RECENT ADVANCES IN HIGH-SPEED DIGITAL IMAGE RECORDING OF VOCAL CORD VIBRATION

Shigeru Kiritani
Research Institute of Logopedics and Phoniatrics, University of Tokyo, Tokyo, Japan

ABSTRACT

The paper describes recent dvances in the high-speed digital image recording of vocal cord vibration. One is the introduction of a large size image memory which enables data recording of longer duration and thus, observations of involuntary, sporadic changes occuring in certain pathological voices. Another is the development of the system of higher frame rate (4500frames per second with 256 x 256 pixels). Examples of the data analysis conducted by these sytems are presented.

INTRODUCTION

To study voice source characteristics in speech, it is important to record vocal cord vibration simultaneously with the speech signal and to analyze the relationship between the pattern of the vocal cord vibration and the acoustic characteristics of the speech signal. The system of high-speed digital image recording developed by the present authors is convenienet for this kind of studies and the system has been used at our institute for the studies of voice source characteristics in normal speech as well as in pathological voices[1-3].

In order to further fascilitate such studies, several technical improvements were introduced to our original system. The present paper reports on the recent advances in our system.; one is the use of a large size image memory for recording glottal image in longer phonation, and another is a development of a new system with higher frame rate and higher resolution. Characteristics of the improved system together with the examples of the data obtained by this system will be presented below.

Fig. 1 shows a block diagram of the high-speed digital image recording system. The syste consists of an oblique angled solid endoscope, a camera body containing an image sensor, and a digital image memory. The laryngeal image obtained through the endoscope is focused on the image sensor. The image sensor is scanned at a high frame rate and the output video signal is fed into the image memory through a high-speed A/D converter. Stored images are then reproduced consecutively as a slow-speed motion picture. In our original system, we used a commercially available image sensor. In order to achieve a high frame rate, it was necessary to scan only a selected part of the sensor. The frame rate was 2500 per second with the number of pixels 126 x 32.

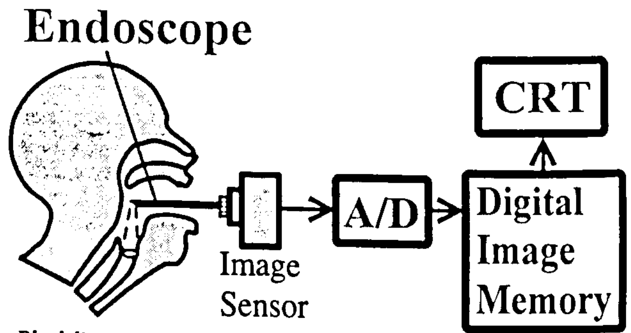


***Figure 1** Blockdiagram of high-speed digital image recording system.*

LARGE-SIZE IMAGE MEMORY –Image Recording of Longer Duration–

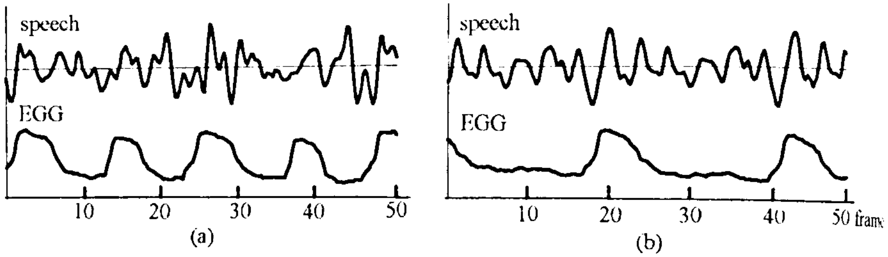
High-Speed observation of vocal cord vibration has generally been conducted for very short periods during sustained phonation, typically a fraction of a second. However, there are several kinds of studies which require data recording of longer duration. One example is the analysis of vocal cord vibration during running speech which includes consonants. For this purpose, a high-speed digital image recording system combined with a flexible fiberscope is now being used at our laboratory. This kind of study requires the recording of laryngeal behavior in the natural utterances for duration of a few seconds.

