Analysis of Human Vocal Fold Vibration by Means of Glottal Velocity Measurement and High-Speed Imaging

Shiro ARII**, Hideyuki KATAOKA***, Yoshitaka OCHIAI**, Kensaku HASEGAWA***, Toyohiko SUZUKI**, Hiroya KITANO*** and Hiroshi MIZUMOTO**

** Graduate School of Engineering, Tottori University
4–101 Minami Koyama-cho ,Tottori, Tottori 680–8552, Japan
E-mail: arii@ike.tottori-u.ac.jp
*** Faculty of Medicine, Tottori University
36–1 Nishi-machi, Yonago, Tottori 683–8504, Japan

Abstract

Human voices originate from the vibration of vocal folds in the larynx. In most previous studies on voice generation, a planar sound source was assumed for a laryngeal sound source and the effects of airflow in a larynx were neglected. However, no direct in vivo measurement of human glottal velocity has been reported. Therefore, detailed study of this airflow is necessary to elucidate mechanisms of human phonation. In the present study, airflow just above the glottis was experimentally analyzed to clarify the phonation mechanism and seek better modeling of vocal folds. This experiment focused on direct measurement of the airflow velocity by means of a tiny hot-wire probe and simultaneous observation of vocal fold movement by means of a high-speed digital camera. Experimental results show that the periodic change in the airflow velocity occurs out of phase with the opening of the glottis, although both have the same fundamental frequency. This is important because it provides crucial information to create better simulation models of the phonation mechanism.

Key words: Vocal Fold Vibration, Voice Source, Airflow Velocity, High-Speed Imaging, Sound and Acoustic, Modeling, Biomechanics

1. Introduction

Characteristics of human voice sounds are traditionally explained using the source–filter theory. In this theory, temporal variation of glottal opening during phonation is considered to be fairly well approximated by a triangular periodic change. Such waveforms are commonly used to represent the rate of air flow passing glottal constriction. Thus, the pressure variations of vocal folds are assumed to have a waveform similar to that of the volume flow rate. Therefore, in most previous research on voice generation, a planar sound source was assumed for a laryngeal sound source and the effects of airflow in a larynx were neglected(1), (2).

Recently, high-speed imaging using a rigid endoscope was performed to observe directly the motion of vocal folds during phonation(7)–(9). Some methods were applied to measure airflow rate; however, these methods were limited to the study of laryngeal physiology(3)–(6). Consequently, no direct in vivo measurement of human glottal velocity has been reported. The effects of the change in the area of glottal opening caused by the vibration of vocal folds during phonation on the fluctuation of airflow velocity just above the glottis have not been shown experimentally. Thus, the important assumption of the source–filter theory has not been confirmed.

In order to clarify the relationship between the glottal opening area and glottal velocity during phonation, this paper describes in vivo measurements of airflow velocity just above
the glottis and the digital high–speed imaging of the vibrating vocal folds in a human. The mutual interaction between the external mucous membrane of the vocal folds and the airflow was subjected to more detailed experimental analysis.

2. Principles of phonation and laryngeal sound source models

In this section, we briefly summarize the principles of phonation and laryngeal sound source models studied in previous research\(^1\), \(^2\). Humans produce sounds by oscillating the vocal folds in the right and left sides of the larynx, which is located at the inlet of the trachea where it branches into the esophagus. Thus, the larynx’s essential function is to prevent foreign matter from being swallowed in the trachea. The vocal folds are composed mainly of an internal vocalis muscle and an external mucous membrane. This membrane freely oscillates to the airflow through the larynx.

The glottis is the space between the vocal folds. It is almost closed during phonation, whereas it is completely open during breathing without phonation. The opening and closing of the glottis is made possible by the movement of the laryngeal cartilage caused by the actions of internal laryngeal muscles that consist mainly of the cricothyroid muscles and vocalis muscles.

Airflow from the lungs opens the glottis effectively by its pressure, and then the air starts to flow through the larynx. At that instant, the sub-glottis pressure decreases and the glottis closes because the high sub-glottal pressure ceases. The vocal folds’ oscillation is considered to be a repetitive sequence of the opening–closing movement of the vocal folds. The temporal variation of the opening area of the glottis has a significant correlation to the generation of voice sounds. In this way, the complex structure of the vocal folds and their interactions with the airflow and the oscillation frequency mainly determine the particular feature of each individual vocal sound.

