Computer Analysis of High-Speed Motion Pictures of the Vocal Folds Vibrations

Masahiro TANABE, Kazutomo KITAJIMA and Wilbur J. GOULD*

Since the Bell Telephone Laboratories first succeeded in producing high-speed motion pictures of vibrating vocal folds in 1937, a number of modifications in the filming system and analysis technique have been reported. The analysis techniques have tended to follow one basic approach; the film is projected one frame at a time onto a screen, and measurements of the relative vocal fold size and motion are made.

These approaches have in common the fact that it is extremely difficult by purely visual means to retrieve and correlate all of the data that may be present in a single frame, even though the frame itself covers but a relatively short period. In most instances, accordingly, these approaches are used for measurement or quantification in relation to a single point on the vocal folds.

Recently, in an endeavor to overcome some of these inherent drawbacks, Hayden and Koike (1972) reported a technique for frame-by-frame film analysis employing a PDP-8 computer. Although the system they used greatly simplified frame-by-frame analysis, the measuring system nevertheless requires teletype keyboard operation and it therefore remains time consuming. The time required for data analysis is wholly dependent upon the quality of the filming, and film processing systems, and both of these may vary widely.

It seemed to us that some of these drawbacks might be eliminated, and we therefore sought to devise a method that would both reduce the time required for analysis and reduce the role of subjective (operator) judgement in the study of high-speed cinematographs of the vocal folds. The purpose of this paper is to describe the elements of the data reduction system we have devised. The system itself incorporates filming modifications designed to produce sharper images with better definition of detail and hence of potential points of reference.

To date, filming techniques have remained relatively unchanged. A high intensity, continuous light is used for illumination, resulting in an exposure time approximately 20 to 100 microseconds per frame. This long exposure time causes the image to appear unsharp, and heat from the high intensity bulb is also a problem. For this reason care must be taken not to damage the tissue being

Masahiro TANABE (田辺正博) and Kazutomo KITAJIMA (北嶋和智): Department of Otolaryngology of the Kyoto University, Faculty of Medicine.
* Wilbur J. GOULD: Vocal Dynamics Laboratory, Lenox Hill Hospital, New York.

observed.

As shown in figures 1 and 2, our filming system employs a high intensity, high speed strobo (model 501) developed by the E.G. & G. Company. The lamp flash rate, controllable up to 6000 flashes per second, is syncronized to the camera (Hycam model K2001) shutter resulting in one flash per frame. In this manner, the exposure time has been reduced to less than 2.1 microseconds per frame. The light from the flash tube is directed into a front surface mirror which has a one inch circular hole cut through the center for viewing. The light from the front surface mirror is then reflected onto a laryngeal mirror, for illumination and view-

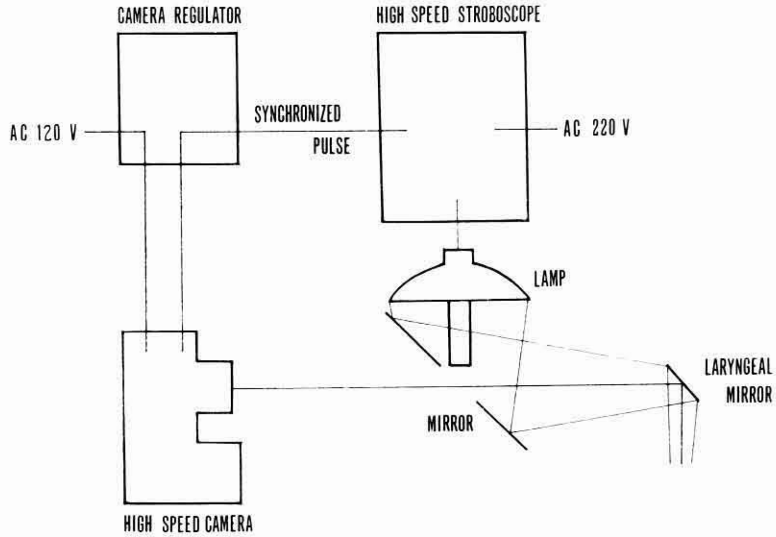

Figure 1. Schematic presentation of filming system.

