Ann.Bull.RILP(1986)No.20, 17-22

IMPROVEMENTS IN THE HIGH-SPEED DIGITAL IMAGE RECORDING SYSTEM FOR OBSERVING VOCAL FOLD VIBRATION

Hiroshi Imagawa, Shigeru Kiritani, Kiyoshi Honda
and Hajime Hirose

I. Introduction

In order to facilitate high-speed recording for images of vocal fold vibration, we have been developing a high-speed digital image recording system using a solid endoscope and a solid-state image sensor. Figure 1 shows a block diagram of the system. The system consists of a solid endoscope with an oblique-view angle, a camera body containing an image sensor and an image processor. In our last Annual Bulletin RILP (1985), we reported on the prototype of the system. Since then, we have made some improvements in the system to attain a higher frame rate. In this report, the characteristics of the improved system are described.

It will be very valuable if the present technique can be applied to laryngeal observations using a flexible fiberscope. Such a system will make it possible to observe vocal fold vibration during the production of various speech sounds. A preliminary test on the application of the present technique to the fiberscope was performed. The present paper also describes the results of this preliminary test.

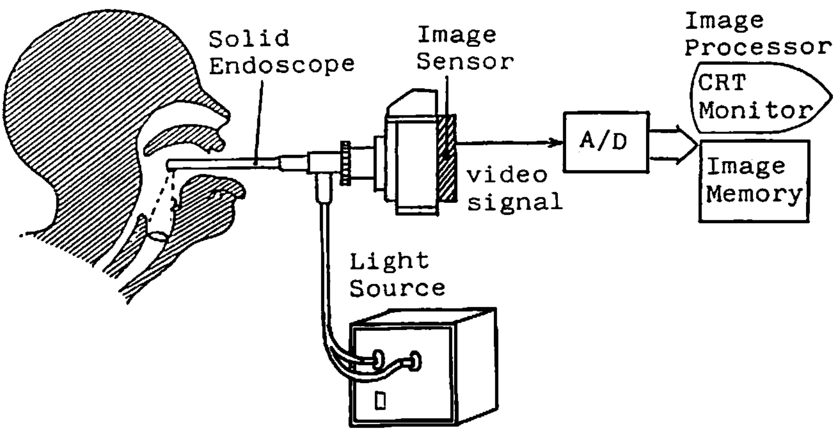


Figure 1. Digital image recording system using a solid endoscope.

II. Improvements in the system using a solid endoscope

The major points of the improvement are as follow.

1) A new endoscope was constructed in which the diameters of the image-guide and the light-guide were enlarged compared to the solid endoscope used before, so that a brighter image could be obtained. The fibers in the light-guide are divided into two bundles. Each of the bundles is connected to a light source with a 250-watt halogen lamp. Consequently, the brightness of the image from the new endoscope is three times as high as that from the endoscope used before.

2) The scan rate of the picture elements has been increased from 1.25 MHz to 2.6 MHz. As a result, for 45 X 50-element images, the frame rate was 1000 per second. A higher frame rate can be obtained by sampling only a selected number of scan-lines out of the entire 50 scan-lines in the image sensor. When the picture elements in an image is reduced to 45 X 25, the frame rate is 2000 per second. When the picture elements in an image is reduced to 45 X 10, the frame rate is 5000 per second (Table 1).

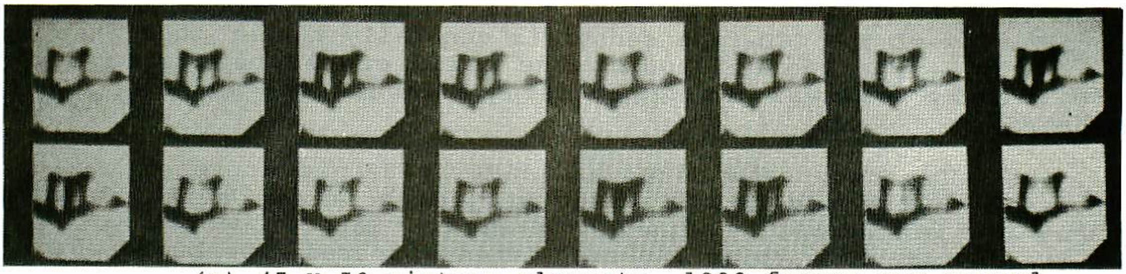
Table 1. Relationship between the number of picture elements and the frame rate in the system using a solid endoscope.

