High-speed Digital Image Recording System for Observing Vocal Cord Vibration

Shigeru Kiritani, Hajime Hirose and Hiroshi Imagawa
Research Institute of Logopedics and Phoniatrics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113, Japan

Introduction

Observation of vocal fold vibration is highly important for a study of the physiology and pathology of voice production. The observation and measurement of vocal fold vibration have generally been made by means of high-speed motion picture systems. However, this system is usually massive, and is not suited for flexible data collection covering various phonation types or various pathogical conditions. In addition, for the simultaneous recording of the voice signal, special consideration is necessary to reduce the mechalical noise from a high-speed camera. For the study of voice source characteristics, it is essential to record the vocal cord vibration simuitaneously 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 stroboscopic system is more compact and is useful for observing vocal fold vibration in clinical situations. However, a preccise cycle-by-cycle analisis of the vocal fold vibration is not posible when the vibration is irregular, as is often the case with pathological voices.

In order to overcome these difficulties, a new method to obtain high-speed digital imaging of vocal fold vibrations has been developed by the present authors using a solid-state image sensor and a digital image memory combined with an oblique angled solid endoscope or a fiberscope. This system is relatively free from mechanical noise and permits the simultaneous recording of speech signals. Since the entire system is also compact and easy to handle, the prospects for its application are very promising.

First, the system was constructed using a solid endoscope. The glottal image by the solid endoscope is brighter than that by the fiberccope and it was easier to realize the high frame rate image recording with the endoscope than with the fiberscope. The use of a solid endoscope system is naturally limited to the study of sustained phonation. The system is now being used at our institute especially for the observation and analysis of pathological vocal cord vibration.

As for the observation of laryngeal behavior during running speech, the fiberscope is now widely used in the field of experimental phonetic. In recent years, the possibillity of applying the technique of high-speed digital imaging to fiberscopic observation has been explored. It has now become possible to record the vocal coed vibration during running speech through the fiberscope. Preliminary studies of vocal cord vibration during the production of consonants are now being conducted. In the present paper, the characteristics of the high-speed digital imaging system will be described, together with examples of analyses of vocal cord vibration in the sustained phonation of the pathological voice and in the production of the consonants in running speech.

High-speed Digital Imaging System

Figure 1 shows a block diagram of the solid endoscope sysetm. 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 from the image sensor is fed into the image memory through a high-speed A/D converter. Stored images are then reproduced consecutively on a CRT monitor as a slow motion picture. Several different types of image display are possible under digital control.

The maximum frame rate which can be realized by the present system is determined by the brightness of the image obtained through the scope and the speed of which the pixels can be scanned by the image sensor. In order to get an image sufficiently bright for high-speed recording, a specially designed solid endoscope was constructed. This scope is wider and shorter than the conventional endscope in clinical use in order to get a greater transmission of light. The cross section of the scope is nearly rectangular and the image guide is located at the center. The light guide consists of two bundles of glass fibers. They are located on the left and the right sides of the image guide and each is connected to a seperate light source. These light sources are two 250W halogen lamps.

The fiberscope system also employs a specially designed fiberscope which was epuipped with a larger number of light guide fibers than a regular scope. The outer diameter is 0.49 cm. In this scope, the space, originally used to pass the biopsy forceps (the largest circle in the diagram of the tip of the scope in Figure 2) is filled with light guide fibers. As a result, an intensity approximately 3-4 times higher in illumination is obtainable compared to the conventional fiberscope.

To realize a higher frame rate, only a selected part of the image sensor is scanned to reduce the number of pixels sampled. In most analyses of vocal cord vibration, the lateral movements of the vocal cords are the main concern. Thus, when the total number of usable pixels is limited due to a limitation in the scanning speed, it is advantageous to arrange a greater number of pixels along the lateral dimension of the glottis than along the front-back dimension of the glottis, and to keep the resolution along lateral dimension higher. In the present system, the shape of the scan area is selected as 128 (horizontal) x 32 (vertical) pixels, at which a frame rate of 2000 per second is achieved. With the present sensor, it is possible to read out two pixels in adjacent horizontal scan line at one time, summing the image signals from the two pixels. Thus the aspect ratio of the image field is 2 to 1, although the ratio of the number of pixels in the horizontal dimension to those in the vertical dimension is 4 to 1.