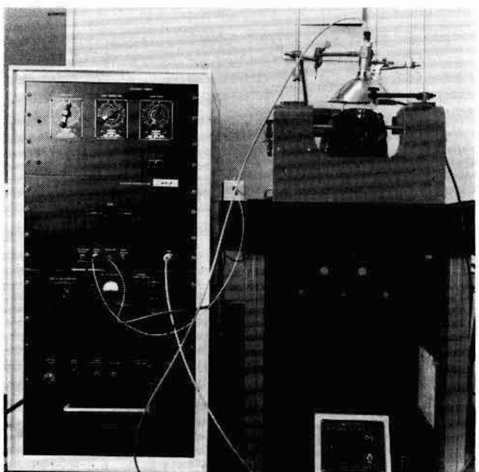

Figure 2. High speed strobo (model 501) and a flash lamp.

ing of the vocal folds. The laryngeal mirror is fixed into position, and the subject must be maneuvered into the proper position for optimum viewing of the vocal folds.

Previous techniques required an intricate system of heat absorbing glass, and in some cases a water-bath filter. All of these additional lenses make the filming systems complex, cumbersome, and expensive. Since the interrupted light used in our filming system produces a "cold" means of illumination, heat filtering is not necessary. Fig 3 shows a series of pictures taken during one complete vibratory cycle obtained using this filming system.

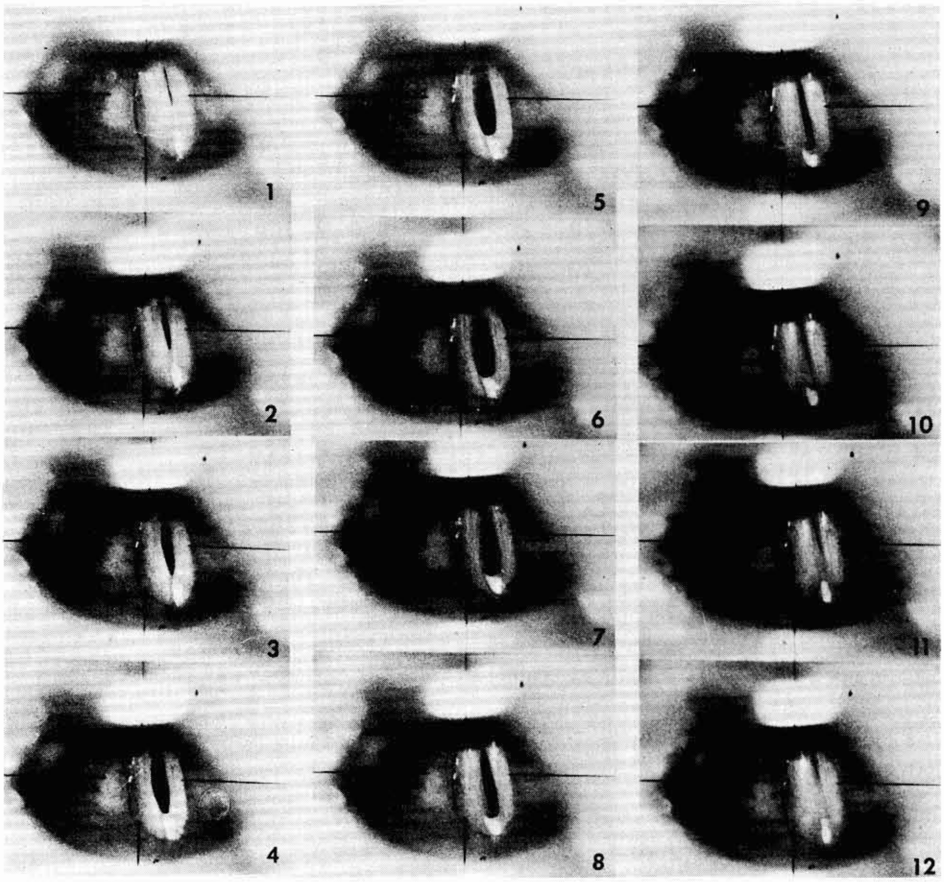

Figure 3. A series of laryngeal pictures taken at 3000 frames per second.

The data reduction system (figure 4) includes an analysis projector capable of projecting film frame-by-frame, and a PDP-12 computer for reading and processing the visual data. The film is projected onto the face of a display scope of the computer, which is programmed to simultaneously project coordinate points. The precise positions of the coordinate points are adjustable by the operator, using potentiometers, and they are positioned to divide the vocal folds into six segments.

