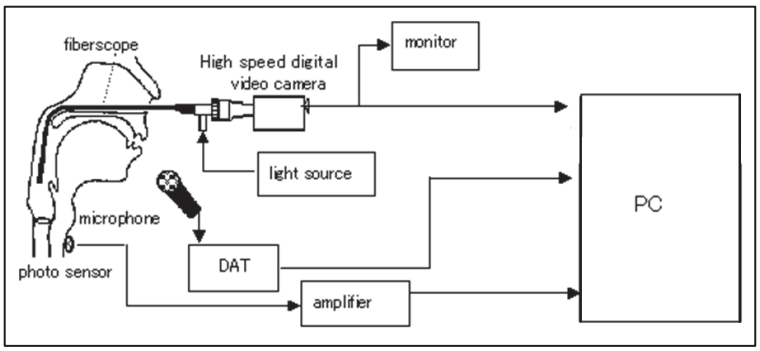


Figure 1: Block diagram of the experimental setting.

### 2.4 Data Analyses

Glottal images were analyzed using Adobe Photoshop 10. Audio and PGG signals were analyzed using the software SONY PCscan II and Praat (ver 6.0.29). Vowels were judged as devoiced when the audio signal did not show one or more waves which typically correspond to vocal fold vibration. Figure 2 shows examples of glottal images on the left, and audio and PGG signals on the right, of /ekiki/ uttered by TM.

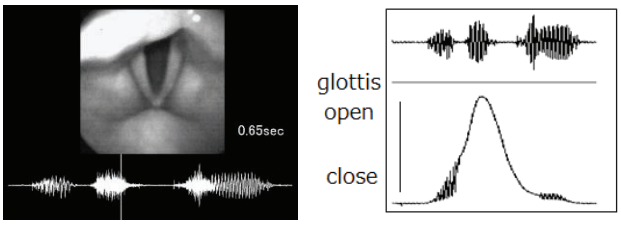


Figure 2 – Left panel shows the glottal image during /etiti / by TM. The vertical line on the acoustic signal denotes the timing of glottal images. The right panel shows the acoustic and PGG signal of the same token.

## 3. RESULTS

### 3.1 Devoicing Rate

Table 2 shows the devoicing rate of the target vowel /i/ in the test words. The result showed that 1) vowel devoicing occurred more frequently for the Tokyo speakers than for the Osaka speakers; 2) within the Tokyo speakers, devoicing was consistent in the 'Typical' environment but not in the 'Atypical' and or 'Before geminates' environments; 3) one token by TM showed 'Consecutive' devoicing; 4) non-high vowels did not devoice for any of the speakers. Three of the speakers read the words with the unaccented LHHH pattern; OF read four mora words with an unaccented pattern and three mora words with an accented HLL pattern.

Table 2 – Devoicing rate according to test words and speakers. Numerator of fraction denotes the number of devoicings out of the total number, indicated in the denominator.

| devoicing condition | test words | TM | TF | OM | OF |
|---|---|---|---|---|---|
| Typical | /ek*i*ki/ | 2/2 | 2/2 | 0/2 | 0/2 |
| | /et*i*ti/ | 2/2 | 2/2 | 0/2 | 0/2 |
| | /et*i*te/ | 2/2 | 2/2 | 2/2 | 2/2 |
| Atypical | /esisi/ | 1/2 | 0/2 | 0/2 | 0/2 |
| | /esise/ | 2/2 | 2/2 | 0/2 | 0/2 |
| Before geminates | /ekk*i*ki/ | 2/2 | 1/2 | 0/2 | 0/2 |
| | /ett*i*ti/ | 2/2 | 0/2 | 0/2 | 0/2 |
| | /ess*i*si/ | 0/2 | 0/2 | 0/2 | 0/2 |
| Consecutive devoicing | /t*i*t*i*ti/ | 1/2 | 0/2 | 0/2 | 0/2 |
| Non-high vowel | /ekeke/ | 0/2 | 0/2 | 0/2 | 0/2 |

