Ann.Bull.RILP(1987)No.21, 9-23

FURTHER DEVELOPMENT IN HIGH-SPEED DIGITAL IMAGE RECORDING SYSTEM FOR ASSESSMENT OF VOCAL CORD VIBRATION

--WITH SPECIAL REFERENCE TO ITS CLINICAL APPLICATION--

Hiroshi Imagawa, Shigeru Kiritani and Hajime Hirose

1. Introduction

We have been developing a high-speed digital image recording system for observing vocal cord vibrations using a laryngeal endoscope and an image sensor. We have already reported on the development of the system in our Annual Bulletins RILP No.19 (1985) and No.20 (1986). This paper presents several improvements introduced in the system during the past year. Figure 1 shows a block-diagram of the system. The major points of the new improvements are the following.

1) With respect to the system using a solid endoscope, the image sensor with 50x50 picture elements was replaced by a new image sensor with 100x100 picture elements.

2) With respect to the system using a fiberscope, a specially designed fiberscope was constructed to obtain a brighter image. The diameter of this scope is slightly larger than that of the scope in ordinary clinical use. At the same time, a new method for scanning the picture elements in the CCD sensor was devised to realize higher frame rates.

3) With the aim of applying the present technique to the observation of pathological vocal cord vibration in practical clinical situations, a compact special-purpose system for high-speed digital image recording was developed.

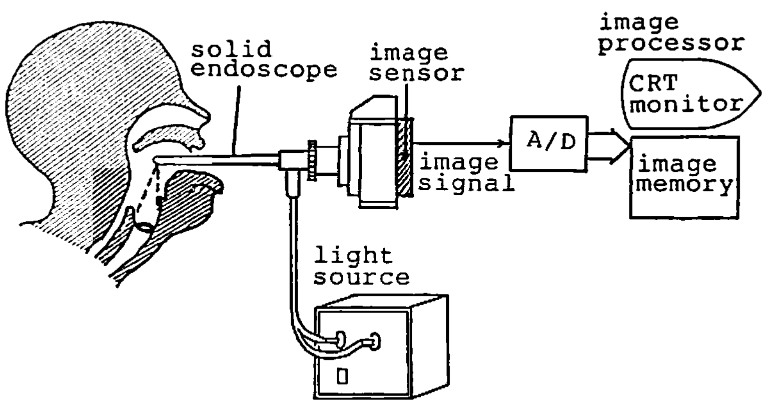


Figure 1. Block-diagram of high-speed digital image recording system using a solid endoscope.

2. The solid endoscope system

Table 1 shows the principal characteristics of the present system. In our previous system, an MOS type image sensor with 50x50 picture elements was used. In the current system, a new sensor with 100x100 picture elements is used, and the resolution of the image is improved. The basic scan rate of the picture elements in the sensor is 10 MHz, and the frame rate for images with all 100x100 elements is 800 per second. In this sensor, an additional waiting time is necessary while the horizontal scan is switched from one scan-line to the next scan-line. As was described in our previous report, in our system a higher frame rate for the practical observation of vocal cord vibration is realized by sampling only a selected number of scan-lines out of all the scan-lines in the image sensor. With the new sensor, when 36 scan-lines are sampled out of 100 scan-lines, the resulting frame rate is 2000 per second. When 17 scan-lines are sampled, the resulting frame rate is 4000 per second. The size of the memory in the image memory is 768 kilobytes. Consequently, when the frame rate is 2000 per second, successive image data of 200 frames can be stored at one time. When the frame rate is 4000 per second, successive image data of 410 frames can be stored at one time.

