HIGH-SPEED DIGITAL IMAGE RECORDING FOR THE OBSERVATION OF THE VOCAL CORD VIBRATION

SHIGERU KIRITANI HIROSHI IMAGAWA HAJIME HIROSE

Research Institute of Logopedics and Phoniatrics
University of Tokyo, Bunkyoku 7-3-1, Tokyo, Japan

Abstract

A new method for the high-speed digital image recording of the vocal fold vibration was developped using a solid endoscope and an image sensor. Video signals from the image sensor are digitized and stored in a digital image memory. Stored images are displayed on a monitor oscilloscope. Frame rates of 2000/sec and 4000/sec are realized for the images with 100x37 and 100x17 picture elements, respectively. Compared to the ordinary high-speed motion picture system, the present system is compact and enables flexible data collection.

A similar system using a fiberscop was also developped. Although images by the fiberscope are darker, a frame rate of 2000 per second was achieved for the images with 100 x 17 picture elements. The system makes it possible to observe vocal cord vibrations during the consonants. Preliminary observations on the transitional characteristics of the vocal cord vibrations during the implosion and explosion of the consonants are performed.

Introduction

This paper presents a new technique for high-speed digital image recording for the observation of vocal cord vibration.

For the study of the voice source characteristics, it is essential to record the 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 speech signal. Observation of the vocal cord vibration has generally been performed by using a high-speed motion picture. However, that method requires special equipment and is not suited for flexible data collection under various modes of phonation. For the simultaneous recording of speech signal, special considerations are necessary on the acoustic shielding of the mechanical noises from the high-speed camera.

In the present study, a system of high-speed digital image recording was developed. The sytem is small and compact and, thus, enables flexible data collection. Simultaneous recording of speech and other physiological signals can be performed easily.

Solid endoscope system

Fig.1 shows a block diagram of the system. The system consists of an oblique-angled solid endoscope, a camera body containing an image sensor and an image processor. The output video signal from the image sensor is fed into the image processor through a high-speed A/D converter. Stored images are then displayed on a CRT monitor as an array of small images which represent sequential time frames. It is also possible to display stored images as a slow motion picture. The image processor in the present system contains about 750k byte of image memory. Generally, for one shot of image recording, about 100 frames of image data are sampled and stored by the image processor.

Maximum frame rate that can be realized by the present system is determined by the brightness of the image obtained through the endoscope and the speed of scanning the picture elements in the image sensor.

In order to get a brighter image, a new model of the solid endoscope was constructed. The diameter of the scope was larger than that of the ordinary scope for the clinical use. The cross

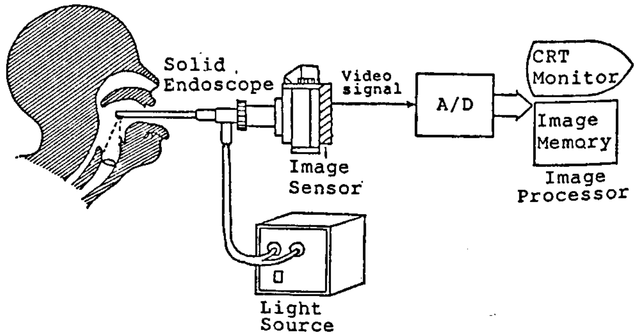


Fig. 1 Blockdiagram of the solid endoscope system.

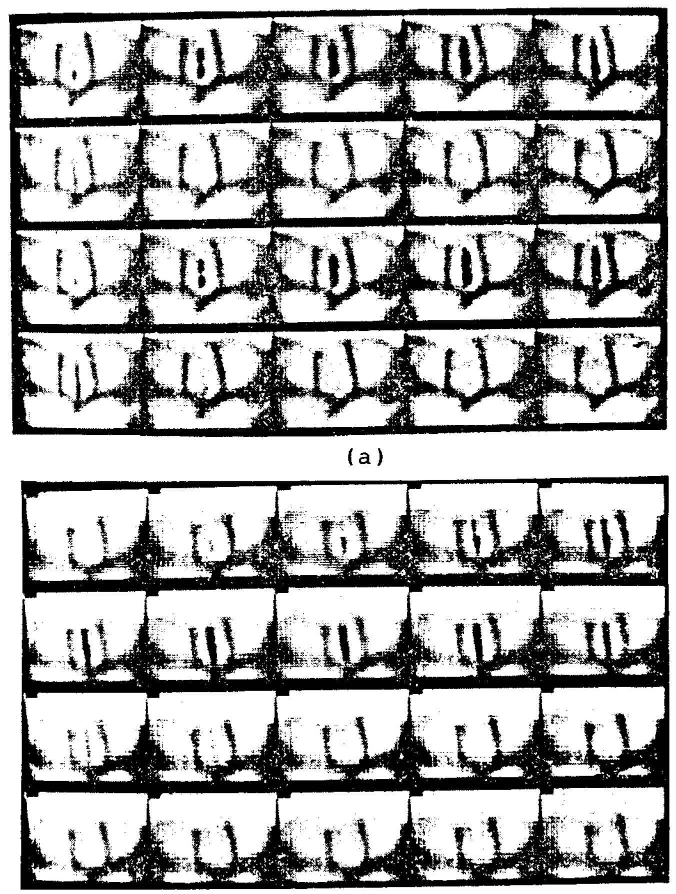
section of the tube of the scope is an ellipse. The image guide is located at the center of the tube and on both sides of the image guide are light guides. The two light guide cables are connected to the separate light sources.