### 3.2 Glottal Images

#### 3.2.1 Typical Devoicing Environments

Figure 3 shows a series of the glottal images with the audio signal of /ekiki/ by TF. The degree of glottal opening increases during the period of devoiced CVC, reaching a maximum opening after the release of /k/ near the devoiced vowel. It then decreases towards /k/ in the third mora. The degree of glottal opening during /k/ in the second mora is much larger than that in the third mora. This suggests that the glottal opening during the devoiced vowel is positively controlled, probably not due to the overlap of two glottal openings for each consonant.

Figure 4 shows a series of glottal images with the audio signal of /etiti/ by TM. Again, the degree of glottal opening increases during the period of devoiced CVC, reaching a maximum opening after the release of /t/ near the devoiced vowel and then decreasing towards the /t/ in the second mora. The degree of glottal opening during /t/ in the first mora is much larger than that following the devoiced CVC. This suggests that the glottal opening during the devoiced vowel is positively controlled; it is not due to the overlap of two glottal openings for each consonant.

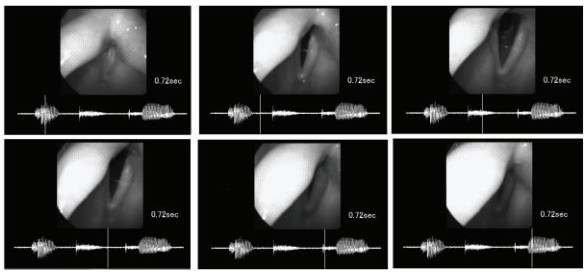

Figure 3 – Glottal images during /ekiki/ with devoiced /i/ by TF. The vertical line on the acoustic wave denotes the timing of each glottal image.

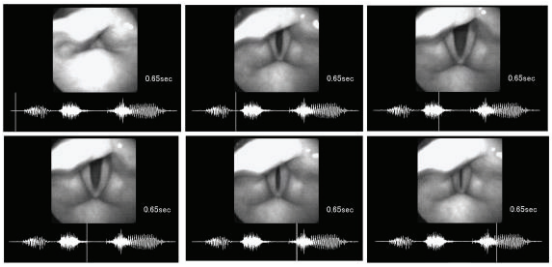

Figure 4 – Glottal images during /etiti/ with devoiced /i/ by TM.

Figure 5 shows the audio and PGG signal of /ekiki/ by TF and /etiti/ by TM corresponding to Fig 3 and 4, respectively. The pattern of the glottal opening is a smooth monomodal shape during the voiceless CVC with its peak at around the voiceless vowel for both tokens. This confirms the analysis of the glottal images mentioned above.

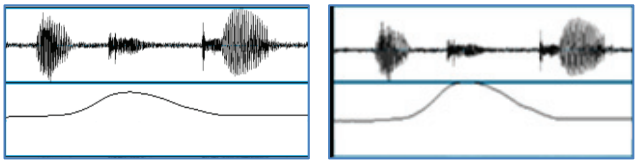

Figure 5 – Audio and PGG signals of /ekiki/ by TF (left) and /etiti/ buy TM (right). The duration of each panel is 700 ms.

For Osaka speakers, VD occurred only in /etite/ but not in the other words. For /etite/, two tokens were devoiced for both speakers. Figure 6 shows the glottal images with the audio signal of /etite/ by OM. The glottal opening is large during the devoiced /tit/, which is similar to the Tokyo speakers, but the maximum opening comes earlier at the release of /t/. Figure 7 shows the glottal images with the audio signal of /etiti/ by the same speaker. For this test word, /i/ in was not devoiced in both tokens. The degree of glottal openings during the /t/s in /etiti/ are much smaller than those of the devoiced /tit/ in /etite/ in Fig 6.

