Ann.Bull.RILP(1986)No.20, 11-15

SIMULTANEOUS HIGH-SPPED DIGITAL RECORDING OF VOCAL FOLD VIBRATION, SPEECH AND EGG

Shigeru Kiritani, Hiroshi Imagawa and Hajime Hirose

For the study of voice source characteristics in various modes of phonation, it is essential to record vocal fold vibrations simultaneously with output speech signals and to analyze the relationship between the pattern of vocal fold vibration and the acoustic characteristics of speech. In order to facilitate such data collection, we have been developing a new method for the high-speed digital recording of vocal fold vibration using a solid endoscope and a solid-state image sensor [1),2)]. Compared to an ordinary high speed motion picture system, these system is very compact and does not produce any mechanical noise, thus making it very easy to record simultaneously the speech signal and other physiological signals. Below, a few examples of the simultaneous recording of the vocal fold image, speech signal and the EGG signal obtained in a preliminary study will be presented.

Figures 1 and 2 show examples of recorded data for sustained phonations by a male and a female subject, respectively. The images were recorded at a rate of 2000 frames per second with 25 scan-lines per frame. The image data of the 25 scan-lines were displayed on a monitor osilloscope as images of 50 lines by reproducing each scan-line twice. The curves in each figure are, from top to bottom, the speech, the LPC residue and the EGG signal. The speech and EGG signals were sampled by a micro-computer at a rate of 10 kHz simultaneously with the image recording. The sampling pulses for the A/D conversion were generated from the 50 kHz End of the line signal in the scanning circuit of the image sensor. The 15 time marks in the figure show the timing of the corresponding vocal fold images. (In the present image sensor, the image signals of the individual elements are read out through an x- y-addressing method. Thus, the actual periods of exposure for the individual picture elements differ from each other. In the image, sampling proceeds from left to right, and from top to bottom. Each time mark in the figure represents the middle of the exposure period for the picture element at the center of each image.) A microphone for recording the speech signal was placed 10 cm from the subjects' lips. In the figures, no adjustments are made for the time lag of the speech signals.

Figure 1 shows the chest voice and falsetto of a male subject. They were produced at approximately the same fundamental frequency (180Hz). The image data in the figure show that, for the chest voice, the left and right arytenoids are adducted more tightly than for the falsetto. The position of the maximum width of the glottal opening is located at the anterior part of the vocal folds. Compared to this, in the case of the falsetto, the glottis appears to be open up to a more posterior part of the vocal folds. At the same time, the closure period is shorter for