The system was originally constructed based on a personal computer using an attached high-speed frame memory board. However, to realize a more flexible, high-speed control of the image acquisition/display process, a stand-alone high-speed digital video unit has been developed. The unit contains an image memory of 64M bytes and can store 16000 frames of image data 8 seconds in duration at a frame rate of 2000 per second. It can be operated as an ordinary video recorder, with functions such as PLAY at variable slow motion speeds, FAST FORWARD and FAST REWIND, STILL, and REPEAT of the selected time intervals. Data storage into the image memory is started by the shutter button on the camera head.

In experimental situations, the speech signal and the EGG signal are digitized simultaneously with image recording by a seperate personal computer. These signals are sampled synchronously with the image recording using sampling pulses generated in the camera head. At the time of the reproduction of the image data by the Digital Video Unit, the sampled data are sent to the Digital Video Unit and are displayed simultaneously with the glottal images.

Figure 3 shows an example of the image display on the video monitor. The display shows the vocal cord vibration during sustained phonation of the vowel /e/ by a normal male subject. The glottal images in successive time frames are displayed as a two dimensional array. At the bottom of the screen are shown the speech waveforms and the EGG signals. The image at the upper left corner is the first frame in this time sequence. The following 4 frames are displayed in the same horizontal row, and the succeeding frames are displayed on the next row. The image at the lower right corner is the last frame. The fundamental frequency of the voice is about 200 Hz, and about two cycles of vibration can be seen in this display. The vocal cord vibration in this figure shows regular symmetrical movements.

Vocal Cord Vibration in Diplophonia
-Observation by the solid endo-scope system-

Figure 4 show an example of simultaneous recording of the vocal cord vibration and the speech signal made of a case of diplophonia (unilateral recurrent nerve palsy). The speech signal shows quasi-periodic variations in the amplitude and waveform which repeat at about 11 vibratory cycles.

In order to obtain clear information on the temporal pattern of the vibration in such cases, an effective measure of the glottal width was defined as follows and its temporal course was displayed. Namely, in the glottal image, the area where the brightness was darker than a selected threshold was taken as the area of the glottal opening. A horizontal scan line at an appropriate position on the glottal opning was selected. By examining the brightness curve on this scan line, the width of the glottal opening was measured and displayed as a short line segment whose upper and lower end represents the position of the edge of the left and the right vocal cord at each time frame, respectively.

Display of the time sequence of the glottal width measure in Figure 4 reveals that the left and the right vocal cords vibrate apparently at different frequencies, and that the phase difference between the movements of the two vocal cords varies with time. At around the letter A in the figure, the movements of the vocal cords are nearly in phase, and the glottis shows a period of complete closure. At around cycle No.6 indicated in the figure, the phase difference is larger, and the glottal closure becomes incomplete. At cycle No. 9, the phase difference is nearly 180 degrees, and the left and right vocal cords move almost in parallel. In this cycle, the inward movement of the right vocal cord is incomplete.

During cycle No.10, the right vocal cord remains in an outer position. Then cycle No.11 of the right vocal cord starts with a light outward movement followed by a clear closing movement, which is nearly synchronous with closing movement of the left vocal cord. At this point, the left and right vocal cords resume synchrony, and the vibration shows a period of glottal closure. Thus, in this case, the 10th vibratory cycle of the right vocal cord almost disappears and is apparently fused with the preceding and following cycles.

The temporal change in the pattern of vocal cord vibration described above explains the pattern of the temporal change in the speech waveform. At around A, where the vocal cord vibration exhibits complete closure, there is a strong excitation pattern in the speech waveform. The amplitude is large, and the format oscilations are very clear. When the glottal closure become incomplete, the excitation becomes weaker. The waveform is noisy, and its amplitude becomes smaller.

