A method of applying Fourier analysis to high-speed laryngoscopy

Svante Granqvist[a)]
Department of Speech, Music and Hearing, Royal Institute of Technology, Stockholm, Sweden

Per-Åke Lindestad[b)]
Department of Logopedics and Phoniatrics, Karolinska Institute, Huddinge University Hospital, Stockholm, Sweden

(Received 25 May 2000; accepted for publication 18 June 2001)

A new method for analysis of digital high-speed recordings of vocal-fold vibrations is presented. The method is based on the extraction of light-intensity time sequences from consecutive images, which in turn are Fourier transformed. The spectra thus acquired can be displayed in four different modes, each having its own benefits. When applied to the larynx, the method visualizes oscillations in the entire laryngeal area, not merely the glottal region. The method was applied to two laryngoscopic high-speed image sequences. Among these examples, covibrations in the ventricular folds and in the mucosa covering the arytenoid cartilages were found. In some cases the covibrations occurred at other frequencies than those of the glottis. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1397321]

PACS numbers: 43.70.Jt [AL]

I. INTRODUCTION

The periodicity of the vocal-fold vibrations is a highly relevant aspect of voice production, both in normal and in pathological voices. The periodicity, or lack of periodicity, is typically described in terms of jitter and/or shimmer, period-to-period correlation, or by spectral characteristics (see, e.g., Titze and Liang, 1993; Hess, 1983). Analysis of such characteristics is generally applied to acoustic signals recorded from microphones, but can also be used on signals derived from physiological events such as EGG or airflow recorded from flow masks.

Another method for acquisition of physiological data is direct visual inspection of the vocal folds by means of laryngoscopy. This method is often combined with stroboscopy (see, e.g., Švec, 2000) to visualize the vibrations of the vocal folds. However, aperiodicities associated with some voice qualities make stroboscopy inappropriate, since it requires periodic vibration and a single, measurable fundamental frequency. One way to circumvent this problem is to use high-speed imaging, a technique that has been used for many decades (see, e.g., Moore *et al.*, 1962; Dunker *et al.*, 1964). High-speed imaging therefore also offers an informative description of aperiodic vocal-fold oscillations, since each vibratory cycle is documented in terms of a sequence of several images. Recently, digital high-speed imaging of the larynx has become more commonly used, mainly because of reduced costs and improved light sensitivity of high-speed cameras with digital storage (see, e.g., Kiritani *et al.*, 1988; Hammarberg, 1995; Kiritani, 1995; Köster *et al.*, 1999; Eyshold *et al.*, 1996). Moreover, modern cameras can now produce images of acceptable quality when combined with standard clinical optical instruments.

A problem associated with video recordings in general, however, is visual interpretation. In particular, it is sometimes hard to observe which parts of the larynx are vibrating, at what frequencies, and in what phase relation to other vibrations. One solution is to use kymography, which can be acquired directly from a single-line camera (Švec and Schutte, 1996) or by extraction from high-speed image sequences (Tigges *et al.*, 1999; Larsson *et al.*, 2000). Such kymographic images give a good view of the movements of the vocal folds, periodic or nonperiodic, but only for part of the image, i.e., the single line. Another method for data extraction from high-speed images is detection of the edges of the glottis (e.g., Larsson *et al.*, 2000), which can be used for glottal area and flow calculations. Such methods are mostly applicable, however, only to the glottis, as they generally cannot reveal oscillations in other parts of the larynx.

This paper presents a new method to visualize periodic or quasiperiodic oscillations in the entire laryngeal area. The basic idea is to extract time signals from consecutive images. The resulting signals are Fourier transformed and can be displayed in different ways, for example in terms of a coloring of the laryngeal image. The method can be seen as a way of condensing data from all of the images within an image sequence and for all locations within those images into one or a few more informative images, displaying relevant oscillatory frequencies.

II. MATERIAL

Digital high-speed recordings were made at the department of Logopedics and Phoniatrics at Huddinge University Hospital. The subject was a male with a healthy voice (co-author P-Å L). Laryngoscopy was performed using either a flexible (Olympus ENF, P3) or a rigid (70° Hopkins 8706 CJ, Karl Storz) endoscope. The light source was a Storz 600 with a halogen lamp. The endoscopes were connected to the cam-

[a)]Electronic mail: svante.granqvist@speech.kth.se
[b)]Electronic mail: Per-Ake.Lindestad@logphon.hs.sll.se