Human voice originates following the intermittent airflow through the larynx caused by vibration of the vocal folds produces the origin of the human voice. In the traditional theoretical model of the sound source of the human voice, assuming that the pressure of the lungs is constant, the air flows through the larynx while the glottis is open and is interrupted while the glottis is closed. The time-based variation of the volume flow rate of the airflow through the glottis is the same as that of the open area of the glottis. Thus, the time-based variation of the sound pressure at the glottis is simply treated as the time-based variation of the open area of the glottis, as shown in Fig. 1.

Although humans cannot control the vibration of the mucous membrane of the vocal folds directly, they can control the laryngeal muscles and expand or contract the vocal folds indirectly. The frequency of the vibration of the vocal folds changes with the deformation of the vocal folds. Therefore, humans can control the pitch of the laryngeal sound. It is said that the mean airflow rate required for phonation is 100 to 200 mL per second in the case of a human adult, and the dynamic pressure of the airflow is 5 to 10 mm H\(_2\)O. The air pressure in the lungs may not oscillate, and thus, the vocal folds’ oscillation is essentially a consequence
of the dynamic interaction between the elasticity of the vocal folds and the airflow through them, that is, a self-resonating oscillation of a fluid-structure interaction.

3. Measurement of airflow velocity just above glottis and high-speed imaging of vocal folds’ vibration during phonation

3.1. Measuring system for airflow velocity

A schematic of the measuring system for the airflow velocity is shown in Fig. 2. The system consists of small hot-wire probes, a constant temperature anemometer (CTA; KANOMAX MODEL 1011) and a flexible transnasal endoscope (OLYMPUS ENF Type T3). The outputs from the CTA are monitored through the experiments via an oscilloscope (IWATSU SS-7602) and stored in a personal computer (IBM ThinkPad Pentium III 500MHz) after an analog-to-digital (A-D) conversion with a 16-bit conversion rate (Interface CBI-320416). Furthermore, to record the outputs synchronized with the images of a high-speed camera, one more A-D converter with an 11-bit conversion rate (PHOTORON MCDL BOX) is connected to the high-speed camera. The sampling rates of the A-D conversion used in the personal computer and the MCDL were 100 and 20 kHz, respectively. The output of this converter is stored in the memory of the high-speed camera. The hot wire of the probes, which is 5 µm in diameter, is made of tungsten. The CTA and probes have a frequency response good enough to measure the air velocity within an accuracy of 1.00% in a frequency range of up to 10 kHz. At the tip of the insertion tube of the endoscope, a lens and light source for monitoring an object and an instrument channel outlet are installed.

The small hot-wire probes were designed and fabricated in the laboratory for this partic-

![Diagram](image-url)
ular application in the present research, so that they could be inserted into the larynx through an instrument channel of the flexible insertion tube of the endoscope. The maximum diameter of the probe was 1.8 mm. It was attached to the tip of a stainless steel coil that was 1 m in length and 1.6 mm in diameter. Figure 3 (a) and (b) shows one of the probes and the probe inserted into the endoscope.

Calibration curves were obtained for all the hot-wire probes and then used to convert output voltage into airflow velocity. The small wind tunnel shown in Fig. 4 (a) was used for the calibration. It consists of a diverging nozzle and a flow straightener section of honeycomb sheet. At the test section in the downstream of the straightener section, static and pitot pressure were measured, from which the airflow velocity at that section was calculated. At the same time, the output of the hot-wire probe was recorded via the CTA. The airflow velocity and the hot-wire voltage output produced the calibration curve for the probe. The test was performed for a range of velocities lower than 20 m/s, taking into consideration the general airflow velocity range in the larynx during phonation. A typical calibration curve of a hot-wire probe is shown in Fig. 4 (b).

3.2. High-speed digital imaging system

High-speed images of the vibrating vocal folds during phonation are taken using the lens of the endoscope. The insertion tube of the endoscope is installed through the nasal cavity, and the tip of the tube is fixed at the proper position just above the glottis, as shown in Fig. 2. The eyepiece of the endoscope is connected to a high-speed camera (Photoron FASTCAM-MAX...
120K) using a C mount adapter (OLYMPUS MC-R44), and the images taken by the endoscope are shown on a monitor during the experiment. The optical fiber for the light source of the endoscope is connected to a light source device using a 300 W xenon lamp (OLYMPUS EVIS Universal light source CLV-U40).

