

Modal Analysis of Vocal Fold Vibrations Using Laryngotopography

Ken-Ichi Sakakibara^{1,2}, Hiroshi Imagawa²,
Miwako Kimura², Hisayuki Yokonishi², Niro Tayama^{2,3}

¹ Department of Communication Disorders, Health Sciences University of Hokkaido, Japan

² Department of Otolaryngology, University of Tokyo, Japan

³ Department of Otolaryngology, Head and Neck Surgery, National Center for Global Health and Medicine, Japan

kis@hoku-iryu-u.ac.jp, imagawa@m.u-tokyo.ac.jp,

mkimu-tky@umin.ac.jp, yokonishi-tky@umin.ac.jp, ntayama@hosp.ncgm.go.jp

Abstract

In this paper, we propose a method for analyzing spatial characteristics of the larynx during phonation by high-speed digital imaging.

The laryngotopography was applied to the high-speed digital images of normal subjects, and patients with paralysis and cyst. The results show various modes of vibration of the vocal folds particular to the patients with paralysis and cyst and usefulness of the laryngotopograph for clinical purposes.

Index Terms: vocal fold vibration, high-speed digital imaging, modal analysis

1. Introduction

Vocal fold vibration patterns regulate acoustical characteristics of voices, such as fundamental frequencies and voice quality. Therefore, observation of vocal fold vibratory patterns plays an important role in a better understanding of relationships between the human voice production mechanism of laryngeal sources and vocal characteristics.

A high-speed digital imaging technique has been employed for direct observation of vocal fold vibratory patterns and various analysis methods for high-speed images have been proposed. As analysis methods, glottal width functions, glottal area functions, and kymographs, are widely used [1]. These analysis methods are effective in analyzing temporal characteristics of vocal fold vibratory patterns, however, these do not clarify spatial characteristics of the larynx during phonation.

In this paper, we propose *laryngotopography*, which is a method for extracting spatial characteristics of the larynx by high-speed digital imaging. The laryngotopography is based on idea in [2], and the laryngotopography has been improved and implemented in our clinical studies.

2. Laryngotopography

2.1. High-speed digital imaging

The high-speed digital camera Photoron, FASTCAM-1024PCI at a frame rate 4500 fps, image resolution of 400×512 pixels, 8-bit grayscale, and memory size of 12 GB allowing sampling duration of 11.1 s, was employed. A rigid endoscope (#4450.501, Richard Wolf) was attached to an attachment lens (f = 35 mm, Nagashima Med. Instrument Corp.) connected to

the camera. High-speed digital images were simultaneously recorded with EGG and sound signals.

2.2. Analysis method

Laryngotopography is a method for analyzing high-speed digital images based on pixel-wise Fourier analysis of a time-varying brightness curve for each pixel across images [2]. The procedure of analysis is as follows: (i) a rectangular area is selected; (ii) a time-varying raw brightness curve is extracted from the images (Figure 1); (iii) for each pixel, the average level of brightness of consecutive frames (512 or 256 frames) are calculated; (iv) the brightness curve is normalized by subtracting the average level from the original raw brightness curve; (v) The Hamming window is applied to the normalized brightness curve and discrete Fourier transform (DFT) of 1024 points by adding zero-padding to increase a frequency resolution is applied to the normalized brightness curve.

When the high-speed imaging is conducted at 4500 fps, the frame size for analysis is 0.114 s (512 frames) or 0.057 s (256 frames) and the frequency resolution is 3.7 Hz.

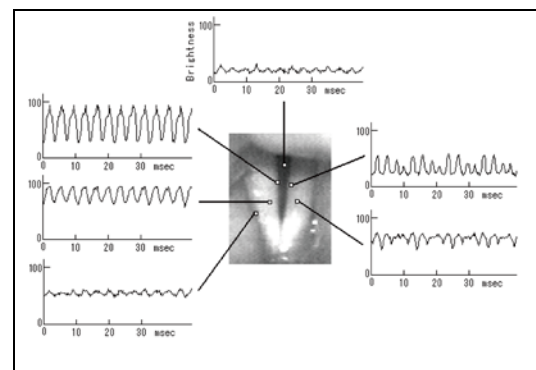


Figure 1: Examples of brightness curves on several pixels.

2.3. Implementation and visualization

Values, such as a frequency, amplitude, and phase, obtained by the pixel-wise Fourier transform were visualized in several manners. Some of our proposed visualizations are similar to the method presented in [2], and some are novel.