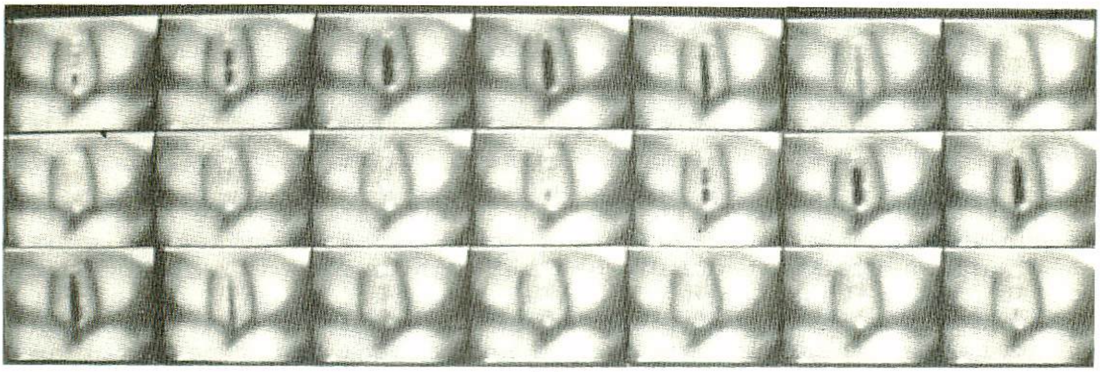
Figure 2 shows examples of images recorded at 2000 and 4000 frames per second. The photographs show the glottis of a male subject during sustained phonation of the vowel /e/. The fundamental frequency of the voice was about 200 Hz. In figure 2(a), the frame rate is 2000 per second, and one cycle of the

Table 1. Principal characteristics of the high-speed digital image recording system using a solid endoscope.

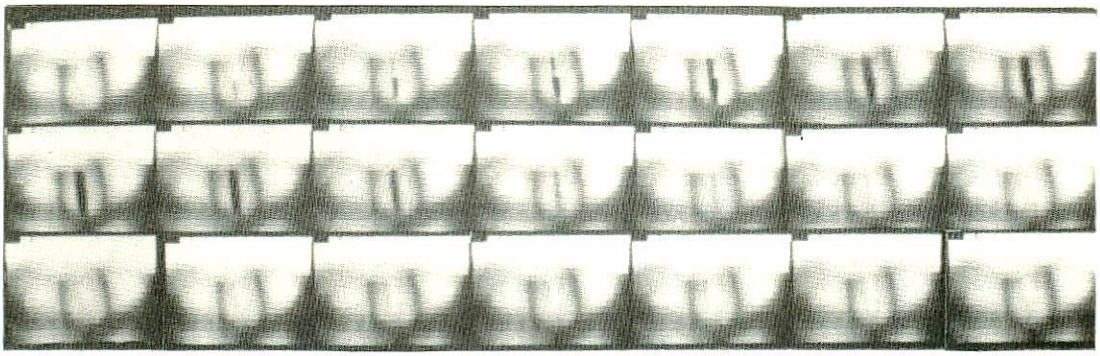
Light source : 250W halogen lamp x2
Image sensor : MOS type
100x100 picture elements
Clock 10MHz
Image memory : 768k bytes

Scan elements	Frames/sec	Storage
100x36	2000	200
100x17	4000	410 (frames)

vocal cord vibration corresponds to 10 frames. In figure 2(b), the frame rate is 4000 per second, and one cycle of the vocal cord vibration corresponds to 20 frames. Naturally, in figure 2(b) the resolution of the image in the vertical direction is worse than that in figure 2(a). However, this kind of image is still useful for observing the patterns of the opening and closing movements of the glottis. Preliminary experiments using the present system have confirmed that for most subjects vocal cord vibration can be recorded at a frame rate of 2000 per second or 4000 per second.



(a)



(b)

Figure 2. Examples of the images recorded by a solid endoscope. Male subject, sustained phonation of /e/.
(a) 2000 frames/second (100X36 picture elements).
(b) 4000 frames/second (100X17 picture elements).

3. The fiberscope system

In order to obtain a brighter image, a new, specially designed fiberscope was constructed. In the case of commercially available laryngeal fiberscopes for general clinical use, the outer diameter of the scope is between 3.2 mm and 4.2 mm. The outer diameter of our new fiberscope is 4.8 mm, and the view angle is 43 degrees. The distance between the tip of the scope and the vocal folds is generally between 2 cm and 4cm. With this new fiberscope, the brightness of the image is approximately 5 times as high as with our previous fiberscope.