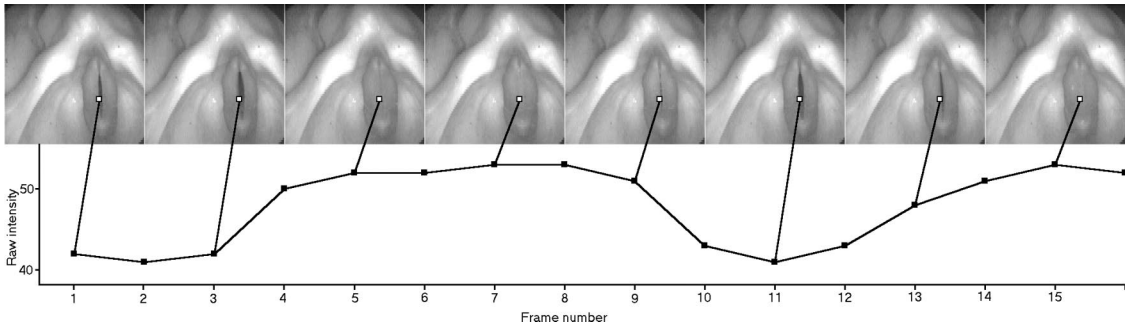

FIG. 1. Derivation of a single time sequence from an image sequence. The lower graph displays the time-varying intensity at the indicated pixel in the upper series of images. Alternatively, a similar intensity curve can be derived from the average intensity over an entire line, rather than from a single pixel. For clarity, every second image in the sequence has been excluded in the figure.

era system via an Olympus AR-L2 C-adapter. The camera system was a Weinberger Speedcam that was run at a frame rate of 1904 images per second, memory allowing recording of 2048 frames, or 1 s, approximately. The image resolution was 256×64 pixels. Images were stored and analyzed digitally. A Hamming window of 270 ms (512 points) was used for all Fourier transforms.

III. APPLICATIONS

The extraction of the time sequences can be applied either to a specific pixel or to a specific line in the image. This procedure involves three steps. First, the raw intensity data are extracted, i.e., the level of gray at the pixel in question, or the average level of gray of the line in question. This extraction is performed on each image with fixed line or point coordinates on a set of consecutive images, as illustrated in Fig. 1. The resulting data points are used to form consecutive samples in a time signal. Figure 2 (left) provides an example. Second, this time signal is scaled according to the average amplitude so as to eliminate the effect of different illumination in different parts of the image, Fig. 2 (middle). Third, the average of the signal is removed, Fig. 2 (right), since a large nonzero average component would obscure the lower part of the spectrum.

The time sequences thus obtained are then Fourier transformed. This results in a spectrum for each pixel or line. The output of these Fourier transforms contains amplitude and phase information as a function of frequency. Such Fourier analyses can be visualized in several ways simultaneously and preferably in conjunction with kymography and standard playback of the image sequence. Since there are four to five parameters involved (*x* position, *y* position, frequency, amplitude, and possibly phase), some of them have to be left out if a two- or three-dimensional graph is chosen. Four alternative modes seem particularly informative.

A. Single-pixel, all frequencies

In this mode of visualization, a single pixel is selected with a cursor. The Fourier transform of the light intensity at that pixel in the selected and consecutive images is displayed in a two-dimensional graph. The graph is an ordinary line spectrum display with frequency in Hz on the *x* axis and amplitude in dB on the *y* axis. For example, if the pixel or line is positioned at an oscillating glottis, a peak will appear in the amplitude spectrum at the glottal vibration frequency. This is illustrated in Fig. 3 pertaining to a high-pitched phonation. The figure shows a fundamental oscillation frequency of 210 Hz. $F0$ analysis of the corresponding sound, recorded simultaneously, yielded an $F0$ of 214 Hz.

B. Line average, all *y* positions, all frequencies

In this mode, the time curve is extracted by averaging the intensity for each line in the images. This yields a set of time sequences, one for each line. The sequences are Fourier transformed and the result is displayed in a three-dimensional graph with frequency on the *x* axis, vertical position on the *y* axis, and amplitude on the *z* axis, which is represented as level of gray. Figure 4 exemplifies such a three-dimensional graph, revealing major oscillation frequencies. At the level of the glottis in the image, shown to the left in figure, a dark line can be seen at 210 Hz. This line corresponds to the frequency of vocal-fold vibration; see above. The vertical line at 100 Hz is an artifact caused by

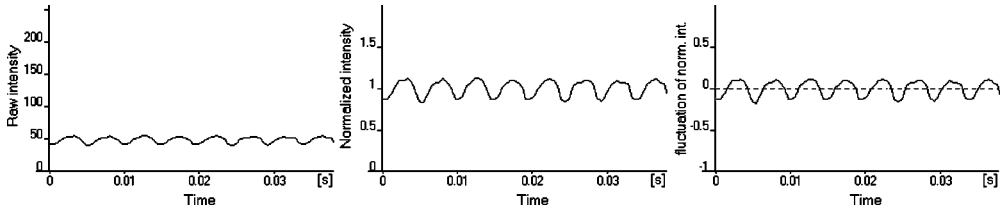

FIG. 2. Normalization of the time sequence, applied to compensate for differences in mean illumination strength and light absorption in different parts of an image. Left panel: raw signal taken from Fig. 1; middle panel: normalization of the same signal achieved by multiplying each value by a constant; right panel: resulting signal after removal of average. In this figure, the time window is longer than that used in Fig. 1.