Vocal Cord Vibration in /h/ and /s/
-Observation by the fiberscope system-

Figure 5 shows the start of vocal cord vibration in the word initial /h/ in /he: he:/. The first glottal approximation takes place at around the frame 11 in the figure. At this moment, the posterior part of the glottis remains open, and even in the next cycle, the glottal closure is still incomplete. Compared to this, in the case of /s/ shown in Figure 6, Complete glottal closure is generally observed in the second cycle, and a tight glottal closure is achieved rather quickly.

As indices of the glottal vibration, time series of the brightness curve at a selected horizontal scan line on the glottal image are displayed in Figure 7. A scan line close to the posterior end of the glottis is shown in this data display. Dips in the brightness curve towards the right correspond to the dark area of glottal opening. This figure shows that, in the case of /h/, the effect of vibration can be seen up to the eighth cycle after vioce onset, whereas in the case of /s/ effect of vibration disappears after the second cycle.

This difference in the pattern of glottal vibration between /s/ and /h/ is clearly reflected in the spectrum of the voice source. Figure 12 also shows the spectrum of the inverse filtered speech waves for the several pitch periods after voice onset. In the case of /s/, the spectrum in the second pitch period already reaches a stationary pattern as that in the following vowel. In contrast to this, for /h/, the spectrum change continues over 5-6 pitch periods.

Concluding Remarks

The present study shows that an application of the high-speed digital imaging technique to the analysis of voice production could prove very useful. The procedure for the recording and analysis of vocal cord vibration in the present system is simple compared to that in high-speed motion picture systems. Simultaneous recording of the vocal cord vibration and the speech signal--and the analysis of the correspondence between them--is valuable to understand the mechanism of production of pathological voice quality.

Furthemore, the high-speed digital imaging system using a fiberscope made possible the observation of the vocal cord vibration during running speech. Although more precise and quantitative studies must be made in the furure, the present study shows that the system can provide useful information for the analysis of voice source characteristics during running speech.

References

Honda, K., 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.

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

Kiritani, S., H. Imagawa and H. Hirtose: "Vocal Cord Vibration and Voice Source Char acteristics -Observations by the High-speed Digital Image Recording." Proc. ICSLP-90, Kobe, Japan. 61-64,1990.

Kiritani, S., H. Imagawa and H. Hirose: "Vocal fold vibration and the speech waveform in diplophonia", Ann. Bull. Res. Inst. Logopedics and Phoniatrics, University of Tokyo, 55-62, 1991.

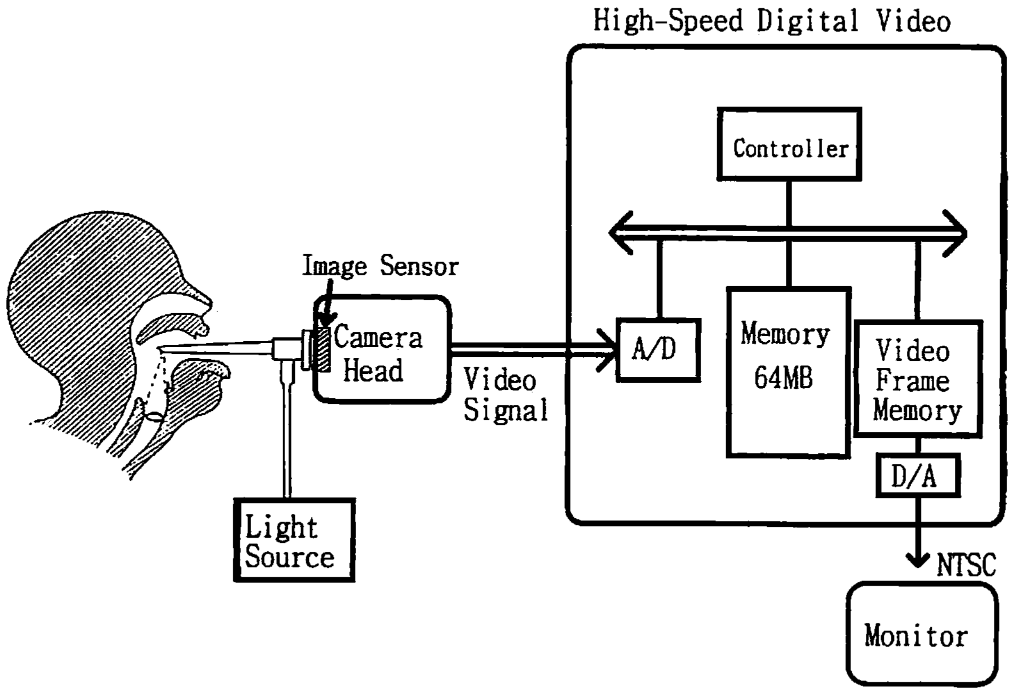

Figure 1. Block diagram of the high-speed digital image recording system using a solid endoscope.

