

## Laryngeal Vibratory Behavior in Traditional Noh Singing\*

YOSHINAGA Ikuyo, KONG Jiangping

### Abstract

The phonation of Noh, a traditional Japanese style of singing, was investigated using electroglottographic and acoustical analyses. The dynamics of the laryngeal vibratory behaviors were analyzed for the singing voice of the Noh play compared with natural speech based on the electroglottography (EGG) parameters, EGG waveform, spectrum and spectrogram. The result shows that Noh singing is characterized by low  $OQ_{egg}$  and high  $SQ_{egg}$ . Three types of phonations are used in the singing with pressed, vocal-ventricular mode (VVM), and growl voices. It was hypothesized that the period doubling observed in the EGG signal was reflective of VVM, which was caused by the phase difference in the vocal and ventricular fold oscillations, while the damped peak amplitude in every other cycle in the EGG signal was the result of the oscillations of the aryepiglottic folds at a frequency of half of the fundamental frequency. Subharmonics generated by the supraglottal oscillations add unique timbre to the sounds. The results suggest that the combination of phonation types is the key factor in generating their peculiar voice qualities.

### Key words:

pressed voice; vocal-ventricular mode (VVM); growl voice; electroglottography (EGG); open quotient ( $OQ_{egg}$ ); speed quotient ( $SQ_{egg}$ )

### Introduction

Noh is a traditional performing art which has been handed down orally since the Nara Era. A typical Noh play involves a small chorus and orchestra, a shite (main role), and a waki (supporting role) wearing masks. The Noh artistic style has a solemn atmosphere. The Shite have 5 schools at present, among which the Hosho and Kanze school are most representative. The Hosho has an excellent reputation for its solid performance style and long history. From the oral arts viewpoint, the singing voice in Noh is influenced by Buddhist chants.

Research on Noh singing is quite rare. The singer's formant has been observed to be 3-4 kHz<sup>[1]</sup>. The duration of the

consonants is lengthened in Noh singing compared with classical European singing. The Mahalanobis generalized distance on the LPC cepstrum was used to evaluate the distances between conversational and singing voices in Noh<sup>[2]</sup>. They showed that the distance between them is much closer than in classical European singing. Thus, expert singers use their characteristic phonation of singing even in their conversational speech.

Recent investigation using high-speed camera, X-ray, Kymography, and other methods has revealed ventricular and aryepiglottic fold oscillations in Asian vocal cultures and in some ethnic and pop music. The ventricular folds oscillate at speeds of the fundamental frequency ( $F_0$ ),  $F_0/2$  or  $F_0/3$  in the vocal-ventricular mode (VVM), with the aryepiglottic folds oscillatory at  $F_0/2$  in the growl voice (also called the voiced aryepiglottic trill<sup>[3]</sup>)<sup>[4-6]</sup>. These kinds of phonations are found not only in the singing technique but also in the vocal fry, voice instabilities, and infant vocalizations. These irregular vocalizations are often interpreted as period-doubling bifurcations, with the corresponding acoustical signals often showing sudden jumps to subharmonic regimes<sup>[7,8]</sup>.

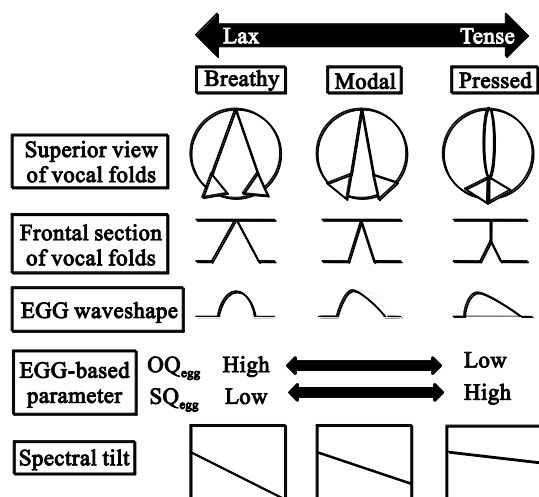
Electroglottography (EGG) has been widely used to analyze vocal fold functions since Fabre reported on its use in 1957<sup>[9]</sup>. Five Chinese phonation types are described using EGG parameters<sup>[10]</sup>. Table 1 shows that the open quotient ( $OQ_{egg}$ ) and speed quotient ( $SQ_{egg}$ ) are key factors distinguishing the five phonation types. For example, the vocal fry is described by high  $OQ_{egg}$  and  $SQ_{egg}$ . This suggests that they are also very important parameters for voice quality assessments.

Figure 1 gives a simplified illustration which shows the relations between the voice qualities and physiological, EGG and spectral properties. The relationship between the EGG waveform and the frontal section of vocal folds is described according to Ref. [11]. The relationship between the superior view of vocal folds and the spectral tilt is determined with reference to Ref. [12]. Generally, low  $OQ_{egg}$  represents a pressed voice with more contact area in the vocal folds that can be seen at the frontal section of the vocal folds in Fig. 1. A high  $OQ_{egg}$  represents a breathy voice with less contact area in the vocal

folds, because a longer glottal release duration will result in more airflow. Low  $SQ_{egg}$  indicates the glottal closing is slower and the voice has less energy without the forceful glottal closure. This can be described as a steep spectrum tilt from an acoustical point of view. A high  $SQ_{egg}$  has a forceful, quick glottal closure and is described by a small spectrum tilt. Thus, these EGG parameters can be used to judge voice qualities.