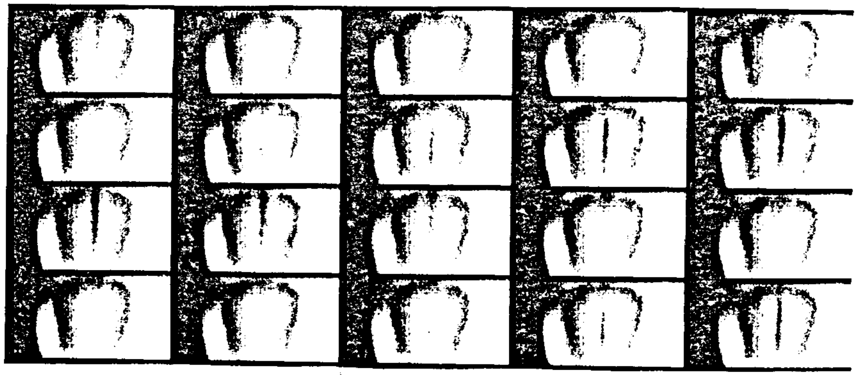
Another example is the analysis of vocal vord vibration associated with sporadic, involuntary voice changes in certain pathological cases. In such studies, recordings of several seconds duration are desired to catch the moments of sporadic changes in the vocal cord vibration such as changes in the pitch frequency.

In order to carry out such studies, a special, large-size didital image memory was constructed. The size of the memory is 64 Mbyte, and it can store 15,000 frames of glottal images with 126 × 32 picture elements. This corresponds to an image recording of 6 seconds at a rate of 2,500 frames per second. Below, two examples of pathological vocal cord vibration are presented which are associated with sporadic, involuntary changes in the fundamental period of the voice occurring during sustained phonations.

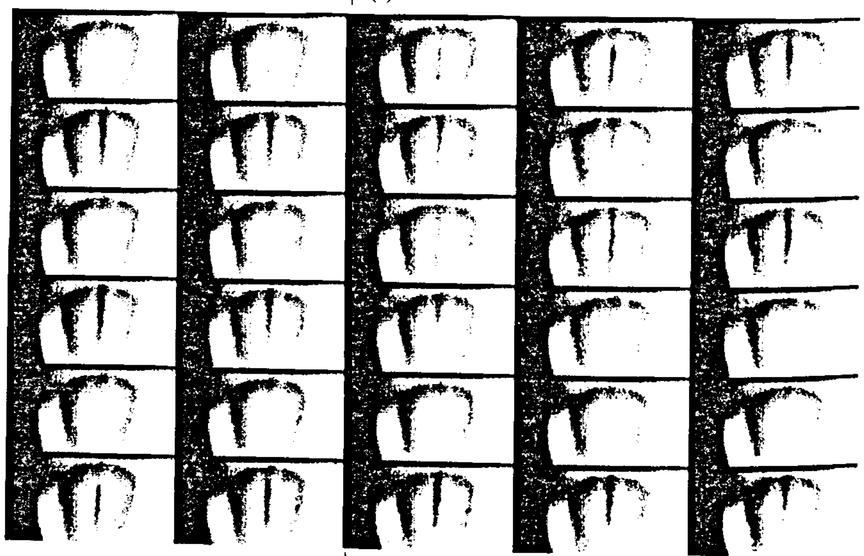
Case 1 shows a rough voice accomparied by sporadic changes in the fundamental period. The subject does not show any apparent pathological change in his vocal cords and he underwent a botulinus toxin injection 4 months prior to the recording for treatment of his spasmodic dysphonia. Figures 2 (a) and (b) show the speech and EGG signals of his voice during the periods of normal pitch and lowered pitch in the same phonation. The pitch period for the lowered pitch is nearly the twice that for the normal pitch. The glottal image during the period of normal pitch shows a clear, tight closed phase in each vibratory cycle. In contrast, the glottal image during the period of lowered pitch shows that the glottal closure is incomplete in every other vibratory cycles, there is no apparent movement of the EGG signal or excitation pattern in the speech wave in these cycles. Thus, it can be concluded that in this phonation, weakening of the closing movements of the vocal cords in every other cycles brings about the apparent doubling of the fundamental period in the speech signal. An a Apparent fundamental period in speech and EGG signals actually corresponds to 2 cycles of vocal cord vibration.