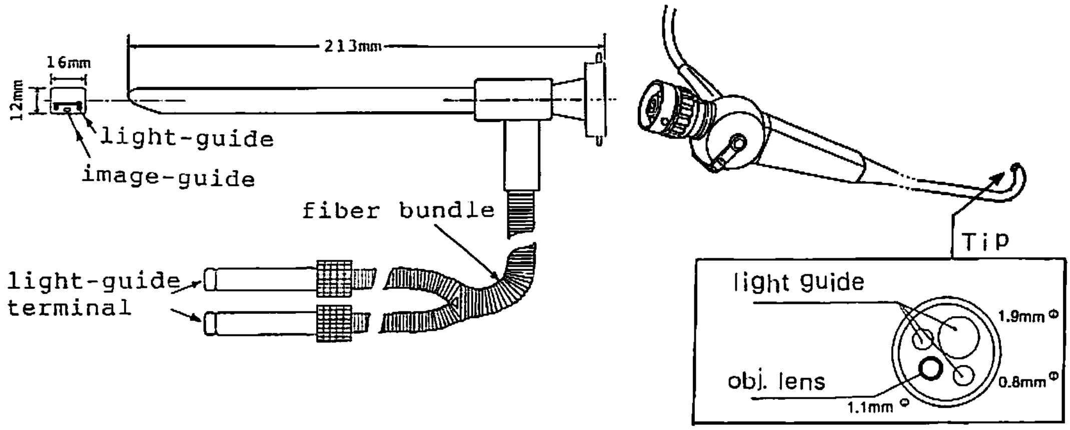

Figure 2. A schematic view of the solid endoscope and the fiberscope.