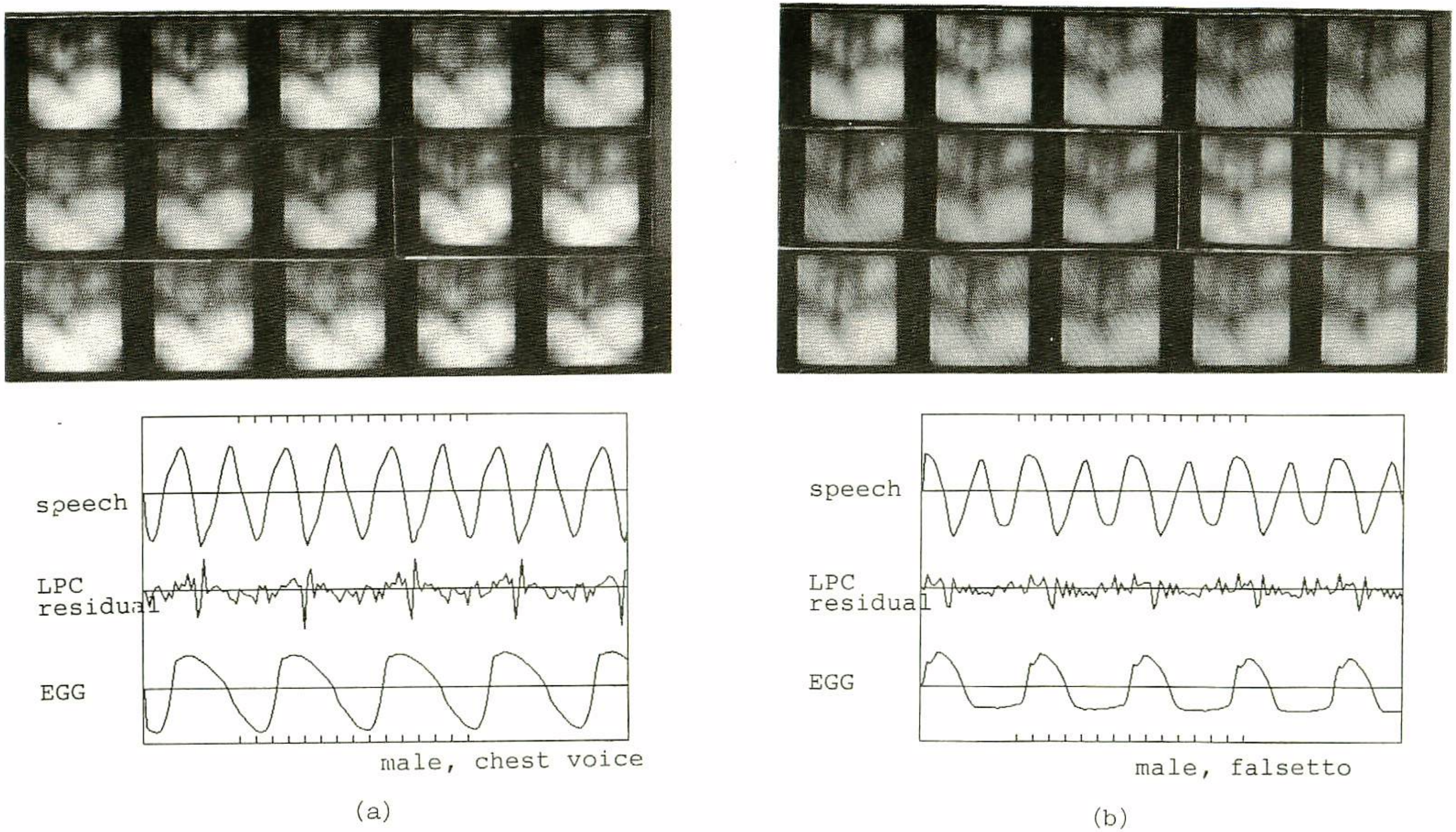