The voice of case 2 is characterized by intermittent cessations in voicing and is accompanied by marked changes in voice quality as well as in pitch frequency. Figures 4 (a) and (b) show the speech and EGG signals for his



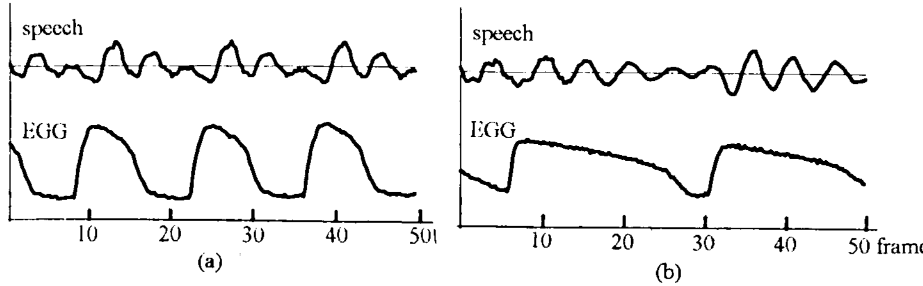


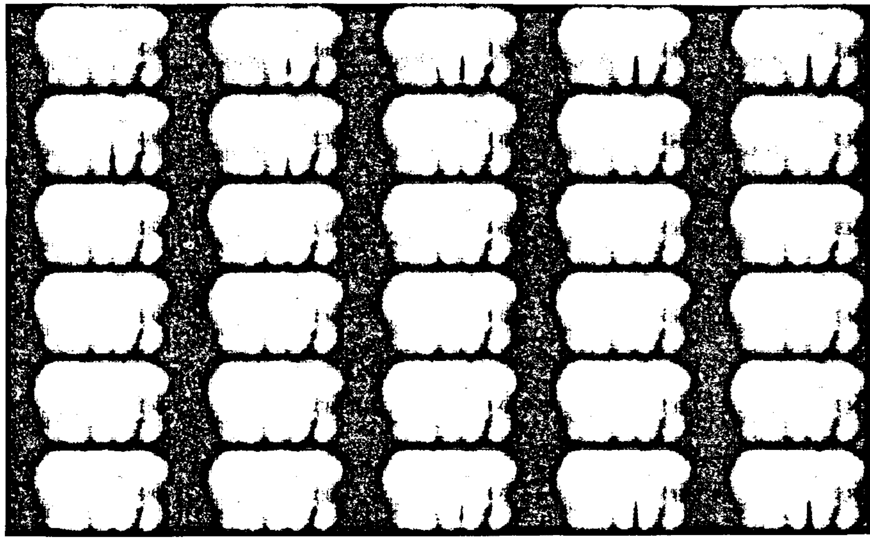
(a) 1-12 frame



(b) 1-30 frame

Figure 2 Sporadic change in the fundamantal period of voice; case 1. (a) Period of normal pitch. (b) Period of lowedred pitch. Frame rate 2500/second.





(b) 1-30 frame

Figure 3 Sporadic change in the fundamantal period of voice; case 2. (a) Period of normal pitch. (b) Period of lowedred pitch. Frame rate 2500/second.

voice during the periods of normal pitch and lowered pitch. In this case also, the fundamental period of the speech wave and the EGG signal for the lowered pitch are about twice those for normal pitch. However, the glottal images confirms that the vocal cord vibration during the period of lowered pitch has a long period of glottal closure which is accompnied by the short period of glottal opening. In this case, an elongation of the fundamental period of the speechwave results from the longer closure period in the vocal cord vibration.

Thus, in the present study, both the voices of case 1 and case 2 show involuntary, sporadic changes in the fundamental period of the speech wave during sustained phonation. In the period of lowered pitch, the pitch period is nearly twice that in normal pitch. However, high-speed recording of the glottal image revealed a characteristic difference between the 2 cases. In case 1, it was due to the weakening of the closing movement in every other vibratory cycle, and in case 2, it was due to a true elongation of the vibratory cycle.

A NEW HIGH-SPEED, HIGH-RESOLUTION SYSTEM

As described above, in our original system, the maximum frame rate was limited to 2,500 frames per second with 126×32 pixels. Recently, a new system with a higher frame rate and higher resolution was developed in cooperation with Photron Co. Ltd. Photron Co. Ltd has produced a specially designed image sensor which incooperates a technique of parallel read-out of image signals to obtain a high frame rate. The sensor contains 256×256 picture elements and can be scanned at a rate of 4,500 frames per second. When the image area is restricted to 256×128 picture elements, the frame rate is 9,000 per second. As an example of data analysis obtained by the new system, an analysis of vocal cord vibration in a simulated diplophonic voice is presented below. The voice was produced by a normal subject simulating a diplophonia.