Picture Element	Frame Rate (/sec)
45 X 50	1000
45 X 25	2000
45 X 10	5000

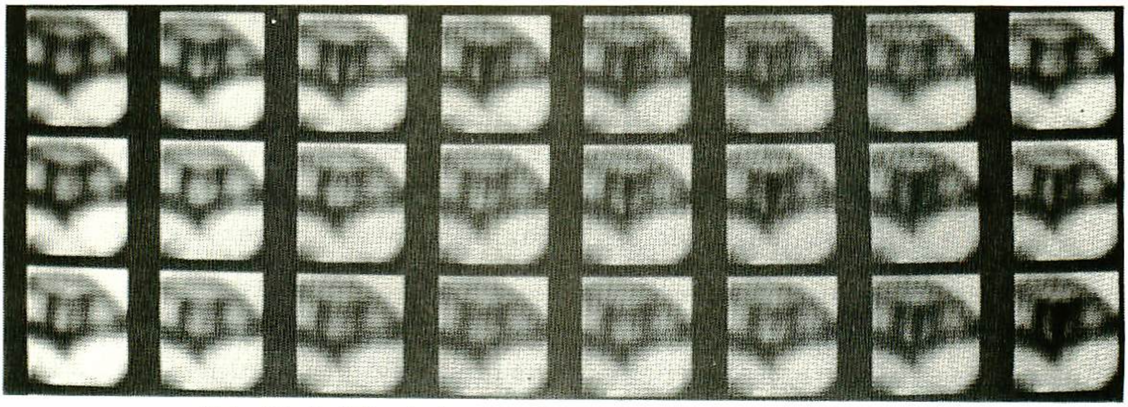
Figure 2 contains photographs of part of the CRT screen displaying an array of images, which were recorded at the three different frame rates given in Table 1. These examples show the glottis of a male subject during a sustained phonation of the vowel /e/. The fundamental frequency of the voice was about 200 Hz. Consequently, in Figure 2(a), the maximum opening of the glottis is seen at every 5th frame. In Figure 2(c), the frame rate is 5000 per second, and one cycle of the vocal fold vibration corresponds to 27 frames. An image recording at this frame rate is maximum with present light source.

III. Preliminary test of the system using a fiberscope

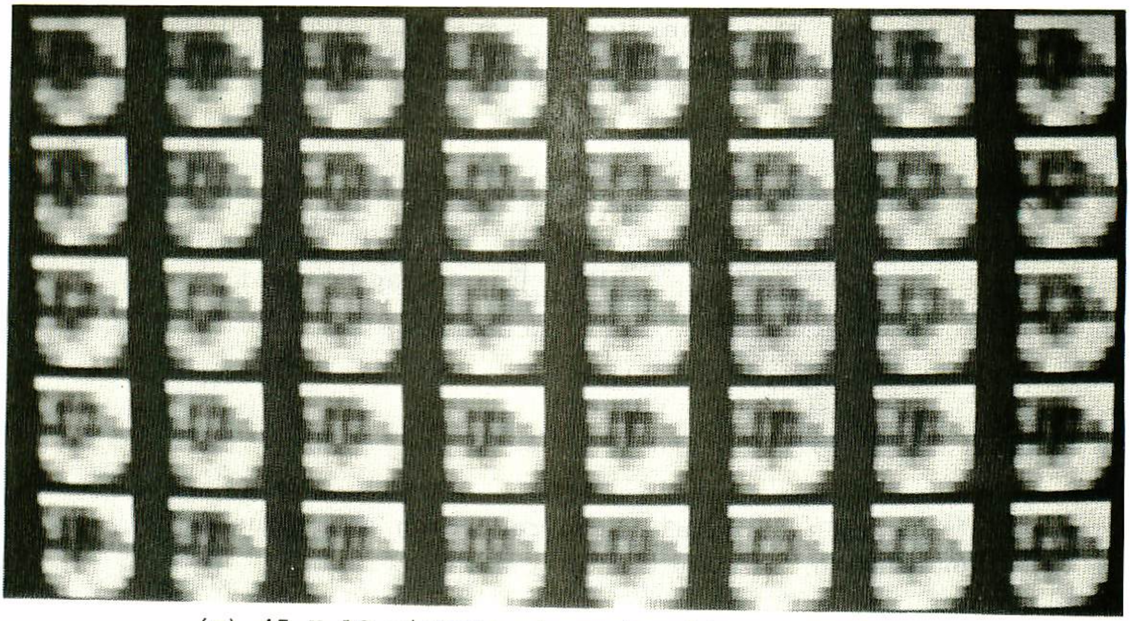
The images that can be obtained by a fiberscope are darker than those obtained by a solid endoscope, because the diameter of the scope to be inserted through the nose is limited. Thus, in order to apply the present method to the fiberscope, it is



(a) 45 X 50 picture elements, 1000 frames per second.