**Table 1 Distinctive features of the source parameters for the five Chinese phonation types compared with the modal. “-” indicates lower and “+” indicates higher than modal. A high pitched voice is indicated as “High” in this table.**

	$F_0$	$OQ_{egg}$	$SQ_{egg}$
Fry	-	+	+
Breathy	-	+	-
Pressed	-	-	+
Modal	±	±	±
High	+	-	-



**Fig. 1 Simplified illustration of the vocal folds, EGG waveforms, and parameters and the spectral tilt related with the phonation types**

This study used electroglottographic and acoustic analyses to study the laryngeal vibratory behavior and to describe the phonatory characteristics of voice qualities in Noh singing.

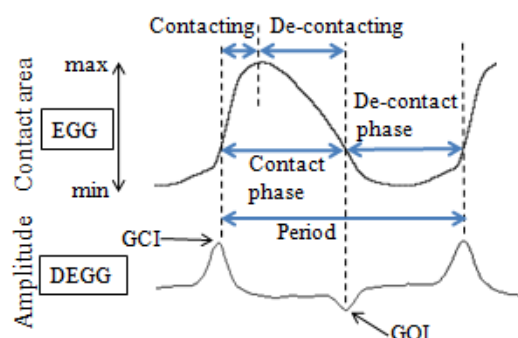
## 1. Methods and Materials

This section describes the calculational method to get the EGG parameters, the details of the voice materials, and the processing procedure.

### 1.1 parameter calculation method

Voice quality is key to judging various singing styles.

Perceptual assessment and a variety of instrumental (acoustical and physiological) methods are used in the definition of voice qualities<sup>[13]</sup>. EGG, which measures the electrical conductance changes between a pair of electrodes placed on the neck, is a noninvasive technique used to observe vocal fold vibratory patterns. The EGG signals provide meaningful information only when the vocal folds repeat contact and de-contact during vibration. Therefore, contact-based analysis is the common algorithm. Because the EGG and airflow waveforms differ from each other qualitatively,  $OQ_{egg}$  and  $SQ_{egg}$  are employed in this study as the EGG-based parameters. A rise in the EGG signal corresponds to the closing of the glottis, while dropping corresponds to opening of the glottis as shown in Fig. 2.



**Fig. 2 EGG waveform and phases of vocal fold contact**

Three EGG-based parameters are extracted:  $F_0$ ,  $OQ_{egg}$ , and  $SQ_{egg}$ . The definitions of  $F_0$  and  $OQ_{egg}$  are described as follows:  $F_0=1/\text{period}$  and  $OQ_{egg}\%=\text{de-contact phase}/\text{period}\times 100\%$ . Although the  $SQ_{egg}$  can be varied in detail across researchers, the definition used in this research is  $SQ_{egg}\%=\text{de-contacting}/\text{contacting}\times 100\%$ <sup>[10]</sup>.

There has been much discussion on the definition of the glottal closing instance (GCI) and glottal opening instance (GOI)<sup>[14-18]</sup>. Three kinds of proposed EGG calculational methods are the criterion-level algorithm<sup>[19]</sup>, the differential of the EGG signal (DEGG)<sup>[15,20-24]</sup>, and a combination of the criterion-level and the DEGG method, called the Hybrid algorithm<sup>[17,18]</sup>. The DEGG is considered the best method reflecting the GCI and GOI, but it is not reliable with imprecise or double GCIs and GOIs<sup>[15]</sup>. EGG waveforms of Noh singers often contain sub-cycles (Fig. 3).

In this case, the DEGG signals show double GCIs or GOIs, and the precise setting of the criterion level is necessary so that each instance can be detected. Therefore, the EGG cycle segmentations were set at 35% to define the GCIs or GOIs as shown in Fig. 4.

## 1.2 Voice material

The Noh singing voice was studied using 2 males and 2 females who were professional Noh singers of the Hosho school. The two males (subjects A and B) were successors of Yoshio Hosho, a living national treasure. They started singing Noh at the age of 4. The females (subjects C and D) also started singing at an early age. They were all in their twenties with about 20 years of singing experience when the recordings were made.

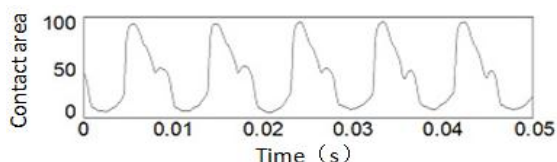


Fig. 3 5-cycle EGG waveform of vowel /e/ at G3 (196 Hz) from Tsurukame (Subject A)

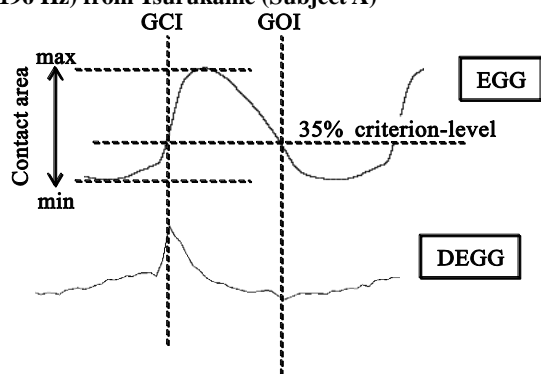


Fig. 4 Definitions of GCI and GOI on the EGG signal by the 35% criterion and a differential method

The voice materials consist of (1) the Noh play “Tsurukame (crane and turtle)” and (2) the sustained vowels /a, e, i, o, u/. Tsurukame is a story describing the New Year’s prayer at the palace of Emperor Xuanzong of the Tang Dynasty. The subjects read and sang the lyrics of Tsurukame in their comfortable pitch. In this case, gaps of  $F_0$  and intensity changes occurred between the singing and speaking. Since  $F_0$  and intensity influence  $OQ_{egg}$  and  $SQ_{egg}$ , this method can’t be used to compare  $OQ_{egg}$  and  $SQ_{egg}$  between singing and speaking with the first set of materials. The subjects spoke 5 sustained vowels /a, e, i, o, u/ using speaking and Noh singing styles. They sustained every semitone from their lowest pitch to the highest and they were requested to use the same loudness in singing and speaking. Thus, the  $F_0$  gap between singing and speaking and the intensity differences were eliminated in case 2. In this case, the recordings covered a range of over an octave, starting at about F2# (92 Hz) for the males and E3 (164 Hz) for the female subjects. Since the recordings for case 2 covered their entire pitch range, it contained the pitch height which is not used in actual Noh

singing. Moreover, the sustained phonation did not contain some of the Noh singing techniques which were used in the actual singing. However, comparison of the singing and speaking using the case 2 material will illustrate the inherent voice qualities of Noh singing phonation. The phonation types of actual Noh singing were studied in case 1.