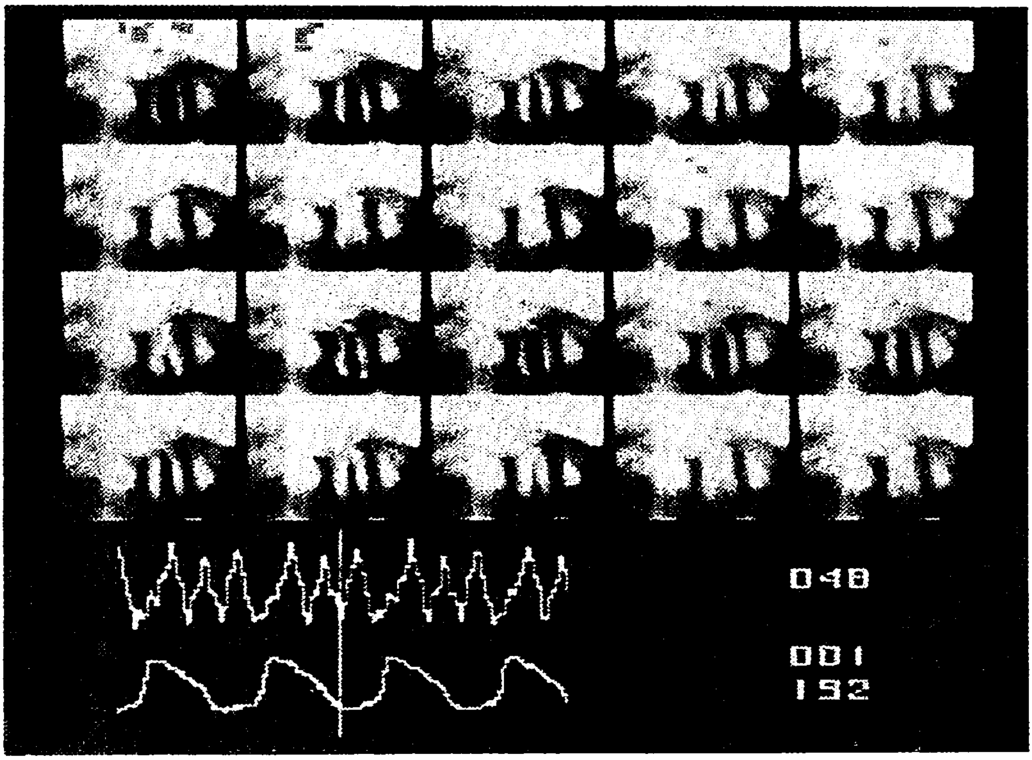

Figure 3. An example of the display of the glottal images recorded at a rate of 2000 frames per second. The vocal cord vibration during sustained phonation by a normal male subject is shown.