Figure 4 shows the speech signal in this phonation. The speech signal shows quasi-periodic variations in amplitude and waveform in 9 pitch periods. A waveform with large amplitude and strong excitation is observed in every 9th period. In between these cycles, the speech amplitude gradually gets smaller. In our previous paper, we reported on the chatacteristics of vocal cord vibration in

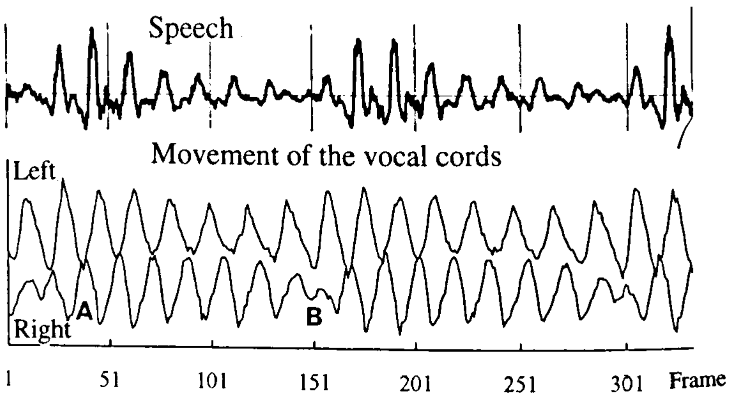


Figure 4 Movements of the vocal cords in a simulated diplophonic voice produced by a normal subject. Frame rate 4500/second.

pathological cases of diplophonia[4]. It was observed that there is a difference in vibratory frequency between the left and right vocal folds, and that the phase difference between the vocal cords varies with time quasi-periodically, resulting in a quasi-periodic variation in the speech signal. However, in that study, due to limitations in image resolution, it was difficult to identify the edges of the vocal cords and to measure the movements of the vocal cords from the glottal images, and only a qualitative measure of the glottal width was presented.

In the present data, owing to the improved image quatily, the edges of the vocal cords can be determined by a visual inspection of the glottal image, at least for this phonation and therefore more detailed information on the movements of the vocal cords can be obtained. Figure 4 (c) shows the result of such measurement. Movement of the edges of the left anf right vocal cords were measured on the selected horizontal scan line shown on the glottal image in the figure. It can be seen in the figure that at around A in the figure, the movements of the vocal cords are nearly in phase, and that the glottis shows a period of complete closure. Then, during successive cycles, the phase difference becomes progressively larger. At around B in the figure,the inward movement of the right vocal fold is incomplete, and one vibratory cycle of the right vocal cord almost disappears, and this process cancels and resets the phase difference between the left and right vocal cords.

The temporal change in the pattern of the vocal fold vibration described above explain the pattern of the temporal change in the speech waveform.

SUMMARY

Recent technical advance in high-speed digital image recording of vocal cord vibration were prresented. Preliminary experiments cerfirmed that data recording of longer duration is useful for observing sporadic phenomena occurring during sustained phonations in certain pathological cases. A new system of higher frame rate made possible more quantitative analysis of movements of vocal cords.

REFERENCES

[1] K. Honda, S. Kiritani, H. Imagawa and H. Hirose: "High-speed Digital Recording of Vocal Fold Vibration Using a Solid-state Image Sensor," in Laryngeal Function in Phonation and Respiration, College-Hill Publication, 485-491, 1987.

[2] H. Imagawa, S. Kiritani and H. Hirose: "High-speed Digital Image Recording System for Observing Vocal Fold Vibration Using an Image Sensor," J. Medical Electronics and Biological Engineering, 25, 284-290, 1987.

[3] S. Kiritani, H. Imagawa and H. Hirose: "Vocal Cord Vibration and Voice Source Characteristics---Observations by the High-speed Digital Image Recording," Proc. ICSLP-90, Kobe, Japan, 61-64, 1990.

[4] S. Kiritani, H. Imagawa and H. Hirose: "High-speed Digital Image Analysis of Vocal Corld Vibration in Diplophonia," Speech Communication, 13, 23-32, 1993.