In the fiberscope system, a CCD sensor which has 376x488 picture elements is used, and the scan rate of the picture elements is 10 MHz. As in the case of the solid endoscope system, a higher frame rate is achieved by sampling only selected scan-lines. With the fiberscope system, a further special technique was introduced in the scanning of the CCD sensor in order to speed up the frame rate. It appears that, for the current purpose of observing glottal movement, it may not be necessary to use all 376 picture elements on each scan-line. Thus, to save scan time per scan-line, it was decided to read out signals for only 200 out of 376 picture elements. In the case of the CCD sensor, the signal charges on one scan-line are transferred to the horizontal shift register, and then the signal charge for each picture element is read out one by one through horizontal transfer by shift pulses. Thus, when the signal charges for only 200 picture elements are read out, the signal charges of the remaining picture elements are left in the horizontal shift register and will be superimposed on those of the next scan-line. In order to avoid an artifact due to this superimposition, half of the sensor surface was covered with an aluminum sheet to block input light and to avoid the generation of signal charges.

Table 2. Principal characteristics of the high-speed digital image recording system using a fiberscope.

Fiber scope :	Diameter 4.8 mm View angle 43° Distance 7-70 mm
Light source :	300W xenon lamp
Image sensor :	CCD type Clock 10MHz
Image memory :	768k bytes

Scan elements	Frames/sec	Storage
200x34	1000	100
200x14	2000	234 (frames)

level reached a prespecified threshold, the generation of the sampling pulses was triggered. The generation of the sampling pulse was actually initiated with a prespecified delay time to record the selected portion within the uttarance. The threshold level for the speech envelope and the value of the delay time were selected according to the type of utterances to be examined.

Figure 5(a) shows the vocal cord vibration during the /i/ to /p/ transition in /pi:pi:/. The speech waveform is shown at the top of the figure, and the glottal images are shown for the time period which is indicated by the horizontal bar below the speech waveform. In this example, there are 7 closed periods before the vocal cord vibration ceases and the glottis becomes open for the voiceless consonant /p/. It can be seen in the figure, that, during this transition period, the maximum opening of the glottis is becoming larger and the closed period is becoming shorter, indicating an abductive movement of the glottis. Corresponding to this, the excitation in the speech signal becomes weaker and the high frequency component in the speech waveform becomes less. It appears that at the 7th closed period, the closure is not complete (frame No.71). At the next period, there is a closing movement of the glottis, but the vocal cords do not show any contact even at the time of the minimum opening of the glottis (frame No.85). No excitation was observed in the speech signal which corresponded to this vibration.

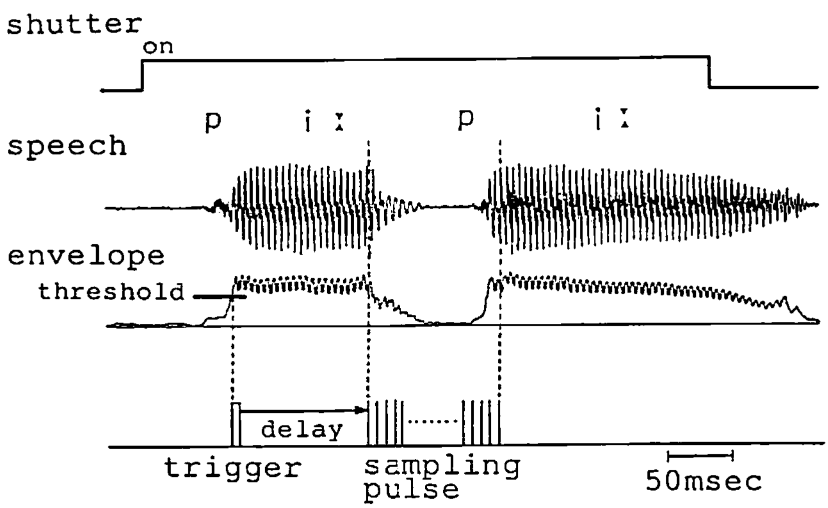


Figure 4. A method for recording a selected period during the utterance.