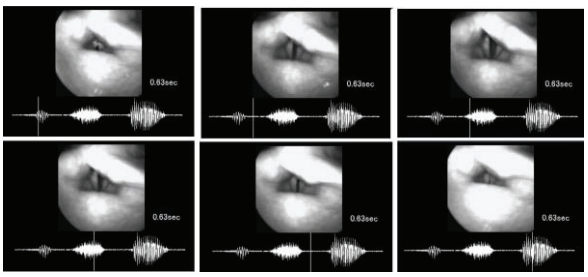

Figure 6 – Glottal images during /etite/ with devoiced /i/ by OM.

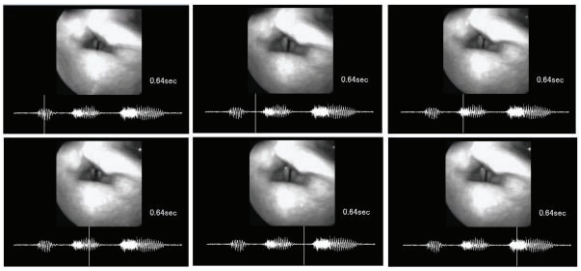

Figure 7 – Glottal images during /etiti/ by OM. /i/ is not devoiced.

Figure 8 compares the audio and PGG signals of /etite/ and /etiti/ by OM. The PGG signal during /tit/ is clearly different depending on whether the vowel is voiced or devoiced. The degree of glottal opening of /tit/ in /etite/ is large while that of the two /t/s in /etiti/ are small. The pattern of voiceless CVC differs from that by Tokyo speakers. Recall that the maximum opening of voiceless CVC by Tokyo speakers is around the voiceless vowel. In contrast, the maximum opening of /etite/ occurs at the beginning of the frication noise of the first /t/. It then decreases to its lowest degree at the end of the devoiced /t/ and keeps the low level of opening during the closure of the second /t/. This pattern can be regarded as a merged shape of one big opening for /ti/ and a small opening for /t/.

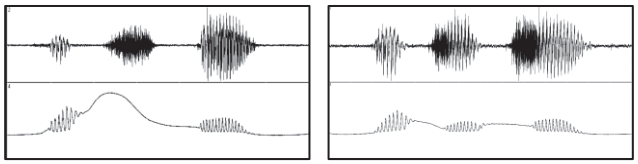

Figure 8 – PGG and audio signal of /etite/ and /etiti/ by OM. The duration of the panel is 700 ms.

Figure 9 shows the glottal images of two tokens of /etite/ and /etiti/ by OF. Although /i/ in /etite/ was devoiced in both tokens, the degree of glottal opening of /tit/ significantly differs between the two, one small (leftmost panel) and one large (second panel). The smaller opening is similar to the two /t/s in /etiti/ shown in the right two panels. This suggests that the glottal opening is re-organized in the second token to produce a voiceless CVC with a large opening, but it is as small as a single /t/ in the first token. Figure 10 shows the audio and PGG signals of the same three tokens of Fig 9. The PGG pattern corresponds to this analysis.

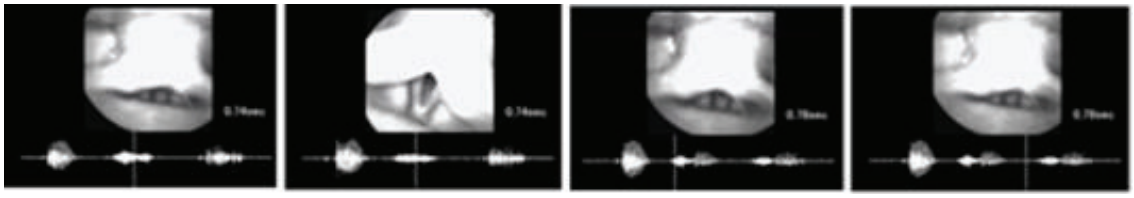

Figure 9 – Glottal images of /etite/ and /etiti/ by OM. Left two panels are images near the timing of the devoiced /i/ from two different /etite/ tokens. The right two are at the first and second /t/s in a /etiti/ with voiced /i/.