The recordings took place at the Nohgaku stage in Tokyo, Japan. The EGG signal was obtained on an EGG by LARYNGOGRAPH Ltd BF; Kay, UK. The audio signal was recorded by a Sony Electret Condenser Microphone. These signals were simultaneously recorded and digitized on 16 bits at a sampling frequency of 44.1 kHz.

## 1.3 Data processing

The recorded files were prepared for the acoustical analysis by downsampling to 11.025 Hz. Then, the EGG rumble caused by the up and down laryngeal movements was filtered out by a high-pass cutoff frequency of 60 Hz, because it could confuse the calculated results. The files were divided into smaller pieces to prepare for the batch processing to obtain the EGG parameters. Since a large amount of data processing was needed and the recorded files were different lengths, 30 parameter values were extracted from each file. Wavelet transforms were applied to each file to adjust the noise and warp of the EGG signals, which might also cause miscalculations, before extracting the EGG parameters. The data processing was performed using the Matlab based SpeechLab developed by the Linguistic Lab of Peking University.

Since the singing data files were 11 min long which was twice the length of the speaking data, the parameter were extracted using 1140 data points from each singing and speaking file. For case 2, 30 parameter values were extracted for each sustained vowel.

## 2 Parameter Analysis

The EGG parameters were then analyzed by comparing the singing and speaking parameters to study the Noh singing voice quality.

### 2.1 Pitch range analysis for Tsurukame

The  $F_0$  distributions for the singing and speaking of Tsurukame are shown in Figs. 5 and 6. The discussion

is limited to  $F_0$  to focus on the pitch range. The  $F_0$  distributions for male subjects A and B are shown in Fig. 5, while the  $F_0$  distributions for female subjects C and D are shown in Fig. 6. The singing parameters are shown by 1140 black circles, while the speaking parameters are shown by 1140 gray circles in Figs. 5 and 6. Table 2 lists the average, minimum, and maximum of  $F_0$  for each subject. Semitone as well as hertz was used to measure  $F_0$ . Figures 5 and 6 show a large difference in  $F_0$  between the singing and speaking.  $F_0$  for the singing is significantly higher and its range is wider than for speaking with all the subjects. The  $F_0$  range in the males is 11-12 semitones for singing, and 6 semitones for speaking. Thus, the singing range is almost double the speaking range. The  $F_0$  range for the females is 11-12 semitones for singing and 4-5 semitones for speaking. Thus, the

singing range is more than double the speaking range. The  $F_0$  range for singing is then wider than that for speaking regardless of gender. The average  $F_0$  of the males is about 182-188 Hz for singing and 103-108 Hz for speaking. Thus, the  $F_0$  for the singing is 9-10 semitones higher than that of speaking. The average  $F_0$  for the females is 288 Hz for singing and 193-198 Hz for speaking. Thus, the  $F_0$  of singing is 7 semitones higher than that of speaking for the females.

Therefore, the  $F_0$  range for singing is 11-12 semitones, almost double that of speaking. The average  $F_0$  for singing is 9-10 semitones higher in males and 7 semitones higher in females than that of speaking. Though singing voice in Noh sounds quite low, the  $F_0$  of singing actually rather high than that of speaking.

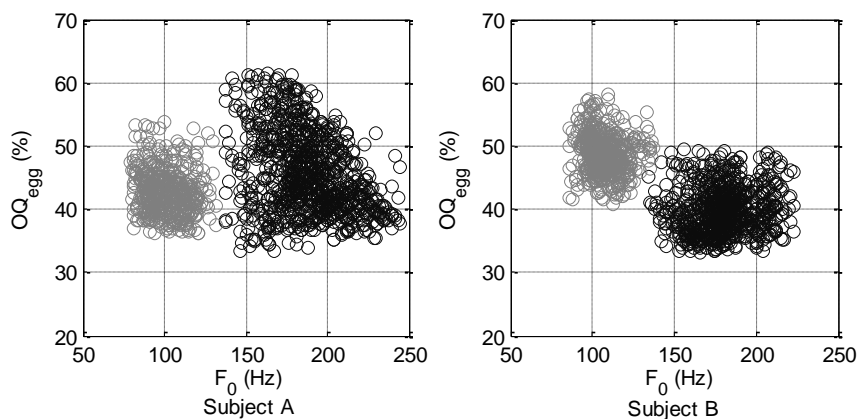


Fig. 5 Distribution of  $F_0$  for singing (black) and speaking (gray) of Tsurukame by male.