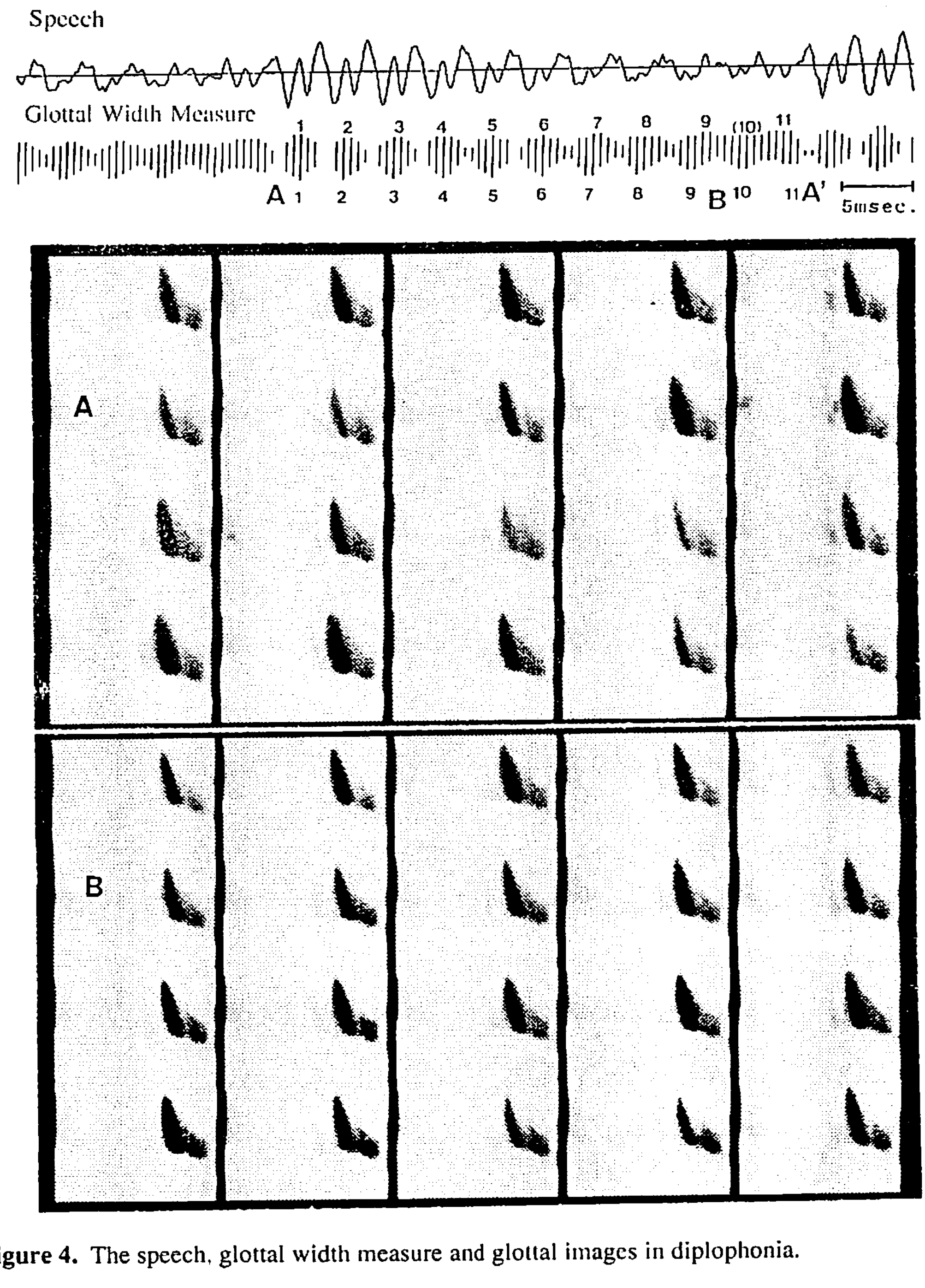

Figure 4. The speech, glottal width measure and glottal images in diplophonia.

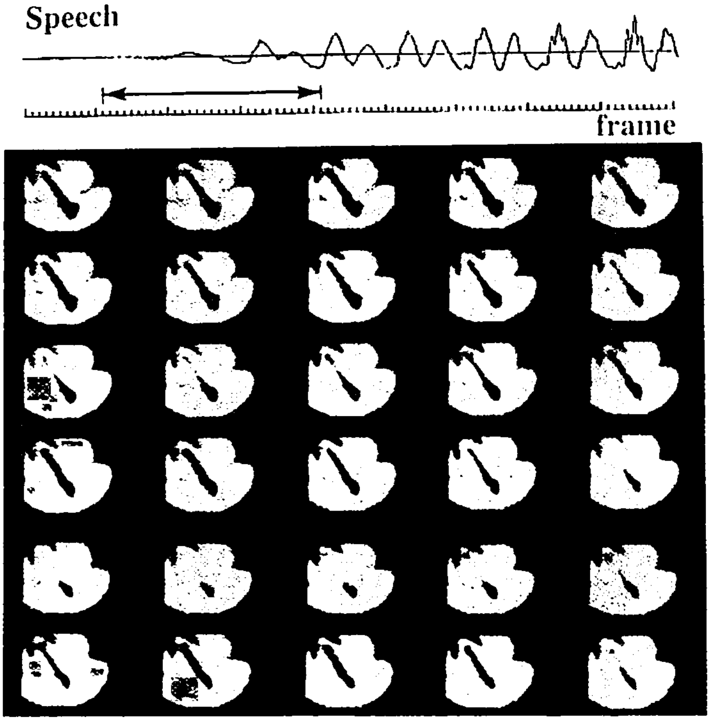

Figure 5. Vocal cord vibration at voice onset in /he:he:/.

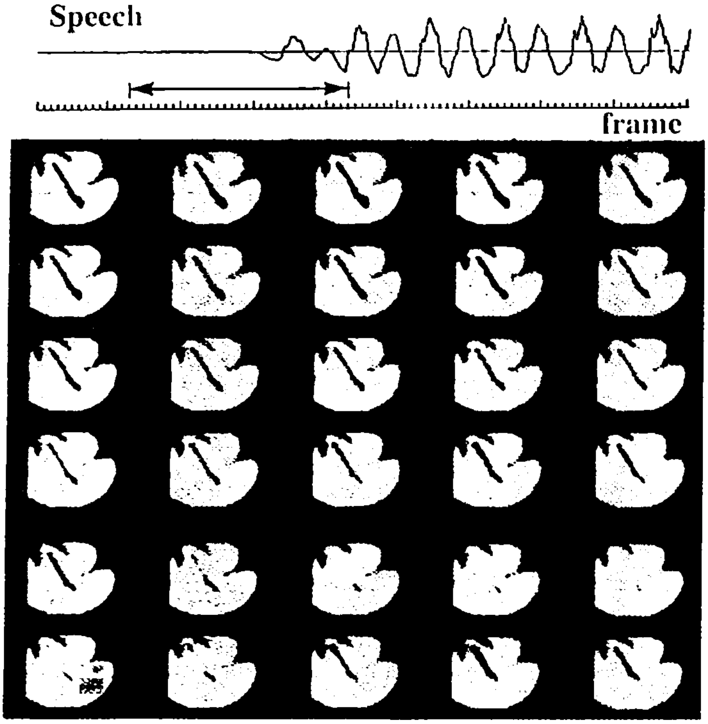

Figure 6. Vocal cord vibration at voice onset in /se:se:/.