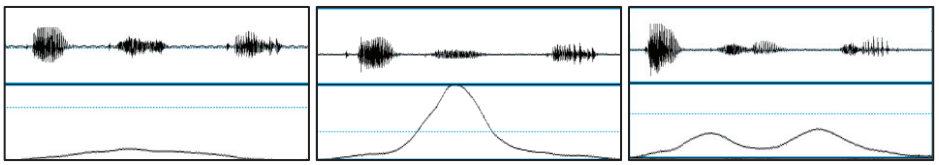

Figure 10 – PGG and audio signal of /etite/ and /etiti/ by OF. Left two panels are /etite/ with devoiced /i/ and right panel, /etiti/ with voiced /i/. The duration of each panel is 700 ms.

### 3.2.2 Atypical Devoicing Environments

For the words /esisi/ and /esise/, both with atypical devoicing environments, VD occurred for two tokens in /esise/ and one token in /esisi/ by TM, two tokens in /esise/ by TF, but none for OM and OF. The results agree with the previous finding that devoicing in atypical devoicing environments is non-systematic and less frequent. Figure 11 shows the glottal images during /esise/ by TF. The glottis opens widely during the devoiced CVC and maximally in the middle at around the devoiced vowel. Figure 12 shows the audio and PGG signals of the same token. The PGG pattern differs from that in a typical consonantal environment shown in Fig 5. While the PGG signal of /ekiki/ in Fig 5 shows a simple mono-modal pattern, that of /esise/ in Fig 12 resembles a plateau-like shape. This pattern can be regarded as a merged shape of two big openings for the first and second /s/s, which results in the devoicing of /i/ between them.

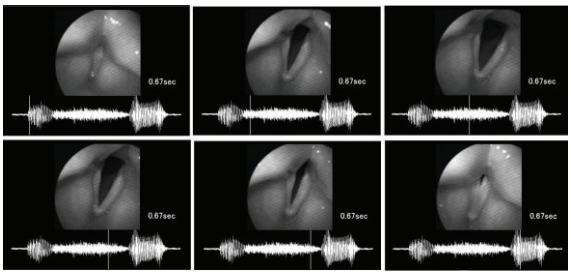

Figure 11 – Glottal images during /esisi/ by TF. The vertical line on the acoustic signal denotes the timing of glottal images.

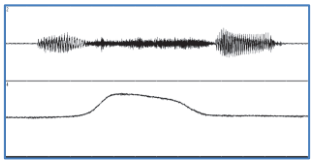

Figure 12 – PGG and audio signal of /esise/ by TF. The duration of the panel is 700 ms.

### 3.2.3 Before geminates

Figure 13 shows glottal images during /ekkiki/ by TM and figure 14, audio and PGG signals of the same token. The glottis opens widely during the devoiced CVCC and maximally at the latter half near the frication of the devoiced vowel. Thus, the glottal opening pattern of devoiced CVCC differs from that of voiceless CVC in terms of the later timing of the peak opening and its right-skewed shape.

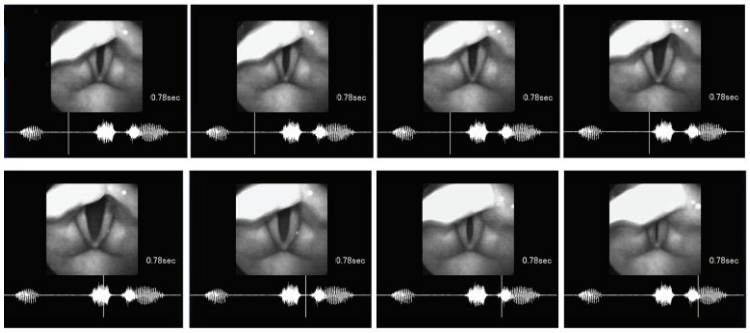

Figure 13 – Glottal images during /ettiti/ by TM. The vertical line on the acoustic wave denotes the timing of glottal images.