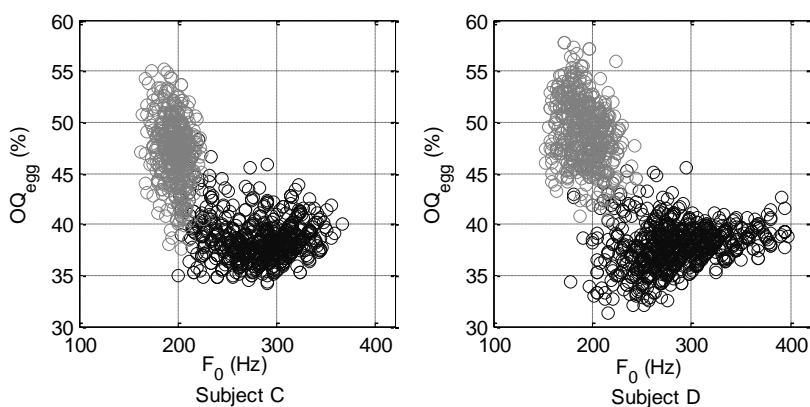


Fig. 6 Distribution of  $F_0$  for singing (black) and speaking (gray) of Tsurukame by female.

Table 2 Average, minimum, and maximum of  $F_0$  in Tsurukame

Subject	Average (Hz)		Min (Hz)		Max (Hz)	
	Singing	Speaking	Singing	Speaking	Singing	Speaking
A	187.7	102.7	89.2	73.4	275.6	149.0
B	181.5	107.7	95.9	74.6	268.9	144.8
C	288.1	197.7	190.1	160.3	376.2	232.6
D	288.2	193.1	161.0	141.7	419.7	263.4

2.2 EGG parameter analysis: Sustained vowel

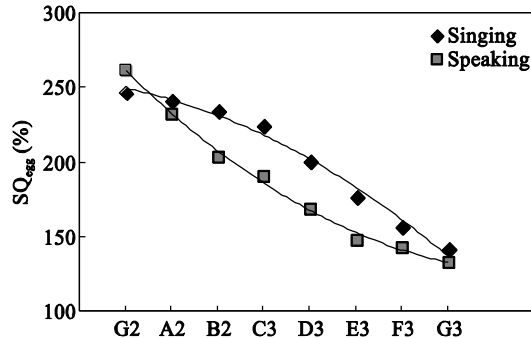
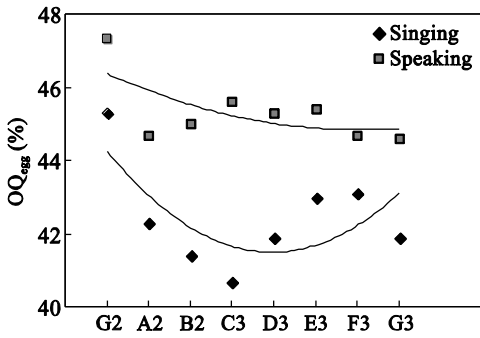


Fig. 7 Average of  $OQ_{egg}$  (left) and  $SQ_{egg}$  (right) for an octave (males)

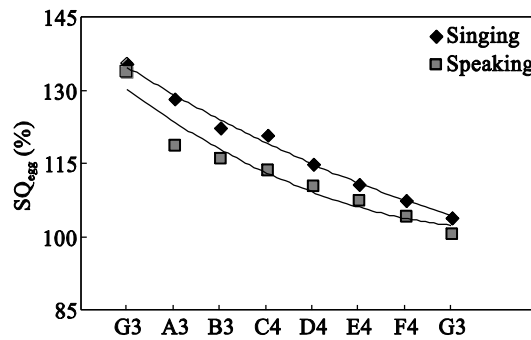
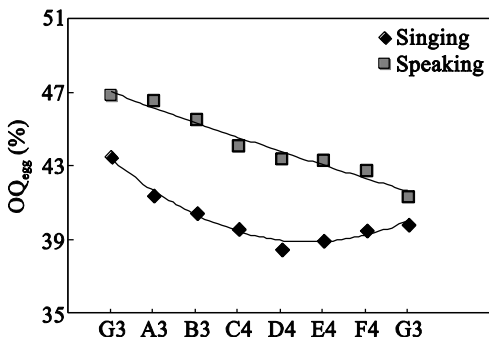


Fig. 8 Average of  $OQ_{egg}$  (left) and  $SQ_{egg}$  (right) for an octave (females)

The pitch range shown in Figs. 7 and 8 covers 1 octave from G2 (98 Hz) to G3 (196 Hz) for the males in Fig. 7, and from G3 (196 Hz) to G4 (392 Hz) for the females in Fig. 8. The average of  $OQ_{egg}$  and  $SQ_{egg}$  is given in black for singing and in gray for speaking. Polynomial fitting curves are also given in Figs. 7 and 8.

The data in Figs. 7 and 8 shows that  $OQ_{egg}$  of singing is lower than that of speaking both in males and females, while  $SQ_{egg}$  of singing is higher than that of speaking. Thus, singing is characterized by low  $OQ_{egg}$  and high  $SQ_{egg}$ . The  $OQ_{egg}$  of speaking tends to drop steadily as  $F_0$  increases. For the singing, the fitted polynomial curves for  $OQ_{egg}$  are rather arched with the lowest  $OQ_{egg}$  at C3 (131 Hz) for the males and D4 (294 Hz) for the females. The lowest  $OQ_{egg}$  for the males reaches 40.7% while that for the females reaches 39.6% with a maximum 5% gap between singing and speaking. Thus, very strong adduction of the vocal folds occurs during singing, especially around C3 for the males and D4 for the females.  $SQ_{egg}$  for both singing and speaking tends to drop steadily as  $F_0$  increases. Thus, Noh singing can be characterized as low  $OQ_{egg}$  and high  $SQ_{egg}$ .

The overall averages and standard deviations, SD, are given in Table 3. The average  $OQ_{egg}$  for singing is 42.5% for the males

and 40.2% for the females, which is 3%-4% lower than that for speaking. Strong adduction of the vocal folds with longer glottis closures is achieved in singing. The average  $SQ_{egg}$  for singing is 202.8% for the males and 118.1% for the females, which is 5%-18% higher than those for speaking. In this case, singing involves quick and forceful glottis closure. The standard deviation of  $OQ_{egg}$  is higher in singing, reaching 3.6% in males. The standard deviation of  $SQ_{egg}$  is also quite different between singing and speaking especially in males, with 23.5% for singing and 14.6% for speaking. These results illustrate the dynamic activities occurring in the glottis during Noh singing.