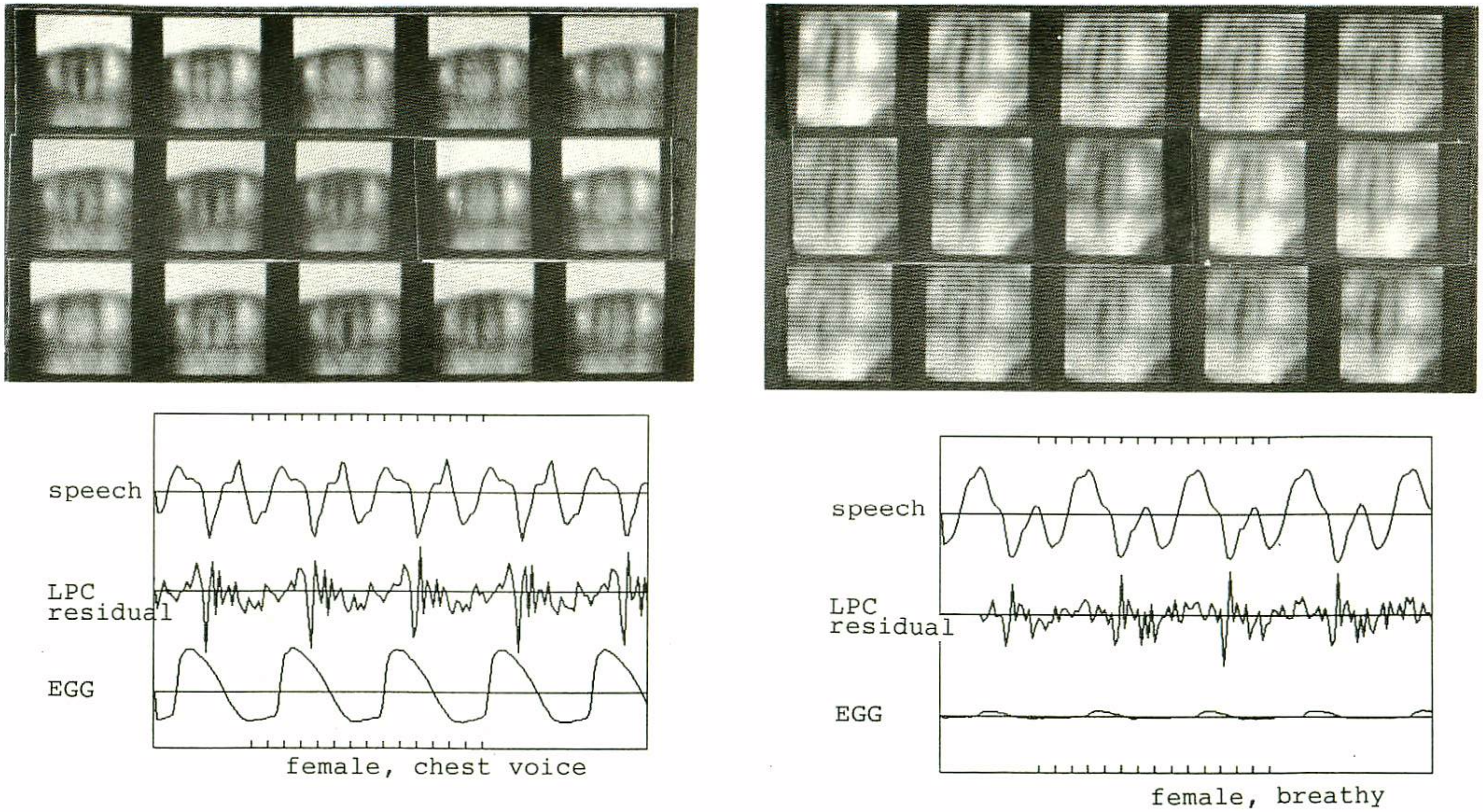