In order to take clear 512 × 512 pixel images at a rate of 2000 frame/s, another endoscope is used to supply a beam of light to the glottis. The optical fiber of this endoscope is connected to another light source device with a 300 W xenon lamp (OLYMPUS EVIS Universal light source CLV-U40). Furthermore, auxiliary thin light guides, as shown in Fig. 5 (a), were designed. The light guide is inserted into the instrument channel of the endoscope, as shown in Fig. 5 (b). The light guide is connected to the third light source device with a 300 W xenon lamp (OLYMPUS EVIS Universal light source CLV-U20D).

3.3. Measuring method

In order to avoid or suppress a laryngeal reflex caused by the insertion of the endoscope tube and allow naturally vocalization during the experiment, the tube was inserted through the nasal cavities. The nasal cavities of the subject were sufficiently anesthetized locally using 4 % xylocaine in advance. The first tube used for the airflow measurement was inserted and fixed at the proper position, where the tip of the tube comes across the epiglottis to a point just above the glottis. The second tube used for auxiliary light was inserted in another nostril and fixed at the proper position, where the tip of the tube is behind the first tube and the beam illuminates the glottis well.

The hot-wire probe was inserted into the instrument channel of the first endoscope. After using the monitor screen to ensure that the probe reached the tip of the tube and was pushed out several millimeters from the instrument channel outlet, the probe was retracted slightly and held just inside the outlet. Recording of the airflow velocity and the high-speed images started the moment the subject started to vocalize. The probe was pushed out from the outlet simultaneously with the vocalization. The position of the probe was carefully controlled above the glottis by checking the monitor screen. The new probe was used in every measurement. After the measurement, the probe was checked to see whether the hot wire was clean, because any adherent such as mucus may damage the frequency response of the CTA. The
data measured when the hot-wire probe was dirty were omitted from the record.

Figure 6 (a), (b), and (c) are given to show the location of the hot-wire probe in relation to the vocal folds. These figures represent the typical phases of the glottal opening. The phase shown in (a) is a relatively opened glottis during breathing. The phase shown in (b) is an almost closed phase during phonation with vibrating vocal folds. The phase shown in (c) is a completely opened phase during deep breathing.

4. Results and discussion

4.1. Results for airflow velocity passing glottis

Figure 7 (a) shows a typical temporal variation of the airflow velocity. The experiments were conducted with the phonation of the vowel “ah.” A man in his early thirties cooperated as the subject. Figure 7 (b) shows the corresponding Fourier spectrum. It is seen that the average velocity is around 2.5 m/s. Naturally, the average velocity increases when a louder sound is pronounced. It was noticed that the fundamental frequency component and its several harmonics mainly comprise the lower frequency part of the signal. The fundamental frequency was identified at 0.13 kHz, which is in accordance with that of the opening of the glottis. It should be pointed out that a very high frequency component of a significant level was found at 7.7 kHz. Thus, the airflow velocity was seen to change mainly with the fundamental frequency of 0.13 kHz and some harmonics, but it was also superposed with a significant level of the 7.7 kHz component. The fundamental frequency of other subjects who cooperated in the experiments differed from that shown in Fig. 7 (a), and their level and frequency of high frequency components differed from that shown in Fig. 7 (b). Nonetheless, the subjects shared similar features.

It was discovered, however, that anesthetization of the vocal folds might change this tendency. Figure 8 (a) shows a typical temporal variation of the airflow velocity after the anesthetization performed on the same subject whose data is shown in Fig. 7, and Fig. 8 (b) shows the corresponding Fourier spectrum. Note that the anesthetization in this experiment
shown in Fig. 8 was applied not only to the area of the nasal cavities but also to all the vocal folds. The level of the high frequency component became very low, and the frequency of the fundamental frequency changed to 0.15 kHz. Anesthetization clearly influenced the produced sound as well as the airflow velocity near the outlet of the vocal folds, and the researchers noticed this change during the experiment. Considering these changes, anesthetization may have hardened the vocal folds and restricted the vibration of the external mucous membrane due to airflow through the larynx.

According to the source–filter theory, the sound spectrum of human sound source consists of the fundamental frequency component and its several harmonics, and the first and second formant, that is, the first and second resonances in the vocal tract, determine the shape of the sound spectrum of vowels. However, it was discovered that high frequency components are found in the airflow velocity measured just above the glottis. The frequencies of these components are different from those of the first and second formants. It is considered that the high-frequency components result from the vibration of the external mucous membrane of the vocal folds, and that this vibration is caused by the airflow passing through the glottis during the closing movement of the vocal folds.