Figure 5(b) shows the vocal cord vibration for /b/ in /pi:bi:/. It can be seen in the figure that the vocal cord vibration is maintained throughout the closue period of /b/. It should be noted here that in these vibrations there exists a clean closure period and there is no observable abductive movement of the glottis, although the maximum opening of the glottis is wider and the vibratory cycle is longer compared to those during the vowel production. In the same figure, the speech waveform during the closure period of /b/ is more pulsative than that during the implosion of /p/. This fact may correspond to the existence of a clean closure period in the glottal vibration for /b/.

Figure 5(c) shows glottal images for the /i/ to /h/ transition in /pi:hi:/. Corresponding to the abductive movement of the glottis, the maximum opening of the glottis becomes larger and the excitation in the speech signal becomes weaker. At the 8th closed period (frame No.69), the posterior part of the glottis does not close and remains open even at the time of the minimum abductive opening of the glottis. After this, the abductive movement of the glottis further progresses and the separation of the vocal cords at the posterior part becomes larger. However, the vibration is still maintained, and at the time of the minimum opening of the glottis, the anterior part of the vocal cords show contact. This type of vibration with incomplete closure continues until near the last frame shown in the figure. It is to be noted that the above type of vocal cord vibration is associated with the generation of the noise component in the speech signal. In the speech waveform, the noise is superimposed on the periodic component. It appears that the amplitude of the noise is modulated nearly cyclically, and that the power of the noise is greater when the glottis is narrowed. Even during the period following the frames shown in the figure, the vocal cord vibration still continues with a wider separation at the posterior part. However, the anterior part of the vocal cords no longer shows contact, and, for this period, there is no appreciable excitation in the speech signal. The generation of the noise component, whose intensity is modulated and which is superimposed on the cyclic excitation, is also observed during the offset of the /h/. This type of noise generation is not observed for the release of /p/ or /ʃ/ (Figure 6).

The above examples show that the present technique of glottal imaging provides an effective means for analyzing the process of source generation in these consonants. Simultaneous recordings of pressure-flow signals will also be very useful for this kind of analysis.

compeleted, the computer starts a slow motion display of the image data. The display is produced by using a commercially available video image I/O board. The board contains a 256x256 bytes video RAM and an 8 bits D/A converter, and it generates an ordinary video signal. The images are displayed on a TV monitor. It is also possible to record the images on an ordinary video recorder.

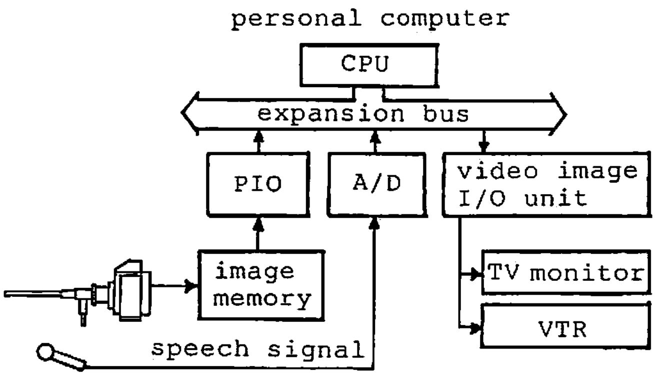
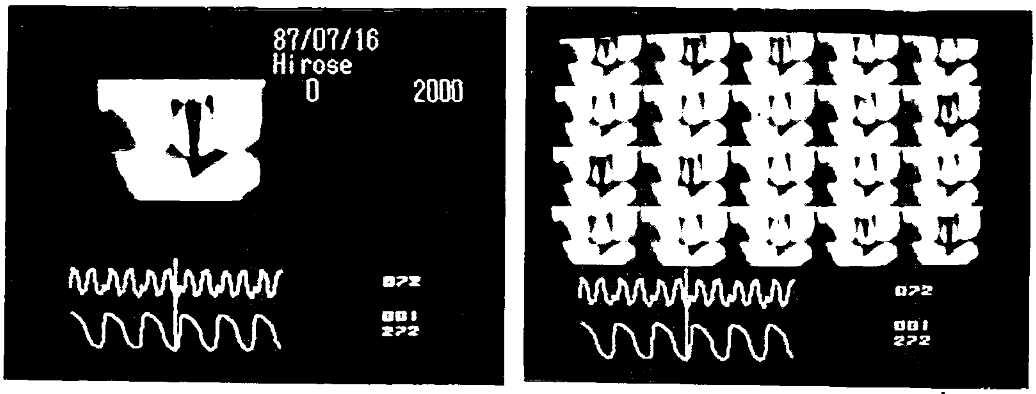


Figure 7. A compact vocal cord image recording system developed as a practical clinical unit.