2.3.1. Amplitude spectrum of a single pixel

A single pixel is selected and the amplitude spectrum obtained by DFT for the brightness curve of the pixel is displayed in a two-dimensional graph (Figure 1, Figure 2(e)).

2.3.2. Spatial distribution of amplitude, frequency, and phase of the maximum-amplitude components

In this visualization method, Fourier transforms are conducted for all pixels, and the maximum amplitude component is determined for each pixel. Then, the amplitude, frequency, and phase of the maximum amplitude components are superimposed at each pixel of one frame in the image sequence.

When the different vibration modes are superposed, the second-maximum amplitude component is also selected and its amplitude, frequency, and phase are displayed in another panel. The second-maximum component is obtained by the following procedure: let f_1 be the frequency of the maximum amplitude component, then the second-maximum amplitude component at the frequency f_2 is determined as satisfying the condition $f_1/2 < f_2 < 2f_1$.

Figures 2 (b), (c), and (d) show the amplitudes, frequencies, and phases of the maximum amplitude components at all pixels, respectively. The amplitude, frequency, and phase of the brightness curve are represented in terms of color saturation. In this case, the frequencies of the maximum amplitude components at the pixels are equal.

2.3.3. Spatial distribution of the amplitude of a single frequency

A single frequency is selected. The amplitude and phase of the component at the selected single frequency at each pixel is displayed in terms of color saturation. Figures 9 (a) and (b) show the amplitude and phase of the components at 251 Hz, and Figures 9 (c) and (d) show those at 313 Hz.

2.3.4. Spatial distribution of amplitude ratio of second-maximum components to maximum components

For each pixel, let A_1 be the amplitude of the maximum amplitude component and A_2 be the amplitude of the second-maximum component, then a ratio of the amplitude of the second-maximum amplitude component to that of maximum amplitude component A_2/A_1 is displayed at each pixel in terms of color saturation. It is reasonable to regard a large A_2/A_1 as occurrence of two different frequency oscillations in the larynx.

2.4. Results

2.4.1. Normal subjects

Figures 2 (a)-(d) are the laryngotopographs of a male subject without any vocal problems, age of 39. Figure 2 (a) shows the selected area for analysis, and Figures 2 (b), (c), and (d) show spatial distribution of amplitudes, frequencies, and phases of the maximum amplitude components, respectively. In (b), the middle portion of the glottis shows the maximum amplitude of the brightness oscillation. Because the glottal closure in this case was complete, the brightness of the pixels at the glottal space dynamically changed low to high. In Figure 2 (c), the distribution of the frequencies is almost uniform at 128 Hz. In Figure 2 (d), the phases of the maximum components on the vocal folds are longitudinally varied. The antero-posterior difference is 60 degrees.

Figures 3 (a)-(d) are the laryngotopographs of a female subject without any vocal problems, age of 25. The distribution of the amplitudes and frequencies of the maximum amplitude components are almost similar to that of the male subject. Figure 3 (d) displays a lateral phase difference. The left-to-right phase difference was 30 degrees.

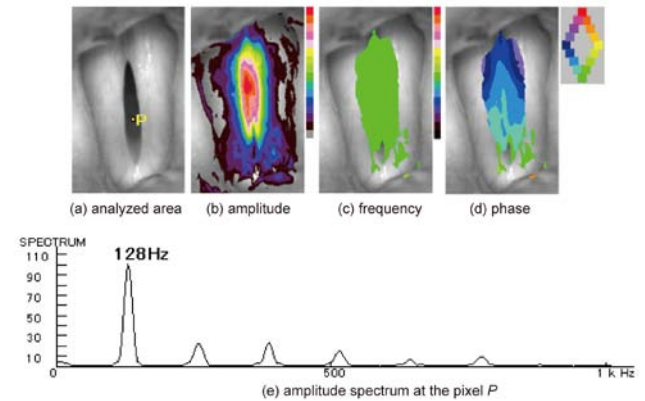


Figure 2: Laryngotopograph of a modal voice of male subject. (a): selected area for analysis. (b): amplitude of the maximum amplitude component. (c): frequency of the maximum amplitude component. (d): phase of the maximum amplitude component. (e): amplitude spectrum of a pixel P in the panel (a).