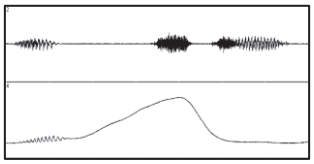

Figure 14 – PGG and audio signal of /ettiti/ by TM. The duration of the panel is 700 ms.

### 3.2.4 Consecutive Devoicing

As for the test words /kikiki/ and /tititi/, consecutive devoicing occurred in only one token /tititi/ by TM. Figure 15 shows the glottal images of the word where the first and second /i/s were devoiced and figure 16, audio and PGG signals of the same token. The glottis opens throughout the voiceless CVCVC and most widely during the first voiceless CV (the third panel in Fig. 15). In this respect, in terms of the timing of the peak opening, the glottal opening pattern of a devoiced CVCVC differs from that of a voiceless CVC in a typical devoicing environment. Namely, while peak opening comes in the middle around the devoiced vowel in CVC, it occurs in the first half in a voiceless CVCVC. Due to this remarkably large opening, the light passed though the glottis exceeded the limit of the experimental setting and the PGG signal was saturated at the first half of the devoiced CVCVC. Nonetheless, it is clear that the PGG signal skewed left. It is also noticeable that the shape of the latter half of the PGG signal shows a bi-modal pattern, i.e. a small plateau follows after a rapid decrease at the end of the CVCV.

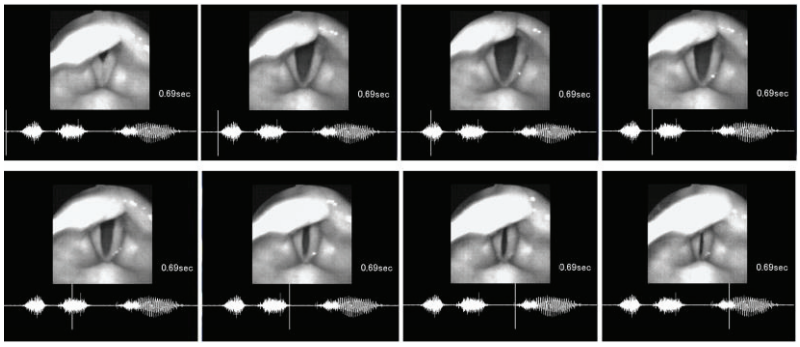


Figure 15 – Glottal images during /tititi/ by TM where first and second /i/s were devoiced. The vertical line on the acoustic wave denotes the timing of glottal images.

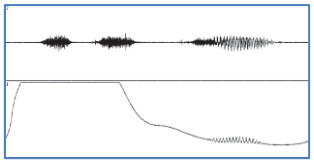

Figure 16 – PGG and audio signal of /tititi/ by TM. The duration of the panel is 700 ms.

### 3.2.5 Non-high Vowel

The non-high vowel /e/ did not devoice at all regardless of the test word or the speaker. Figure 16 shows the glottal view during /ekeke/ by OM and Figure 17, audio and PGG signals of /ekeke/ by OF. As is clear from the images, glottal opening of /k/ is as small as or even smaller than that of the opening phase of a cycle of the vowel /e/, which is clearly indicated in the PGG signal. This means that the glottal opening of the word medial /k/ is small.