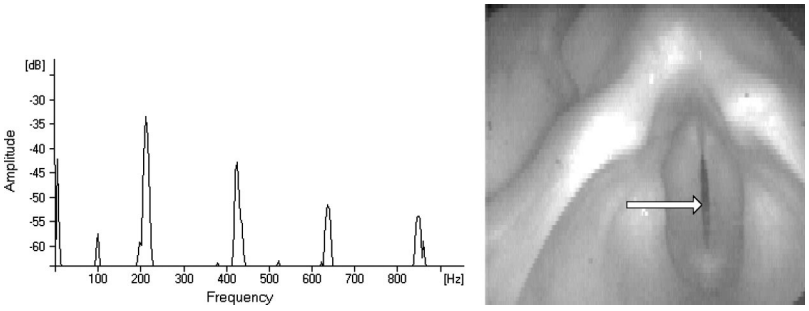

FIG. 3. Fourier transformation (512-point Hamming window) of the signal shown in the right panel of Fig. 2 yielding a spectrum for the selected point. The fundamental at 210 Hz corresponds to the vibration frequency.

light-source flickering. This was checked by analyzing the signal from a light sensor exposed to the light source; this signal contained a 100-Hz component. The flickering was probably due to the fact that the power supply of the lamp was not sufficiently stabilized. The averaging over lines as described above is preferable in cases where oscillatory similarities occur horizontally over the image. If, however, such similarities occur vertically, the averaging could be done over columns.

C. Full image, single frequency

In this mode, Fourier transforms are calculated for each pixel, and a single frequency from the Fourier transforms is selected. The magnitude of the oscillations at that frequency is displayed as color saturation on top of a single image selected from the original sequence. The image is then colored intensely in areas that oscillate at the selected frequency, e.g., in the glottal area. The image in Fig. 5 shows an example, illustrating the amplitude of the intensity variation at the frequency of 210 Hz. The oscillating vocal folds are intensely colored in red, since the light intensity variation in this area is great. The isolated red spots are artifacts caused by light reflections in the wet mucosa. The position of such a glare spot is highly sensitive to a tilting motion of the object, since this causes the light beam to move. At a given pixel this yields a great variation of light intensity and thus high amplitude and intense coloring, even though the motion is not large. Thus, such glare spots are of minor relevance.

D. Full image, single frequency, including phase information

A variant of the last method mentioned is to include the phase information from the Fourier transforms as the hue of the color. In this display mode, those areas of the image in which the intensities oscillate in different phase will be colored differently. The images in Fig. 6 illustrate this display mode for a hyperfunctional breathy phonation as produced by the same subject at a fundamental frequency of about 130 Hz. It must be noted, however, that this phonatory setting is not necessarily typical for a hyperfunctional breathy voice. It can be seen that the vocal folds and the ventricular folds are differently colored, due to an inverse phase relation in the intensity fluctuations between these areas. Careful examination of the high-speed recording confirmed that in this phonation the ventricular folds oscillated at the same frequency as the vocal folds, but in an opposite phase. Thus, the closed phase of the glottis was synchronous with the maximum separation of the ventricular folds. This is further illustrated by Fig. 7.

IV. DISCUSSION

When using the method presented here, certain aspects should be kept in mind. The approach in this paper is based on light-intensity fluctuation. This leads to some problems, since it introduces a nonlinear transfer function from motion to intensity. In particular, edges in the image will yield higher amplitude than flat surfaces, since a motion in a flat