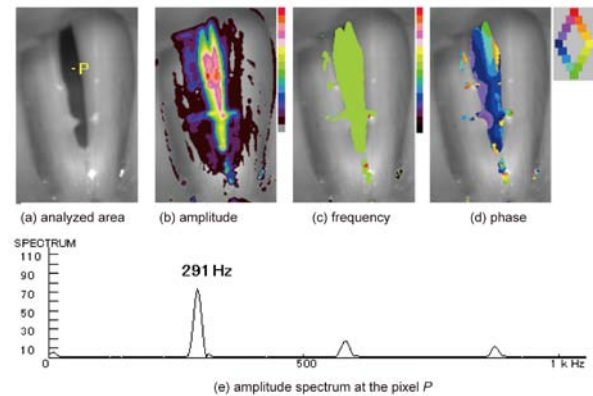


Figure 3: Laryngotopographs of a modal voice of female subject.

2.4.2. Laryngeal nerve paralysis

Figures 4 (a)-(e) are the laryngotopographs of a patient with left vocal fold paralysis, female, age of 22, perceived as diplophonia (patient A). Figures 4 (a) and (b) show a selected area for analysis and distribution of the amplitudes of the maximum amplitude components, as in Figures 2 and 3. Figures 4 (c) and (d) show frequencies of maximum and second-maximum amplitude components, respectively, and Figure 4 (e) shows amplitude-ratios of second-maximum amplitude components to the maximum amplitude components.

In Figure 4 (b), the amplitude of the brightness oscillation in the left vocal fold with paralysis is significantly smaller than that of the right vocal fold. In Figures 4 (c) and (d), the frequencies of the maximum and second-maximum amplitude components are permuted. The frequencies of the maximum components of the left and right vocal folds are 361 Hz and 269 Hz, respectively. In addition, the difference of amplitudes between the maximum and second-maximum amplitude

components is significantly large in the left vocal fold with paralysis. Figures 4 (e) and (f) show the amplitude spectra at pixels *R* and *L* in (a).

Figure 5 shows multi-line kymograph of the patient A. A ratio of the vibratory frequencies of the left vocal fold to the right vocal fold observed in Figure 5 is 3 : 4. This ratio well corresponds with the observation of the laryngotopographs. The difference of the frequencies of two maximum amplitude components is 92 Hz, thus, the estimated fundamental frequency of this case is 92 Hz. From the estimated fundamental frequency, the third and fourth harmonic components to the fundamental are estimated to be the components of 269 Hz and 362 Hz, respectively.

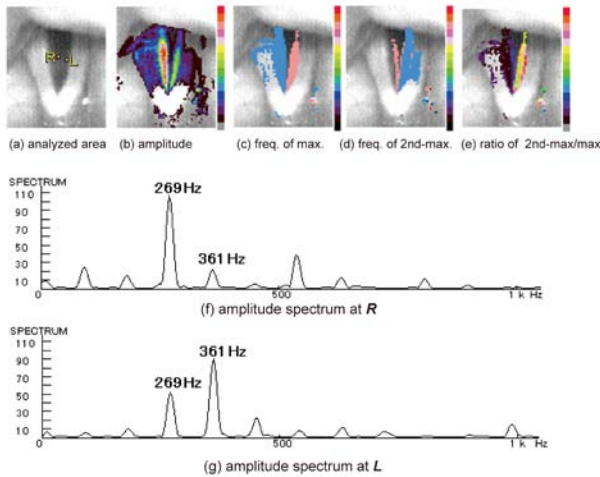


Figure 4: Laryngotopographs of patient with left vocal fold paralysis.

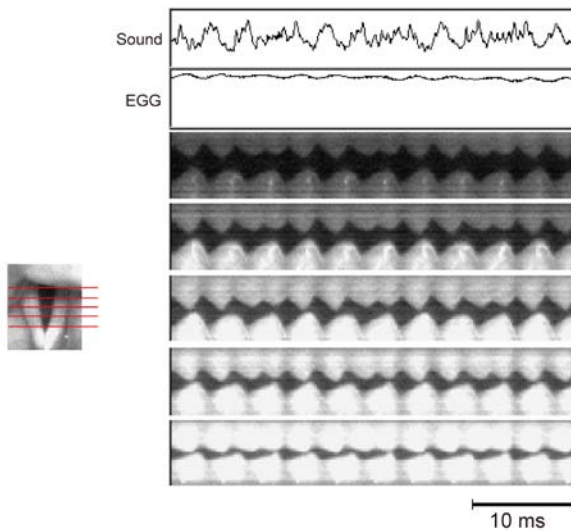


Figure 5: Multi-line kymograph of patient with left vocal fold paralysis. Upper is right and lower is left.