Table 3 Average and standard deviation of  $OQ_{egg}$  and  $SQ_{egg}$

		$OQ_{egg}$ (%)		$SQ_{egg}$ (%)	
		Average	SD	Average	SD
Male	Singing	42.5	3.6	202.8	23.5
	Normal	45.3	2.4	185.1	14.6
Female	Singing	40.2	1.9	118.1	8.9
	Normal	44.2	1.8	113.3	7.2

It is difficult to represent each parameter linearly, because natural human voices always have source variation. However these attempts can help recognize the inherent characteristics of the voice source parameters. The low  $OQ_{egg}$  and high  $SQ_{egg}$  are the

key features of Noh singing phonation.

The source characteristics can be described in terms of the glottal pulse shapes in the time domain, such as in the LF model<sup>[25]</sup>, and can also be described in terms of their spectral effects in the frequency domain. The shape of the glottal source is an important determinant of voice quality<sup>[26,27]</sup>. Figure 9, which is based on the LF model, shows a schematic of the glottal flow and the differentiated glottal flow for speaking and Noh singing. UP and EE are equivalent to the peak volume velocity of the glottal pulse and the excitation strength. A low  $OQ_{egg}$  and a high  $SQ_{egg}$  in the EGG, which are characteristics of Noh singing, are related to a small UP and high EE. Thus, Noh singing has a small peak volume velocity of the glottal pulse with a short glottal release while a strong excitation strength with forceful glottis closure. Thus, Noh singing characterized by the low  $OQ_{egg}$ , and high  $SQ_{egg}$  can be assessed as a pressed voice.

### 3. Phonation Analysis

In this section, the Noh singing phonation types are investigated using the EGG waveforms, EGG parameters, and acoustic spectra and spectrograms.

#### 3.1 EGG parameter analysis for each phonation type

Actual singing phonation involves more voice qualities than just sustained phonation, because it has rhythm and emotional factors. The Noh singing voice was classified as pressed phonation in Section 2 by analyzing the singing and speaking parameters. The voice quality of the actual singing voice from Tsurukame is assessed here.

Some Asian oral characteristics involve oscillations of the ventricular folds and aryepglottic folds. The EGG waveform

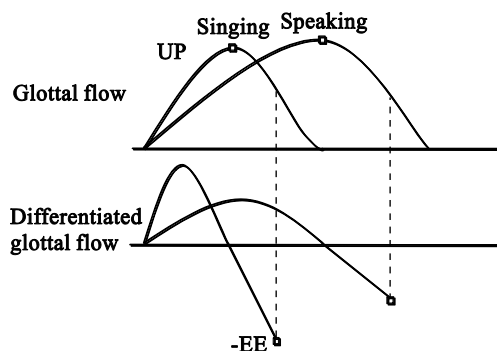


Fig. 9 Schematic of glottal flow and its derivative for singing and speaking

for VVM includes periods doubling<sup>[4,28,29]</sup>, while growl is described as a damped amplitude every other cycle<sup>[29]</sup>. The period doubling is caused by the phase difference between the vocal and ventricular fold oscillations. When the ventricular folds oscillate at  $F_0/2$  or  $F_0/3$ , subharmonics appear at  $F_0/2$  or  $F_0/3$  in the spectrum or spectrogram<sup>[4,29]</sup>. When the growl voice is analyzed by X-ray videofluoroscopy<sup>[6,29]</sup>, the larynx position is usually high, the aryepglottic region is compressed antero-posteriorly, and the tubercle of the epiglottis and arytenoid cartilages come into contact. The corresponding EGG waveform indicates less contact area in every other cycle<sup>[6]</sup>. These supraglottal vibrations are also thought to be used in Noh singing. The EGG waveforms for VVM and growl observed in Noh singing are shown in Fig. 10. The period doubling is found in VVM while damped amplitudes in every other cycle in growl.

The three kinds of EGG waveforms observed in Noh singing are growl, VVM, and pressed voice. Growl appears in the initial part of almost every phrase, while VVM and pressed voice appear during the rest of the sound. The  $F_0$  distributions for

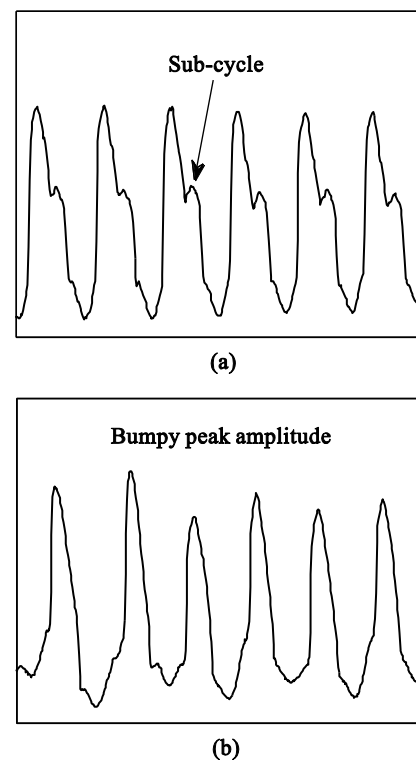


Fig. 10 6 cycle EGG waveform for (a) VVM and (b) growl from Tsurukame (subject A)

each phonation type are shown in Fig. 11.  $F_0$  of growl is the lowest and the range is the narrowest with an average  $F_0$  of

141.6 Hz for the males and 220.0 Hz for the females. The VVM frequency is in the middle with an average  $F_0$  of 165 Hz for the males and 249 Hz for the females. The  $F_0$  of the pressed voice is the highest and the range is the widest. Overlap occurs between each phonation type since each type covers a rather wide range, while it implies that switching between phonation types involves not only the pitch height but also techniques to add certain vocal effects to the singing voice. Thus, the  $F_0$  are characterized as: growl < VVM < pressed.