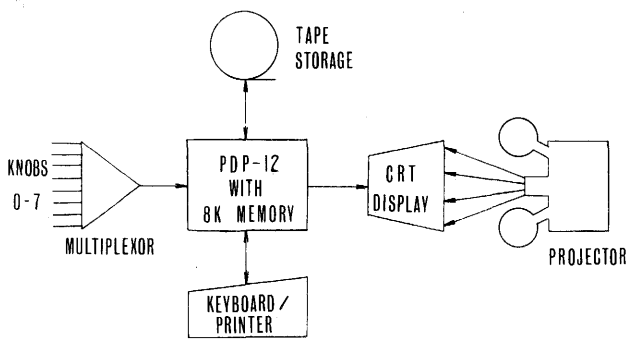

Figure 4. Schematic presentation of analysis system.

Once the coordinates have been positioned, the film is projected onto the display scope and the projector is adjusted to project the anterior commisure of the vocal folds on point A (figure 5). It should be noted that point A is a fixed point, not under the control of any potentiometer.

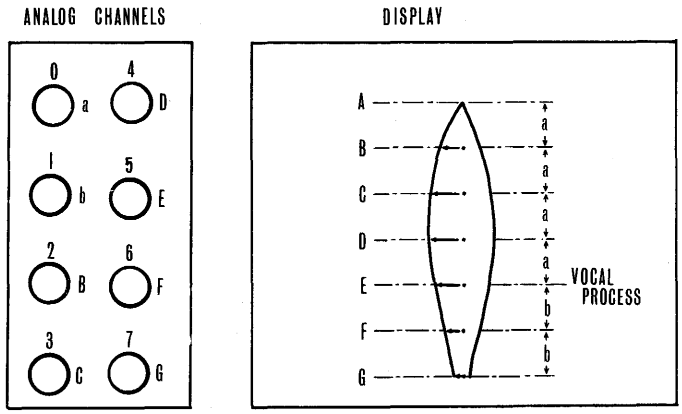

Figure 5. Position of coordinate points on the vocal fold edge.

The vertical position of point E is adjusted by means of potentiometer 0 so that it lies on the vocal process. Since points B, C, and D have been set so as to divide the length from A to E into four equal parts, the vertical positioning of points B, C, and D is automatically determined by the position of point E.

The vertical position of points F and G is then determined placing G on the posterior commisure, by means of potentiometer 1. Since point F will divide segment E—G into equal portions, the vertical position of point F is automatically determined by the position of point G.

The horizontal coordinates of points B, C, D, E, F, and G are controlled by potentiometers 2, 3, 4, 5, 6, and 7 respectively, and these points are aligned to correspond with the edge of one vocal fold. Once the operator is satisfied with the positioning of these points, a signal is given to store the coordinates of each point on magnetic tape. This same procedure is then repeated for the contralateral vocal fold. In most cases, however, the initial three steps may be omitted since the vocal folds will be identical in length.

At this point the film is advanced for processing, frame-by-frame, and the analysis program then computes the following parameters from the various points as a function of time:

1) the excursion curves of both the right and left vocal folds at all measured points.

2) glottal width, length, and area.

The results can either be presented visually as graphs on the display scope or typed out on teletype paper. When the magnification ratio of the vocal folds on the film is given, the absolute value of all these parameters noted above can be derived immediately and typed out.

The glottal width and area, it might be noted, have been analyzed as relative values by previous investigators, but in this study an attempt was made to obtain absolute values for these parameters, this procedure will be discussed briefly.

Since the magnification ratio will differ from one subject to another because of the changing film-object distance, a reference point is incorporated into the procedure employed. To obtain the reference point a metric ruler is filmed with the same lens setting that is used in photographing the larynx. When the ruler is brought into focus, the lens-object distance is the same as that during the actual filming of the larynx, since the filming system used employs a wide open lens setting having a virtual zero depth of focus. The vertical movement of the larynx does not effect the measurement, since the phonation is a vowel phonated at a constant pitch, and the time span covered on the film is short. The filmed ruler is next focussed onto the display scope and the values between each of the scales are read and stored on the magnetic tape. Using this scale in processing stored data, it is possible to obtain the absolute value of each of the parameters noted.

Both the program for reading, as well as that for analyzing the data are written in FOCAL-12, which is equivalent to FORTRAN. The use of the same computer language for both operations greatly simplifies the process since no system change is required to shift from reading to analysis or vice-versa.