Figures 6 (a)-(e) are the laryngotopographs of a patient with left associated laryngeal paralysis, female, age of 57, perceived as diplophonia (patient B). Figures 6 (a) and (b) show a selected area for analysis and distribution of the amplitudes of the maximum amplitude components. Figures 6 (c) and (d) show frequencies of maximum and second-maximum amplitude components, respectively, and Figure 6 (e) shows

an amplitude-ratio of second-maximum amplitude component to the maximum amplitude component. In Figure 6 (c), the frequency of the maximum amplitude components is almost uniformly distributed around 357 Hz. In Figure 6 (d), the second-maximum amplitude components take two different frequencies, 445 Hz at the right and posterior portion of the left and 530 Hz at anterior portion of the left. In Figure 6 (e), the amplitude of the second-maximum amplitude components at the left vocal fold is relatively larger than that of the right vocal fold.

Figure 7 shows multi-line kymograph of the patient B. A ratio of the vibratory frequencies of the left vocal fold to the right vocal fold observed in Figure 7 is 5 : 4. The difference between is 88 Hz. The estimated fundamental frequency of this case is 88 Hz, therefore the fourth and fifth harmonic components to the fundamental are estimated as the component of 357 Hz and 445 Hz, respectively.

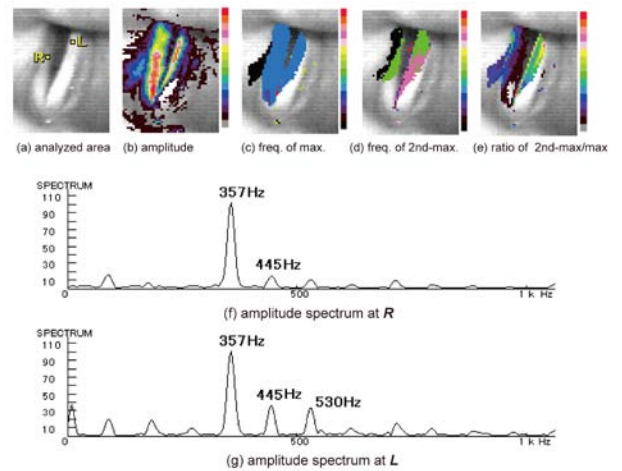


Figure 6: Laryngotopographs of patient with left vocal fold paralysis.

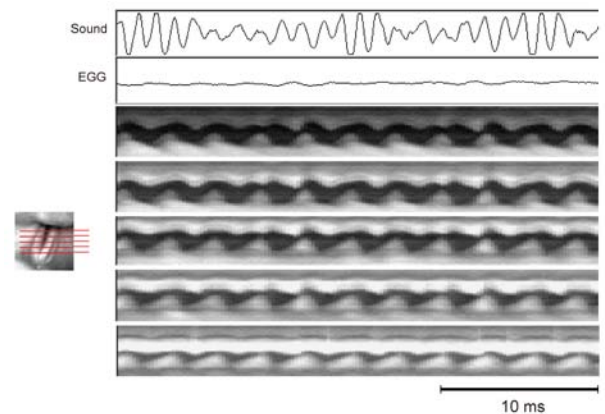


Figure 7: Multi-line kymograph of patient with left vocal fold paralysis.

2.4.3. Vocal fold cyst

Figures 8 (a)-(e) are the laryngotopographs of a patient with the cyst of the right vocal fold, female, age of 70, perceived as diplophonia (patient C). All panels in Figure 8 are similar to those in Figures 4 and 6.

In Figure 8 (a), the amplitude of an oval portion of the cyst in the middle of the right vocal fold is significantly small. The portion of the cyst is considered to be solid and heavy, which is likely to prevent mucosal movement.

In Figure 8 (c) and (d), the frequencies of the maximum and second-maximum amplitude components are permuted. The frequencies of the maximum amplitude components are 251 Hz at the anterior and posterior ends and 313 Hz in the middle portion. In Figure 8 (e), the amplitude-ratio of second-maximum amplitude component to the maximum amplitude component increases as the distance from the cyst increases.

Figure 9 shows distribution of amplitudes and phases of components with the selected frequencies, 251 Hz and 313 Hz. Phases of the component of 251 Hz is uniformly equal to 90 degrees (observed in (b)), in contrast, phases of the component of 313 Hz at anterior and posterior portions are opposite to each other with regards to the cyst border. Namely, the phase difference between anterior and posterior portions is almost 180 degrees.

In terms of Titze and Talkin [2], mode (1, 0) of 251 Hz is considered as longitudinal first-mode vibration, and mode (2, 0) of 313 Hz is considered as longitudinal second-mode vibration, and these two modes are superposed.