The  $OQ_{egg}$  variations in Fig. 12 show that growl has the highest  $OQ_{egg}$ , then pressed, and VVM with the lowest  $OQ_{egg}$ . The average  $OQ_{egg}$  for growl is 53.0% for the males and 45.5% for the females. The average  $OQ_{egg}$  for VVM is 39.4% for the males and 36.7% for the females. There is a significant difference between  $OQ_{egg}$  for growl and that for the other phonation types. For example,  $OQ_{egg}$  for growl is 9%-14 % higher than that for VVM. Moreover, though  $F_0$  for growl and the other phonation types overlap as shown in Fig. 11, less overlap occurs for  $OQ_{egg}$  in both males and females as shown in Fig. 12. Thus,  $OQ_{egg}$  is characterized as: VVM < pressed < growl.

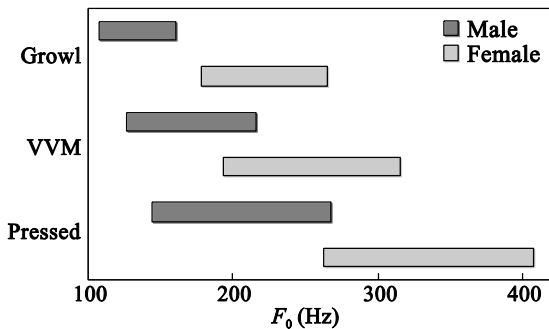


Fig. 11  $F_0$  Distributions for each phonation type

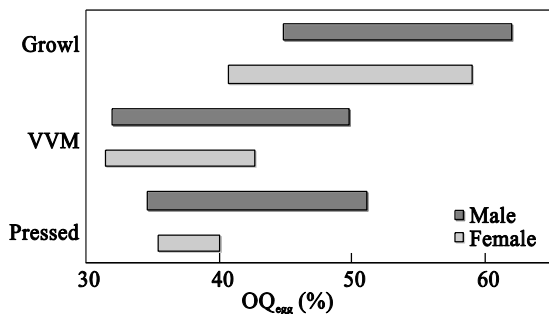


Fig. 12  $OQ_{egg}$  distribution for each phonation type

The  $SQ_{egg}$  for VVM is significantly higher than that of growl or pressed voices in both males and females as shown in Fig. 13 and covers a range of 170.2% in the male subjects. Though the  $F_0$  range of the pressed voice is the widest among

the three phonation types (Fig. 11), the  $SQ_{egg}$  variation is the most stable as shown in Fig. 13. Therefore, the vocal fold vibrations for the pressed voice can be characterized as stable, while the VVM involves dynamic laryngeal behavior result in very unstable  $SQ_{egg}$ .  $SQ_{egg}$  can be characterized as: pressed < growl < VVM.

The distinctive features of each phonation type are shown in Table 4. As noted in Section 3, Noh singing is classified as pressed phonation with low  $OQ_{egg}$  and high  $SQ_{egg}$ , whereas the growl voice has a peculiar phonation quality with high  $OQ_{egg}$  and low  $SQ_{egg}$  caused by less contact area of the vocal folds resulting from the aryepiglottic fold oscillations. Although the EGG parameters indicate a significant difference between growl and the other two phonation types, the supraglottal oscillation is a common feature in both growl and VVM. VVM is characterized by the lowest  $OQ_{egg}$  and the highest  $SQ_{egg}$ . As a result, VVM and pressed voice are classified as the pressed type phonation. VVM is characterized as the more pressed, energetic and dynamic phonation type. On the other hand, growl is characterized as a rather lax phonation type.

### 3.2 Switching of phonation types and acoustic analysis

Subharmonics are often found in voice instabilities,

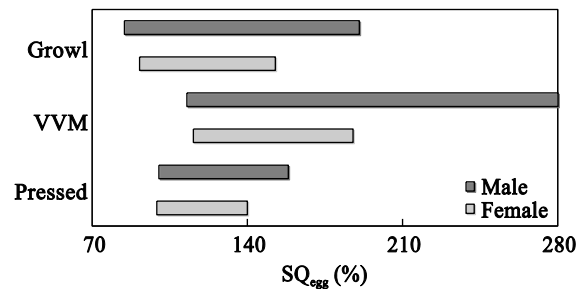


Fig. 13  $SQ_{egg}$  distribution for each phonation type

Table 4 Distinctive features of the source parameters for the three phonation types. “-” indicates low, “+” means high, “±” is middle.

	$F_0$	$OQ_{egg}$	$SQ_{egg}$
Growl	-	+	±
VVM	±	-	+
Pressed	+	±	-

infant vocalizations, some paralinguistic features of speaking, and singing techniques<sup>[4,7,8,29]</sup>. Subharmonics are also observed in Noh singing. Figure 14 shows the switching of

phonation types in actual Noh singing, (a)  $F_0$  (indicated in black), (b)  $OQ_{egg}$  (indicated in gray) and  $SQ_{egg}$  (indicated in black), and (c) corresponding narrow-band spectrogram. The parameters are all normalized. Figures 14a and 14b provide a typical example of each phonation type's source features with growl having high  $OQ_{egg}$  and low  $SQ_{egg}$  and VVM having low  $OQ_{egg}$  and high  $SQ_{egg}$ . Subharmonics are observed both in spectrograms in Fig. 14c and the spectrum in Fig. 15 in the growl region. As seen in Fig. 15, the 63 Hz undertone is relatively weaker than the original harmonic peak at 126 Hz with the spectrum including clear subharmonics, indicated by the black arrows in Fig. 15, up to the high frequency region.