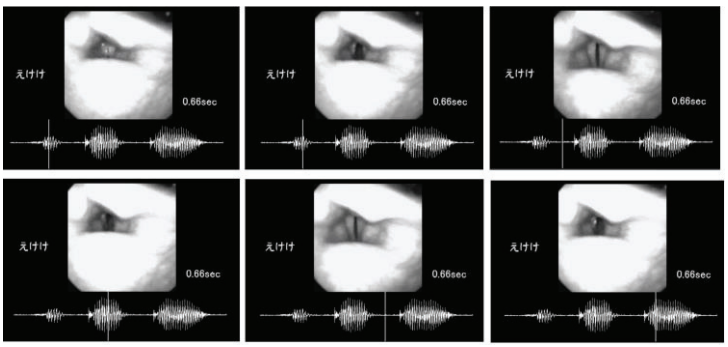


Figure 16 – Glottal images during /ekeke/ by TM vowels /e/s were not devoiced. Vertical line on the acoustic wave denotes the timing of glottal images.

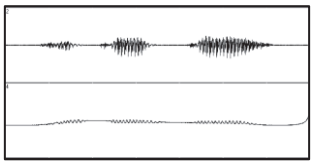

Figure 17 – PGG and audio signal of /ekeke/ by TM. The duration of the panel is 700 ms.

## 4. DISCUSSION

The present results confirm the dialectal differences as well as the differences in consonantal environments with regard to VD rate. As expected, VD was less frequent in the Osaka speakers than the Tokyo speakers. Also, VD rate is more frequent in typical consonantal environments. Even though VD is very infrequent for the Osaka speakers, it occurred only in typical consonantal environments. For the Tokyo speakers, VD occurred in typical devoicing environments regardless of test words, but not systematically in atypical environments before geminates and consecutive conditions. VD in non-high vowels did not occur at all.

In typical devoicing environments, glottal opening during devoiced CVC by the Tokyo speakers was mono-modal and large without exception. Peak glottal opening during devoiced CVCC was also mono-modal and large, although the peak came later than CVC. These results are consistent with previous studies. Such VDs are regarded to be controlled electromyographically (4). In contrast, glottal opening during a devoiced CVC in atypical environments showed a plateau pattern, and during devoiced CVCVCV, a bimodal one. These devoicings by Tokyo speakers may be passively derived.

For an Osaka speaker, notably small glottal opening corresponded to devoiced CVC in one of the tokens. This agrees with the previous result using another Osaka speaker (5). Such VD may not be controlled electromyographically and, thus, occurred passively.

Although the data are limited, the present results demonstrate that the occurrence of VD in Japanese is not a simple phonological event with binary choice but embraces a variety of gradual states which depend on phonetic realization in each token.

## 5. FINAL REMARKS

By using high-speed movies and PGG, this study shows the rich varieties of glottal opening patterns during different segmental conditions. The results shed light on some of the articulatory mechanisms and uncover differences in devoicing patterns, found from time to time in Japanese. Further study is necessary to clarify the whole mechanism of vowel devoicing.


## ACKNOWLEDGEMENTS

This study was partly supported by JSPS Kakenhi 17K02707 and 15K02524.


## REFERENCES


1. Sakuma K. Nihon Onseigaku [Japanese Phonetics],” Tokyo: Kyobunsha, 1930.
2. Kawakami S. Nihongo Onsei Gaisetu [Introduction of Japanese Phonetics] Tokyo:Ohfu, 1994.
3. Fujimoto M. Vowel devoicing, In Handbook of Japanese Phonetics and Phonology, Berlin: De Gruyter. p. 167-214, 2014.
4. Yoshioka, H. Löfqvist, A. and Hirose, H., Laryngeal adjustments in Japanese voiceless sound production, J. of Phonetics 10, p. 1-10, 1982.
5. Fujimoto M., Murano E., Niimi S. and Kiritani.S. “Difference in glottal opening pattern between Tokyo and Osaka dialect speakers: Factors contributing to vowel devoicing,” Folia Phoniatrica et Logopaedica, 54-3, p. 133-143, 2002.
6. Fujimoto M., Glottal opening pattern in devoiced tokens by an Osaka dialect speaker, J. of Phonetic Society of Japan, 9(1), p. 50-59, 2005.