Table 1 summarizes basic characteristics of the present system. Light sources are the two 250W halogen lamps. Number of the picture elements in the image sensor is 100x100 and the sampling rate is 10MHz. When entire picture elements are scanned, the resulting frame rate is about 1000 per second. In order to achieve a higher frame rate, a special scan method was devised in which only the selected scan lines were sampled. When 37 scan lines are sampled out of 100 scan lines, the frame rate is 2000 per second. A higher frame rate of 4000 frames per second is achieved by sampling only 17 scan lines. As far as the brightness of the image is concerned, this frame rate appears to be nearly the maximum that can be achieved for the most subjects. Naturally, the brightness of the image varies, depending on the laryngeal view of the individual subjects. The image memory can store the image data for the period of about 200msecond.

Table 1 Basic characteristics of the endoscope system.

Light source	250W halogen lamp x 2
Image sensor	MOS type
	100x100 picture elements
	Clock 10MHz
Image memory	768k byte

Scan elements	Frame/sec	Storage
100x37	2000	200
100x17	4000	450 (frames)



(a)

(b)

Fig.2 Examples of the recorded image of the vocal cord vibration. (a) 2000 frames /sec (b) 4000 frames /sec

Fig. 2 shows examples of the image recorded at a rate of 200 and 4000 frames per second. The pictures shows the vocal cord vibration of a male subject during the sustained phonation of the vowel /e:/. Fundamental frequency of the voice was about 200Hz. The images of the maximum glottal opening are observed at about every 20 40 frames.

Fiberscope system

It is very valuable if a similar system can be constructed using a fiberscope. Such a system makes it possible to observe the vocal cord vibrations during consonants. In the present study, a pilot system of high-speed image recording using a fiberscope was also developed.

Fig. 3 shows a block diagram of the system. In this system, a video camera is connected to the finder of a single-lens reflex camera to monitor and record the glottal view preceding and following the short period of high speed imaging. The image is sent to the image sensor for high-speed image recording only when the shutter of the camera is open. This monitoring is necessary beacause the fiberscope system mainly aims at the observation of the glottis during the running speech and the period of high-speed imaging is very short.

In order to obtain a brighter image, a new fiberscope was also constructed the

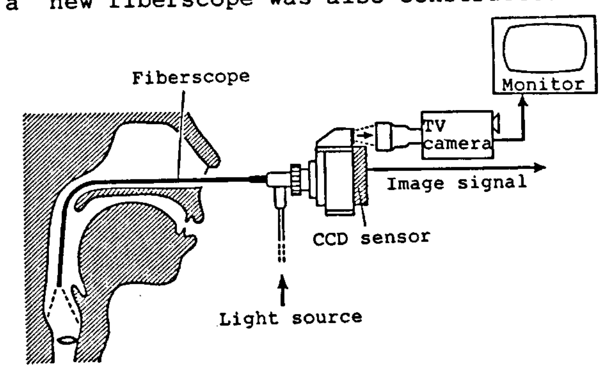


Fig. 2 Blockdiagram of the fiberscope system.