(b) 45 X 25 picture elements, 2000 frames per second.



(c) 45 X 10 picture elements, 5000 frames per second.

Figure 2. CRT display of the high-speed vocal fold images from the solid endoscope.

essential to use an image sensor of higher sensitivity. The sensitivity of the CCD-type sensor is generally higher than that of the MOS-type sensor used in the solid endoscope system. A preliminary test was performed using a fiberscope and the CCD sensor.

The CCD sensor used here had 376 X 488 picture elements. The size of the fiberscope image projected on the sensor was about half of the active area of the sensor. Thus, a central area of the sensor containing 240 scan-lines was used. The number of picture elements in one scan-line of the CCD sensor was greater than that for the MOS sensor used in the solid endoscope system. In order to realize a high frame rate, the number of scan-lines in an image had to be less than that in the solid endoscope system.

In the present system, the maximum scan rate of the picture elements is 10 MHz. Table 2 shows the relationship between the number of scan-lines in an image and the frame rate. The CCD sensor was driven in the following way. When the number of scan-lines is to be reduced to 23, for example, a series of 10 vertical transfer pulses are fed into the sensor, and the signal charges in the ten horizontal rows are accumulated in the horizontal shift register. Then, horizontal transfer pulses are sent to the horizontal shift register to read out the signals. Consequently, the effective sensitivity per scan-line is increased through accumulation of the signal charges.

Figure 3 shows the experimental set-up using a fiberscope and a CCD sensor. The fiberscope image was obtained with a TV camera, through the finder of a reflex camera, and was observed on a TV monitor.

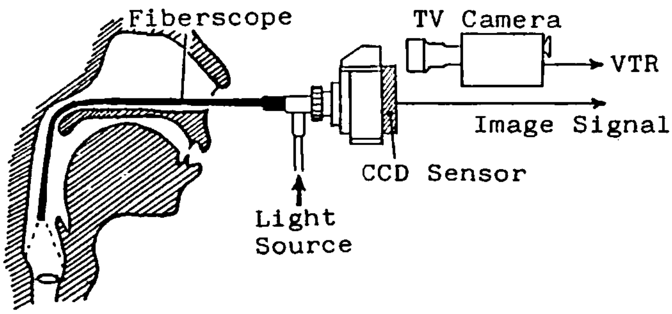


Figure 3. Digital image recording system using a fiberscope.

Table 2. Relationship between the number of pictur elements and the frame rate in the system using a fiberscope.

Picture Element	Frame Rate (/sec)
376 X 23	500
376 X 10	1000
376 X 4	1800

Figure 4 shows examples of images, which were recorded at the three different frame rates given in Table 2. A 300-watt xenon lamp was used as a light source in this recording. A series of glottal images was recorded during a sustained phonation of the vowel /i/ by a male subject. The fundamental frequency of the voice was about 200 Hz.

Since the image in Figure 4(c) consists of only 4 scan-lines, it is difficult to observe the details of the larynx. Figure 5(a) shows one of the CCD images and the VTR image of the glottis taken at the beginning of the high-speed image recording in Figure 4(c). Comparing the VTR and the CCD image, it can be seen that the upper two scan-lines of the CCD image are scanning the glottal area. Figure 5(b) shows the brightness pattern of the second scan-line from the top. The curves for the successive 22 frames are shown. The valley in the central part of the curve corresponds to the dark image of the glottis. The temporal changes in the brightness patterns associated the glottal movements can be observed in this figure.

The essential part of the system which stores and displays the high-speed images can be constructed as compact equipment using a small personal computer combined with an appropriate digital image memory. Such equipment will be useful for the examination of vocal fold vibration in the clinic. Construction of such a system is now being attempted.

References

Honda K., S. Kiritani, H. Imagawa, H. Hirose and K. Hashimoto (1985); High-speed digital recording of vocal fold vibration using a solid-state image sensor. Ann. Bull. RILP, 19, 47-53.

Kiritani S., K. Honda, H. Imagawa and H. Hirose (1986); Simultaneous high-speed digital recording of vocal fold vibration and speech signal. Proceedings ICASSP 86, TOKYO, Vol 3, 1633-1636.