4.2. Correlation between the opening-closing movement of vocal folds and airflow velocity passing the glottis

The pictures of vibrating vocal folds were taken by a digital high-speed camera synchronized with the measurement of airflow velocity. A man in his early twenties cooperated as the subject for these images. The experiments were performed with the phonation of the vowel “ih.” The pictures were taken at a rate of 2000 frames a second. The sampling frequency of the measurement of the airflow velocity was 100 kHz. The series of pictures are shown in Figs. 9 and 10. The pictures show only the area around the vocal folds, and each image is 150 × 150 pixels. Figure 11 shows the recorded airflow velocity synchronized with the high-speed camera. The timing of each picture in the series is shown in Fig. 11 as a black solid circle on
the airflow velocity versus time curve. From this series of pictures, it can be seen that a new closing phase in a periodic cycle starts roughly at 3 ms and ends at around 4.5 ms. An opening phase in the same cycle starts about 5 ms. The next closing phase starts roughly at 8 ms. The period of the periodic cycle, fundamental period, is about 5 ms. According to the Fourier analysis of the corresponding airflow velocity shown in Fig. 11, the fundamental frequency was 195 Hz; that is, the fundamental period was 5.12 ms.

The color of the space between the vocal folds is black in the pictures, because the light cannot be reflected. The area of the opening glottis can be calculated by counting the black pixels in the pictures. Because the distance between the lens of the endoscope and the vocal folds could not be correctly measured, it was impossible to calibrate the opening area of the glottis using one square pixel on the images. Thus, a qualitative description was provided to explain how the airflow velocity changes with the opening and closing timing of the glottis. The number of black pixels in the space between the vocal folds in each picture is shown in Fig. 11 as a red solid circle above the airflow velocity. In Fig. 11, the number of black pixels in each picture is divided by the maximum number of black pixels. The value of the area is equal to 1 at the widest opening and 0 at completely closing.

During the closing phase of the glottis, the airflow velocity was at its highest, and during the opening phase, the airflow velocity was low. This fact clearly explains that the observed closing phase of the glottis is not “a complete closing” of the glottis, but rather “a fully constricted” phase, so that the air flows like a jet from the glottis at the almost highest velocity. The periodic change of the airflow velocity occurs out of phase with the opening of the glottis, though both have the same fundamental frequency.

The experimental results described above are completely different from the traditional theoretical model of the sound source of human voice. The air from the lungs flows all the time during phonation. The vocal folds do not stop the airflow but constrict it. The constricted airflow vibrates the external mucous membrane of the vocal folds. It is considered that this
coupled vibration of the airflow and the vocal folds produces the high frequency component of the airflow velocity. This fact is important since it gives a crucial hint for better simulations of the phonation mechanism.

It can be also considered that the air pressure in the lungs and the trachea fluctuates in synchronicity with the opening and closing of the glottis. In order to create better simulations, it is necessary to consider simultaneously the fluctuation of the following elements during phonation: volume of the lungs, pressure inside the lungs, and the rate of airflow passing through the glottis.
5. Conclusions

The airflow velocity in the larynx during phonation was experimentally studied. The vocal folds’ motion was directly observed simultaneously with the measurement of the airflow velocity. The following conclusions may be drawn from the study.

(1) The periodic change in the airflow velocity occurs out of phase with the opening of the glottis, though both have the same fundamental frequency. This fact is important since it gives a crucial hint for a better simulation model of the phonation mechanism.

(2) The air from the lungs flows throughout during phonation. The vocal folds do not stop the airflow but constrict it.

(3) The constricted airflow vibrates the external mucous membrane of the vocal folds. The coupled vibration of the airflow and the vocal folds results in the high-frequency component of the airflow velocity.

(4) In order to create better simulations of the phonation mechanism, it is necessary to consider the fluctuation simultaneously of the following elements during phonation: volume of the lungs, pressure inside the lungs, and the rate of airflow passing through the glottis.

Acknowledgments

The authors would like to thank the Japan Society for the Promotion of Science for its financial support of this work under Grants-in-Aid for Scientific Research Nos. 16591708 and 17656081.

References