(a) (b)

Figure 8. Example of a display of a vocal cord image on an ordinary TV monitor.
(a) Slow motion display mode.
(b) Special stop-motion display mode.

In the present system, the basic mode is a slow motion display of the stored images. Figure 8(a) shows a picture of the monitor screen during the slow motion display. Below the glottal image is displayed the waveforms of speech signal and EGG signal which are sampled by a personal computer during the high-speed image recording. The waveforms are scrolled synchronously with the image display. Waveforms over the time interval of 25.6 msecond are displayed, and the vertical line cursor shows the moment corresponding to the time frame of the currently displayed image. In the present simple system, a slow motion display at a rate of 9 frames per second can be realized. In addition to the slow motion display, the system provides a special stop-motion display in which the glottal images of successive frames are displayed as an array of small-size images (Figure 8(b)).

During the slow motion display, the keyboad input is always monitored, and when there is keyboad input, the corresponding control is executed. The display modes and the speed during the slow motion display can be controlled through simple operation of the function keys on the keyboad. Table 3 lists the functions of the keys used for the control of the display. Five keys from "f.1" to "f.5" control the speed of the slow motion display. The key "f.6" turns on, and off, the special control mode for adjusting the brightness and the contrast of the displayed image. In this mode, the "↑" and "↓" keys change the brightness level, and the "←" and "→" keys change the contrast of the displayed image. The key "f.7" turns on, and off, another control mode in which the frame number of the start- and end-frames of the slow motion display can be specified. In this mode, when the "←" key is pressed, the frame currently displayed is selected as the start-frame. When the "→" key is pressed, the frame currently displayed is selected as the end-frame. The slow motion display

Table 3. Keys for display control and their functions in the compact vocal cord image recording system.

Key	Function
f.1	Move back one frame
f.2	Move forward one frame
f.3	Set slow motion display speed to slow(3 frames/sec)
f.4	Set slow motion display speed to medium(6 frames/sec)
f.5	Set slow motion display speed to fast(9 frames/sec)
f.6	Contrast and brightness setting mode on/off
f.7	Start- and end-frame setting mode on/off
f.8	Stop motion display on/off
f.10	Stop slow motion display and prepare for next data acquisition

is repeated over the time interval between the start-frame and the end-frame. The "f. 8" key turns on, and off, the mode of the special stop motion display. The key "f.10" resets the system for the next image data acquisition.

We believe that the system described above has sufficient capabilities for the observation of pathological vocal cord vibration and is useful as a practical clinical unit.

This work was supported in part by a Scientific Research Grant (No.61480355, No.61870067) from the Ministry of Education, Sience and Culture of Japan, and Akai Foundation.

References

1) Honda K., S. Kiritani, H. Imagawa, H. Hirose and K. Hashimoto: High-speed digital recording of vocal fold vibration using a solid-state image sensor. Ann. Bull. RILP, 19, 47-53, 1985.
2) Imagawa H., S. Kiritani, K. Honda and H. Hirose: Improvements in the high-speed digital image recording system for observing vocal fold vibration. Ann. Bull. RILP, 20, 17-22, 1986.
3) Kiritani S., K. Honda, H. Imagawa and H. Hirose: Simultaneous high-speed digital recording of vocal fold vibration and speech signal. Proceedings ICASSP 86, TOKYO, Vol 3, 1633-1636, 1986.
4) Kiritani S., H. Imagawa and H. Hirose: Simultaneous high-speed digital recording of vocal fold vibration, speech and EGG. Ann. Bull. RILP, 20, 11-15, 1986.
5) Kiritani S., K. Honda, H. Imagawa and H. Hirose: Observation of pathological vocal fold vibrations using a high-speed digital image recording system. Proceedings of the XXth IALP Congress, 338, 1986.