Figure 1 Vocal fold vibrations of a male subject in (a) chest voice and (b) falsetto.



(a) (b)

Figure 2 Vocal fold vibrations of a female subject in (a) chest voice and (b) breathy voice.

the falsetto than for the chest voice. The acoustic difference between the two types of voice can be seen clearly in the LPC residual waves. For the chest voice, there are clear excitation pulses near the moments of glottal closure. In the falsetto, such excitation pulses are very weak. However, in the EGG curves, there is no indication of a difference in the speed of the closing contact between these phonations. It appears that a more detailed analysis of the closing movement of the glottis is necessary to explain the difference in the pattern of the excitation in these phonations.

Figure 2 shows examples of the phonation of a female subject. This female chest voice shows flat valleys in the EEG curve corresponding to a longer open period. The pattern is similar to that of the male falsetto in Figure 1. However, the LPC residual for this voice has clear excitaion pulses. The image data in Figure 2 (b) show that, in the case of the breathy phonation shown here, the glottis remains open throughout the entire period of the vibration. However, it can be noted that,

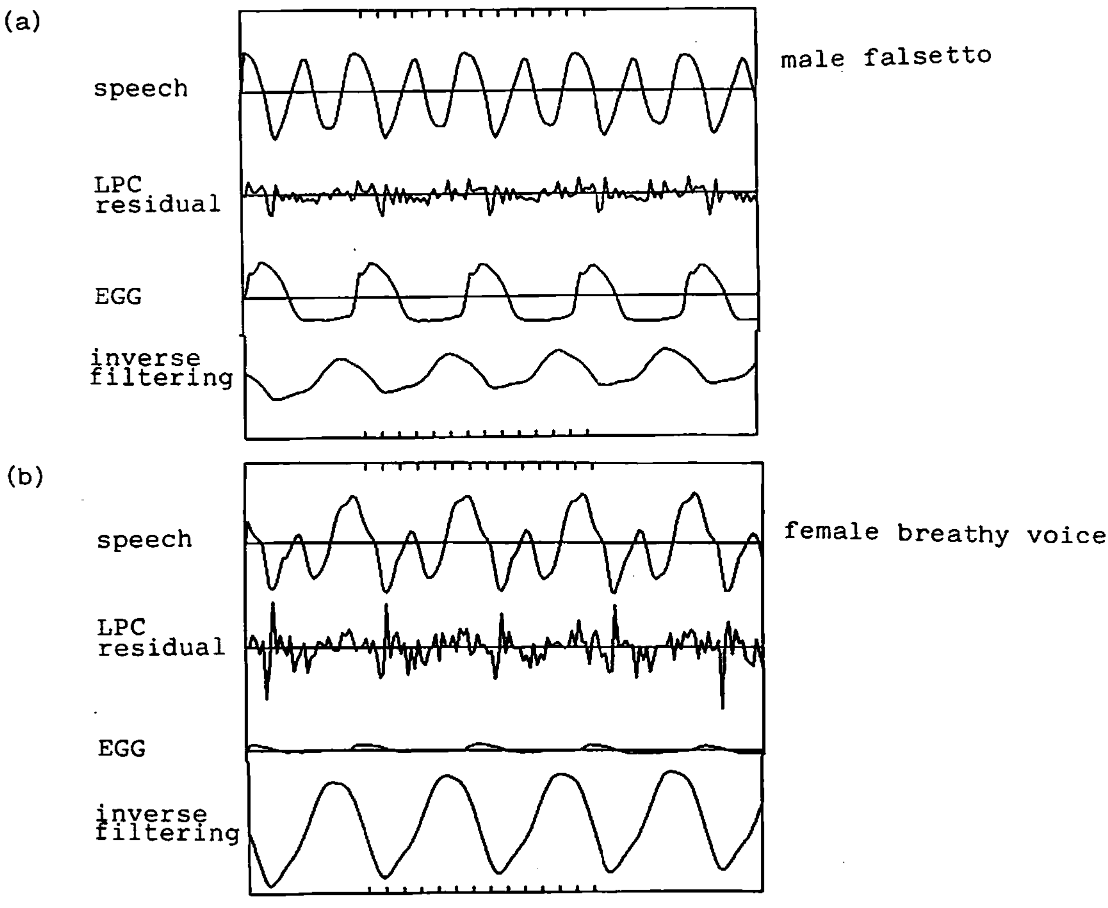


Figure 3 Examples of the inverse filtered speech waves for (a) male falsetto and (b) female breathy voice.

even for such a vibration pattern, pulse-like excitations can still be identified in the LPC residual at approximately the point of the minimum opening of the glottis.

Figure 3 shows examples of inverse filtered speech for the male falsetto and the female breathu voices. Formant frequencies for the inverse filtering were determined by an autocorrelation LPC analysis over 512 sample points of speech. In the case of the male falsetto, the inverse filtered wave has a nearly flat portion corresponding to the closure period of the glottis. Compared to this, the breathy voice by the female subject does not show such a flat portion. However, the waveform of this voice exhibits a considerably sharp deflection near minimum glottal opening, which appears to correlate with the excitation pulses in the residual wave.

A more detailed analysis of the relationship between the pattern of the vocal fold vibration and the characteristics of excitation is now being attempted. We wish to thank Prof. Seiji Niimi and Miss. Keiko Aoyama for their valuable cooperation in these experiments.

REFERENCES

1) K. Honda, S. Kiritani, H. Imagawa, H. Hirose and K. Hashimato: High-speed digital recording of vocal fold vibration using a solid-state image sensor, Ann. Bull. RILP, 19, 47-54 (1985).

2) H. Imagawa, 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).