The frequency of the ventricular fold oscillations is the same as  $F_0$  in both males and females, with  $F_0/2$  also observed in females. Clear subharmonics appear due to the ventricular fold oscillations at  $F_0/2$  in Fig. 16.

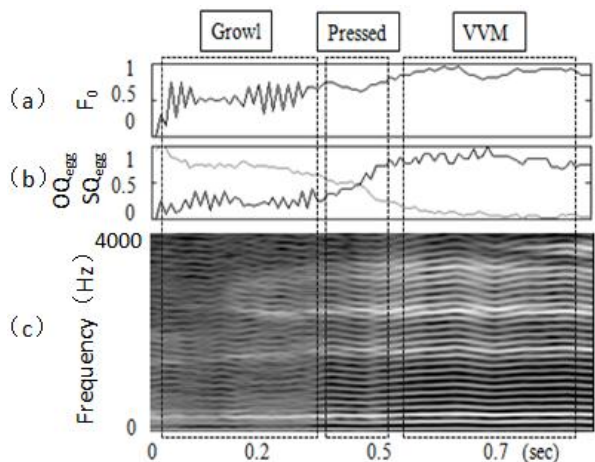


Fig. 14 Switching of phonation types in Tsurukame (subject A) (a)  $F_0$  (black), (b)  $OQ_{egg}$  (gray) and  $SQ_{egg}$  (black), and (c) spectrogram

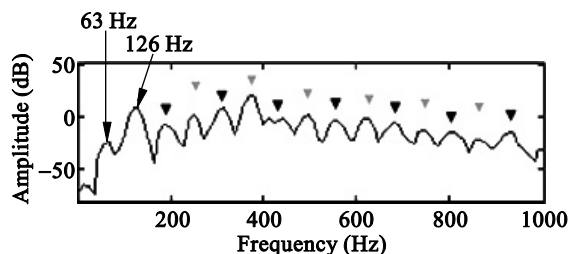


Fig. 15 Spectrum (range 0-1000 Hz) of growl from subject A. Black arrows indicate subharmonic peaks, gray arrows indicate original harmonic peaks.

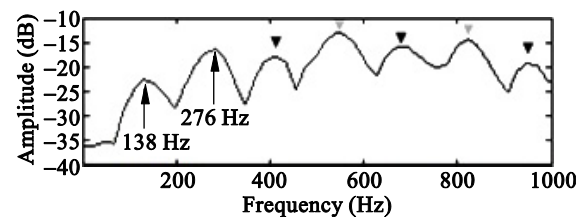


Fig. 16 Spectrum (range 0-1000 Hz) of VVM from subject C with an undertone observed at 138 Hz

$F_0$  is 276 Hz, with the subharmonics at 138 Hz which is equivalent to  $F_0/2$  and other harmonics at higher frequencies.

A vibrato example from Tsurukame is shown in Fig. 17. Since the main perceptual effect of the vibrato depends on the frequency modulation<sup>[30]</sup>, in Noh  $F_0$  also shows sinusoidal modulation in Fig. 17a. Period doubling is observed in the EGG signal in Fig. 17d in the low pitch region in Fig. 17a with low  $OQ_{egg}$  and high  $SQ_{egg}$  in Fig. 17b. Thus, the vibrato involves not only the  $F_0$  modulation but also a combination of phonation types.

Thus, vibrato is achieved by the combination of  $F_0$  modulation and the switching of phonation types. Some VVM and growl subharmonics add low pitched impressions to the sounds.

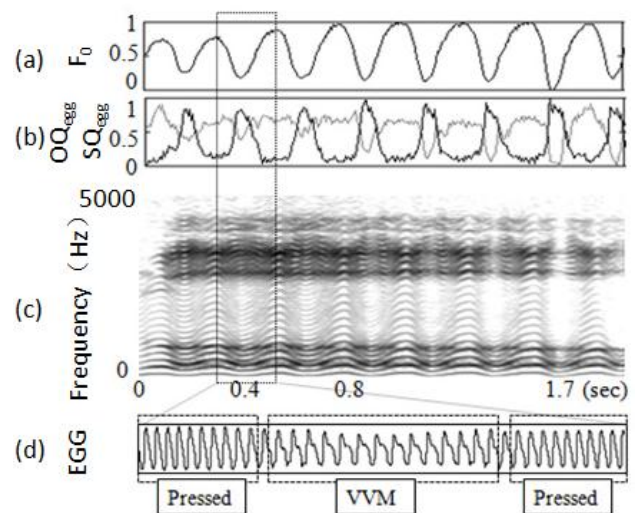


Fig. 17 Vibrato from subject A. (a)  $F_0$  (black), (b)  $OQ_{egg}$  (gray) and  $SQ_{egg}$  (black), (c) spectrogram, and (d) selected EGG signals

## 4. Conclusions

This study analyzed the phonation mechanisms of Japanese traditional Noh using physiological and acoustical methods.



Three types of phonation occur in Noh singing: pressed, VVM and growl. Among the so-called extended vocal techniques, VVM is the representative Asian traditional vocal technique which sounds solemn and ancient, while growl is the animalized sound which delivers passionate and dramatic vocal effects to the singing. Noh singing is characterized by low pitched sounds, however, the actual  $F_0$  is rather high and the pitch range is rather wide. Subharmonics, generated by the ventricular and aryepiglottic fold oscillations at a frequency of  $F_0/2$ , add the low pitched sound effect to the singing. The unique mixed vocal timbre resulting from the combination of phonation types creates a compelling effect to Noh singing.