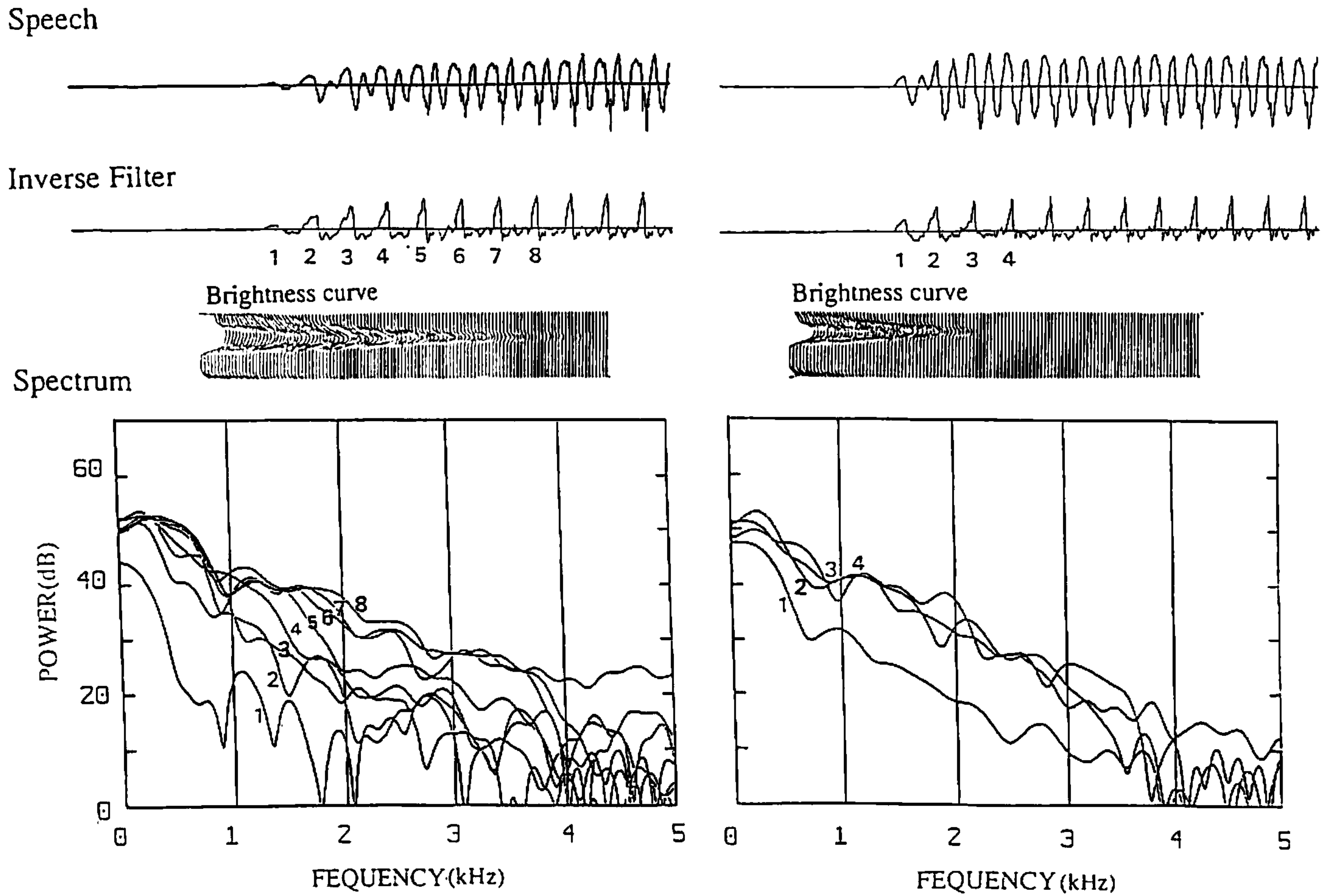

Figure 7. Inverse-filtered source wave and spectrum in /he:he:/ and /se:se:/.