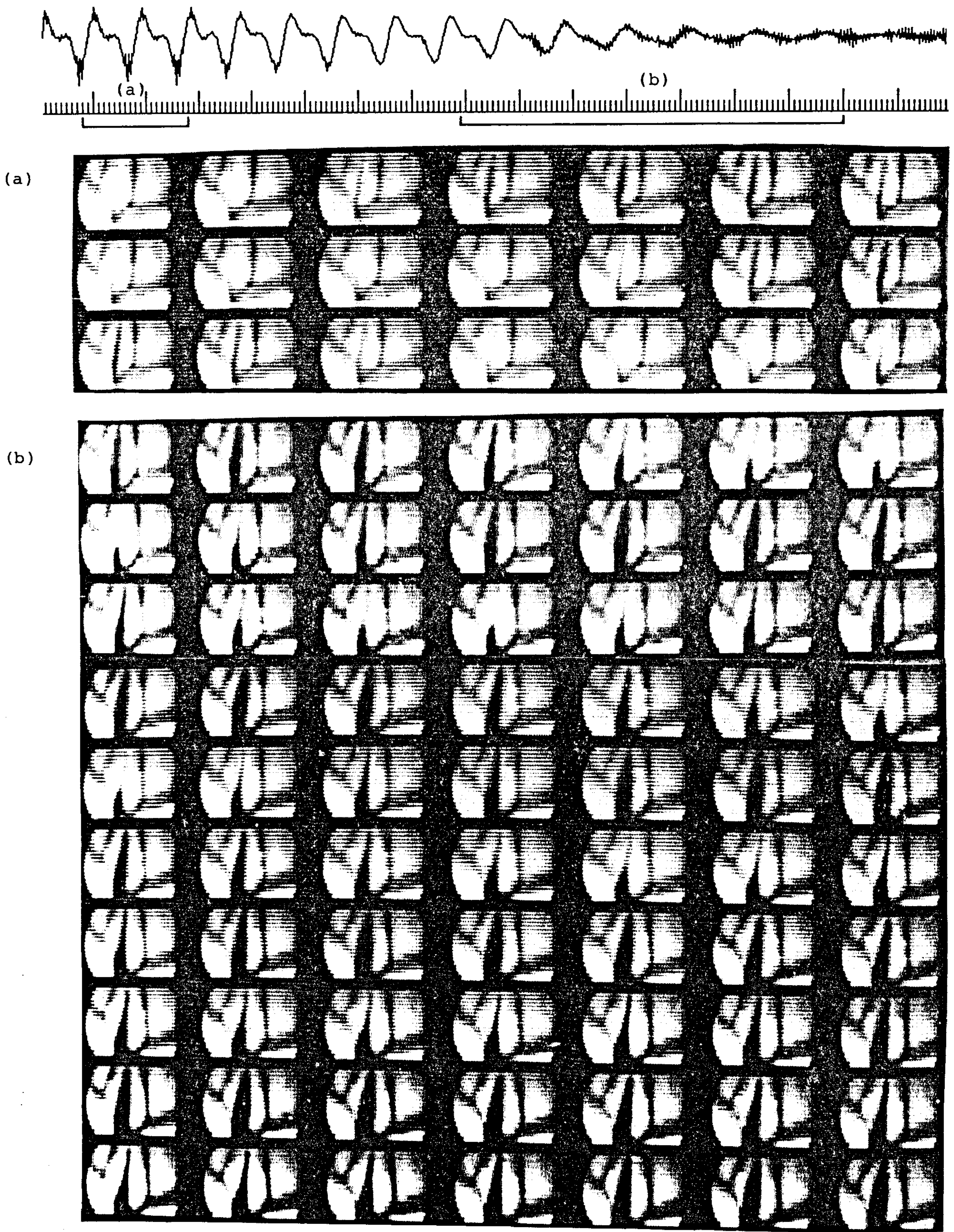


Fig. 4 Vocal cord vibration during transition from vowel to /h/ in the utterance /pi:hi:/.

Table 2 Basic characteristics of the fiberscope system.

Fiberscope	diameter 4.8mm view angle 43_ distance 7-70mm
Light source	500W xenon lamp
Image sensor	CCD type Clock 10MHz
Image memory	768k byte

Scan elements	Frame/sec	Storage
200x34	1000	110
200x14	2000	270 (frames)

diameter of which was slightly larger than that of the ordinary scope. At the same time, a CCD type image sensor was employed in this system, because the image by the new fiberscope was still darker than that by the solid endoscope,. The sensitivity of the CCD image sensor is generally higher than that of the MOS type imagr sensor which was used in the solid endoscope system. However, the comercially available CCD sensors generally contain a large number of picture elememts, 500x500 for example. Thus, in order to realize a high frame rate, it was necessary to develop a special scan method to sample only a very limited portion of the image sensor and to reduce the number of sampled elements.

Table 2 summarizes basic characteristics of the system. The light source is a 300W xenon lamp. The sampling rate of the picture elements is 10Mhz. The period of image recording is about 100ms long. For the picture elemnts of 200x34, the frame rate is 1000 per second. A frame rate of 2000 per second can be achieved with the picture elements of 200x14

By using this system, preliminary tests on the recroding of vocal cord vibrations during consonants were performed. A special triggering method was employed to record the glottal images for the selected period of the consonant in the VCV utterances. First a pulse from the camera shutter sets the entire recording circuit ready. Then, the subject start the utterance. When the beginning of the speech envelope is detected, a trigger pulse is generated. Actual sampling of image signal is started with a delay of the specified interval. By using an appropriate delay time, it is possible to record the glottal images during the selected consonantal period in the VCV utterances.

Fig. 4 shows an example of images recorded by the fiberscope system at a rate of 2000 frames per second. Vocal cord vibration during the transition from the vowel [i] to the consonant [h] in the utterance [pi:hi:] is shown. Fig. 4 (a) represents the stationary vibration during [i:]. During the transition period shown in Fig. 4 (b), the arytenoid cartlidges gradually seperate and the maximum glottal opening is getting larger. At the later period in the figure, the vocal cords are still vibrating but the right and left vocal cords do not contact. It can be seen in the speech wave that, corrsponding to the observed vocal cord vibration, there is a modulation of the amplitude of the noise in the speech signal.

Summary

We have developed a new method of high-speed digital image recording system using laryngeal endoscopes and image sensors. The system is compact and simultaneous recording of speech and other physiological signals can be performed very easily. The system using a fiberscope realized observation of the vocal cord vibrations during consonants. We believe that the system is useful for the study of the voice source characteristics in various mode of phonation.

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, Acollege-Hill Publication, 485 (1987).

2) S. Kiratani, K. Honda, H. Imagawa and H. Hirose: "Simultaneous high-speed digital recording of vocal fold vibration and speech signal.",Proc. ICASSP, Tokyo, 1633 (1986).