Further physiological measurements using other techniques such as high-speed cameras are needed to clarify the laryngeal behaviors of these peculiar phonations. Synthesize and perceptual evaluations are also expected in future work.

## 5. Acknowledgements

The authors thank Yoshio Hosho of the living national treasure from the Hosho school of Noh, his pupils and Mr. and Mrs. Ueda from Toshiba for the recordings and valuable advice on this study.

## 6. References

- [1] Nakayama I. Comparison of vocal expressions between Japanese traditional and western classical-style singing, using a common verse. *The Journal of the Acoustical Society of Japan*, 2000, **56**(5): 343-348.
- [2] Yoshioka N, Nagahata D, Yanagida M, et al. Differences among vowels used in Noh, Kyogen and European classical singing. Technical report of IEICE, 2001, **122**: 1-8.
- [3] Esling J H. Pharyngeal consonants and the aryepiglottic sphincter. *Journal of the International Phonetics Association*, 1996, **26**(2): 65-88.
- [4] Fuks L, Hammarberg B, Sundberg J. A self-sustained vocal-ventricular phonation mode: Acoustical, aerodynamic and glottographic evidences. *KTH TMH-QPSR*, 1998, **3**: 49-59.
- [5] Lindstad P-Å, Södersten M, Merker B, et al. Voice source characteristics in mongolian “throat singing” studied with high-speed imaging technique, acoustic spectra, and inverse filtering. *Journal of Voice*, 2001, **15**(1): 78-85.
- [6] Sakakibara K, Fuks L, Imagawa H, et al. Growl voice in ethnic and pop styles. In: Proceedings of the International Symposium on Musical Acoustics. Nara, Japan, 2004.
- [7] Hollien H, Michel J, Thomas E D. A method for analyzing, vocal jitter in sustained phonation. *Journal of Phonetics*, 1973, **1**: 85-91.
- [8] Titze I R, Baken R J, Herzel H. Evidence of chaos in vocal folds vibration in vocal fold physiology. In: New Frontiers in Basic Science. San Diego: Singular Publishing Group, 1993: 143-188.
- [9] Fabre P. An electrical inscription of the glottis during phonation: glottography from high frequency. *Journal of the National Academy of Medicine*, 1957, **141**: 66-69. (in French)
- [10] Kong Jiangping. On Language Phonation. Beijing: Central Nationalities University Press, 2001: 173-189. (in Chinese)
- [11] Titze I R. Interpretation of the electroglottographic signal. *Journal of Voice*, 1990, **4**: 1-9.
- [12] Stevens K N. Physics of larynx behavior and larynx modes. *Phonetica*, 1977, **34**: 264-279.
- [13] Kent R D, Ball M J. Voice Quality Measurement. California: Singular Publishing Group, 2000: 117-118.
- [14] Baken R J, Orlikoff R F. Clinical Measurement of Speech and Voice. New York: Delmar learning, 2000: 413-427.
- [15] Henrich N. On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *Journal of Acoustic Society of America*, 2004, **115**: 1321-1332.
- [16] Herbst C. Evaluation of various methods to calculate the EGG contact quotient [Dissertation]. Stockholm: KTH Speech, Music and Hearing, 2004.
- [17] Howard D M, Lindsey G A, Allen B. Toward the quantification of vocal efficiency. *Journal of Voice*, 1990, **4**: 205-212.
- [18] Howard D M. Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. *Journal of Voice*, 1995, **9**(2): 163-172.
- [19] Rothenberg M, Mahshie J J. Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech and Hearing Research*, 1988, **31**: 338-351.
- [20] Childers D G, Hicks D M, Moore G P, et al. Electroglottography and vocal fold physiology. *Journal of*

- Speech and Hearing Research*, 1990, **33**: 245-254.
- [21] Childers D G, Krishnamurthy A K. A critical review of electroglottography, *Critical Reviews in Biomedical Engineering*, 1985, **12**: 131-161.
- [22] Childers D G, Larar J N. Electroglottography for laryngeal function assessment and speech analysis. *IEEE Transactions on Biomedical Engineering*, 1984, **31**: 807-817.
- [23] Childers D G, Moore G P, Naik J M, et al. Assessment of laryngeal function by simultaneous, synchronized measurement of speech, electroglottography and ultra-high speed film. In: *The Eleventh Symposium: Care of the Professional Voice*. New York: Voice Foundation, 1983: 234-244.
- [24] Childers D G, Naik J M, Larar J N, et al. Electroglottography, speech and ultra-high speed cinematography. In: *Vocal Fold Physiology and Biophysics of Voice*. Denver: Denver Center for the Performing Arts, 1983: 202-220.
- [25] Fant G, Liljencrants J, Lin Q. A four parameter model of glottal flow. *STL-QPSR*, 1985, **4**: 1-13.
- [26] Ní Chasaide A, Gobl C. Voice Source Variation. *The Handbook of Phonetic Sciences*. Oxford: Blackwell Publishers Ltd, 1997: 427-461.
- [27] Fant G. *Speech Acoustics and Phonetics*. Boston: Kluwer Academic Publishers, 2004: 249-300.
- [28] Esling J H. A laryngographic investigation of phonation type and laryngeal configurations. Working papers of the Linguistic circle. Canada: University of Victoria, 1983: 14-36.
- [29] Sakakibara Ken-Ichi, Imagawa Hiroshi, Niimi Seiji, et al. Physiological study of the supraglottal structure. In: *Proc. International Conference on Voice Physiology and Biomechanics*. Marseille, France, 2004.
- [30] Johan Sundberg. *The Science of the Singing Voice*. Illinois, USA: Northern Illinois University Press, 1987: 163-176.