Some typical analyses obtained using this system are presented in the figures 6 and 7. On both graphs, lines A, B, C, D, E, F, and G represent excursion of each measured point and the designated "area" represent the glottal area as a function of time during one complete vibratory cycle. As can be seen from these graphs, there are a number of advantages in displaying glottal width at a number of points in this fashion. It is readily possible, for example, to compare the amplitude at different points.

In addition, phase differences between different points, shown clearly in figure 6, can be easily analyzed. In figure 6, it should be pointed out the opening

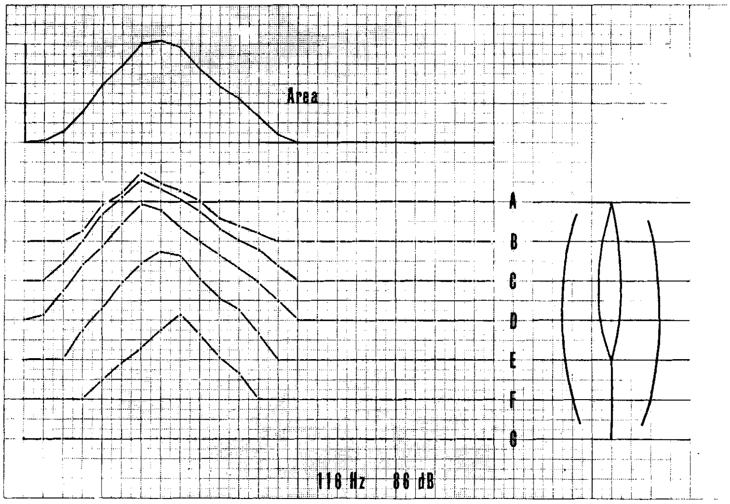

Figure 6. Glottal area and excursion of each measured point at 116 Hz, 86dB[1].

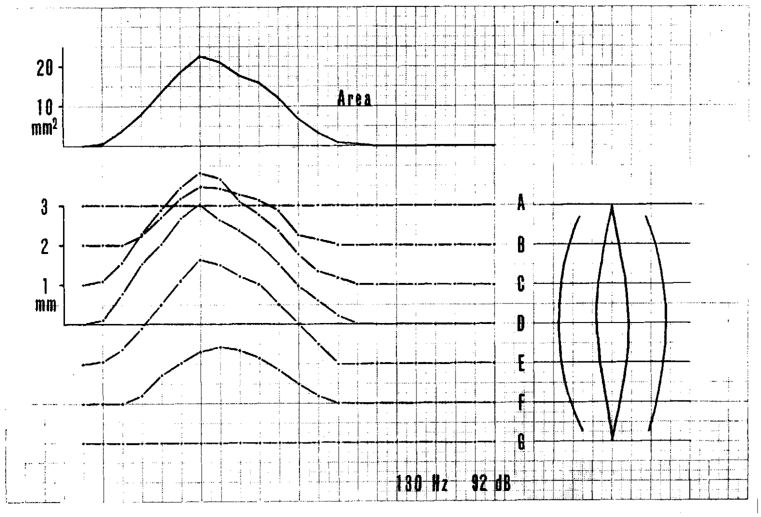

Figure 7. Absolute glottal area and excursion of each measured point at 130 Hz, 92dB[1].

1. Recorded using a Bruel and Kjaer microphone (model 4145) and a B & K SPL meter (model 2203). Microphone was placed 20 cm from subject's mouth.

starts at the anterior portion and travels toward the posterior commisure (longitudinal wave).

The absolute amplitude of the glottal width and the magnitude of the glottal area are seen in figure 7. As already mentioned the absolute values are computed by processing the raw data with the magnification ratio derived from the film of metric ruler made at the same lens setting.

References

1. Bell Telephone Laboratories: High Speed Motion Pictures of the Vocal Cords. New York, Bureau of Publication, Bell Telephone Laboratories. (1937).
2. Hayden, E.H. and Koike, Y.: A Data Processing Scheme for Frame by Frame Film Analysis. Folia Phoniatrica **24**: 169–181 (1972).
3. Moore, P. and von Leden, H.: Dynamic Variations of the Vibratory Pattern in the Normal Larynx. Folia Phoniatrica **10**: 205–238 (1958).

(Aug. 31, 1975, received)