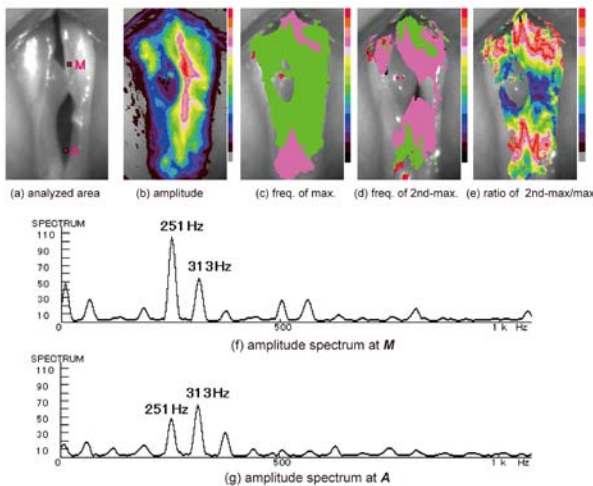


Figure 8: Laryngotopographs of a patient with cyst of the right vocal fold.

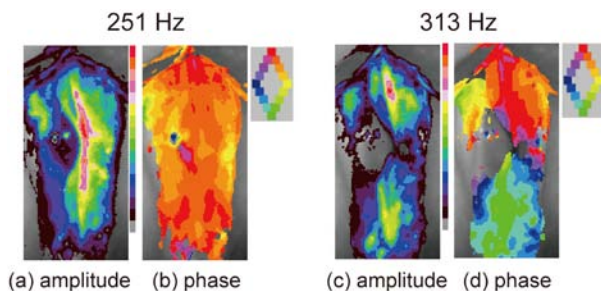


Figure 9: Amplitudes and phases of selected frequencies. (a), (b) are amplitude and phase of the component of 251 Hz, respectively. (c), (d) are amplitude and phase of the component of 313 Hz, respectively.

3. Discussion

In general, it is difficult to observe different modes of oscillation by viewing high-speed imaging movies. None of

the previously used analysis methods, such as glottal area functions and kymograph, is effective for extracting spatial information of vibration, in contrast, the laryngotopography is able to display spatial information related to natural frequencies of a portion of a larynx, amplitude of vibration, and propagation speed of mucosal wave.

By laryngotopography combined with analysis of brightness curves of pixels, intuitive observation of the vibration of the larynx is conducted. In fact, we may be able to speculate from the laryngotopographs of patients that diplophonia can be caused by mechanical superposition of two different vibration modes on the vocal folds. In the case of paralysis, two different modes appeared laterally, and in the case of cyst, two different modes appeared longitudinally. The laryngotopographic observation for the larynx is effective in clinical studies since the result can be clearly displayed as a static image which enables intuitive observation.

On the other hand, we have to consider certain issues in the current laryngotopography technique. All the information obtained from the laryngotopography is pixel-base, not based on a point of interest on the real laryngeal space. Therefore, the maximum amplitude is at the glottal area and the amplitude is relatively small outside the glottal area by the reason that brightness is almost zero on the glottis at open phase. It is necessary to consider the above, not to misinterpret laryngotopographic data. It may be useful to develop some adjustment techniques to associate point of interests of a larynx with pixels, in order to have the laryngotopography combined with an analysis of brightness curves with regards to point of interests instead of pixels. This development will bring us more accurate observation of behaviors of the larynx.

4. Conclusions

The laryngotopographic analysis method for the high-speed images was proposed. It was applied to normal and pathological subjects. The laryngotopographic observations for the larynx are effective in clinical studies.

5. Acknowledgements

The authors thank Svante Granqvist for his valuable discussion about his idea of a pixel-wise Fourier transform. The authors also thank Tatsuya Yamasoba and Takaharu Nito for their helpful discussions and Mika Ito for her valuable comments on this paper. This research was partly supported by Japan and Grant-in-Aid (KAKENHI: 20500161) from the MEXT, Japan, and SCOPE (071705501) of MIC, Japan.

6. References

- [1] S. Kiritani, High-speed digital image recording for observing vocal fold vibration, in *Voice Quality Measurement*, pp. 269—283, R.D. Kent and M. J. Ball Eds., Singure Publ., 2000
- [2] S. Granqvist, P.-Å. Lindestad, A method of applying Fourier analysis to high-speed laryngoscopy, *J. Acoust. Soc. Am.*, 110(6):3193—3197, 2001.
- [3] I. R. Titze, and D. T. Talkin, A theoretical study of the effects of various laryngeal configurations on acoustics of phonation, *J. Acoust. Soc. Am.*, 66(1): 60—74, 1979.