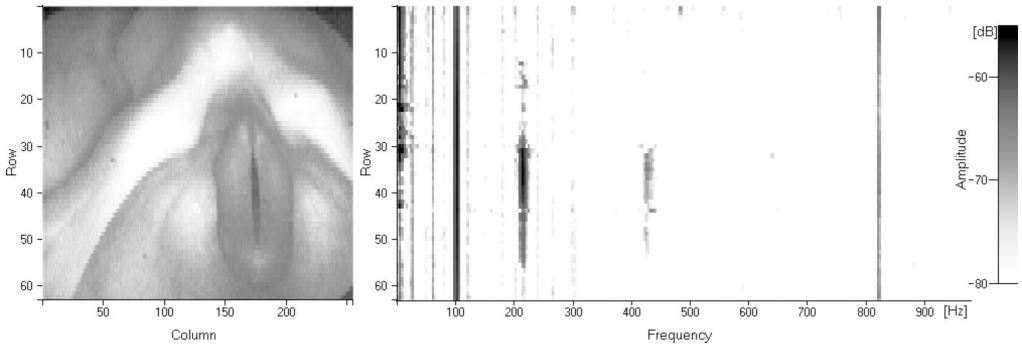

FIG. 4. Three-dimensional representation of spectra at lines rather than at single pixels of the image shown in Fig. 3. The *y* axis is identical with the one used in that image. The vertical bars refer to the oscillation frequencies of light intensity. The bar at 210 Hz corresponds to the vibration frequency of the vocal folds. The weak bar at 420 Hz reflects the second partial of the nonsinusoidal light oscillation at 210 Hz. The bar near 100 Hz is an artifact reflecting the flickering frequency of the light source used in this recording.

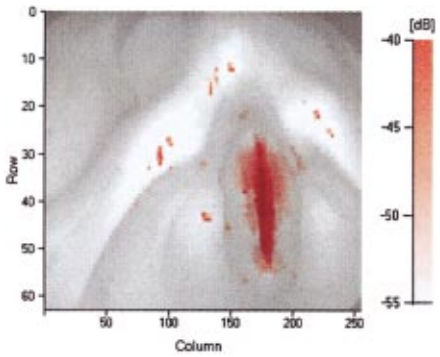

FIG. 5. Amplitude of oscillation, represented in terms of color saturation, at a selected frequency (210 Hz). The coloring has been superimposed on one of the frames from the image sequence.

surface leads to little intensity variation. On the other hand, glare spots will receive high amplitude, as they can move considerably even for a small tilt of the surface. The amplitudes of higher harmonics of the intensity oscillations seem relatively unimportant. To a large extent they are determined by the transfer function from motion to intensity, rather than by higher harmonics in the motion itself. Also, if the structures oscillate at several frequencies, combination tones can appear that are not necessarily present in the motion. However, if these limitations are kept in mind, the images seem quite useful for finding and displaying oscillatory movements in different parts of the larynx. Even though the method does not perform frequency analysis of the actual motion in the larynx, it can still be useful for detecting covibrations, e.g., in the ventricular folds or other structures close to the glottis. These vibrations might appear at the same frequency as that of the glottis, or at other frequencies.

Analysis of the movements of particular structures in an image would be advantageous. This, however, would require identification of such structures within the image. Automatic identification of components within an image is a complicated task, containing many sources of error. The Fourier analysis method can reveal oscillatory movements of the larynx, not obvious from direct inspection or kymography. Relevant information may emerge. For example, structures close to the larynx may oscillate at frequencies other than that of the vocal folds. Such oscillations may interfere with the folds' normal vibration but are not possible to visualize with standard stroboscopy. While high-speed imaging reveals such phenomena, analysis of the actual frequencies at which the structures oscillate is difficult. The case shown in Fig. 6 is an example of this, where vibrations at about a quarter of the fundamental frequency can be seen in the mucosa covering the arytenoid cartilages. The presence of this frequency is verified by the right kymogram in Fig. 7. Similar oscillations have been observed by Švec (2000) who used external excitation to study the resonance properties of the larynx. In this study, large oscillations in the arytenoid cartilages and ary-epiglottic folds were found in the range 50 to 75 Hz. The lower frequency in the present example is probably due to a different phonatory setting with more relaxed ary-epiglottic folds. It could be speculated that such an oscillation could disturb the glottal oscillation, especially near the phonation threshold. In models of vocal-fold vibration it is mostly assumed that the folds are attached to a rigid wall (see, e.g., Flanagan and Landgrat, 1968; Titze, 1973; Liljencrants, 1991; Wong *et al.*, 1991). If the structures to which the vocal folds are attached vibrate, such a model would fail to predict the behavior of the glottis. These types of oscillations are possible to visualize with kymography, but only at a single line of the image sequence. With the Fourier analysis method, the oscillations are visualized for one frequency at a time, but for the entire image.

The method also seems useful in printed representations of oscillations, where a moving playback of an image sequence is not available. Typically, entire sequences of images must be presented, image by image. Using the method presented here, the information contained in such sequences can be condensed to one or a few figures, particularly if a color scale is available.

The method presented here was applied to a limited material provided by one single subject with a healthy voice. The method, however, seems promising, as it offers information on relevant aspects of phonatory vibrations, which are difficult to detect by other means. It should be worthwhile to test the method on a larger material, including different types of phonation and clinical diagnoses.

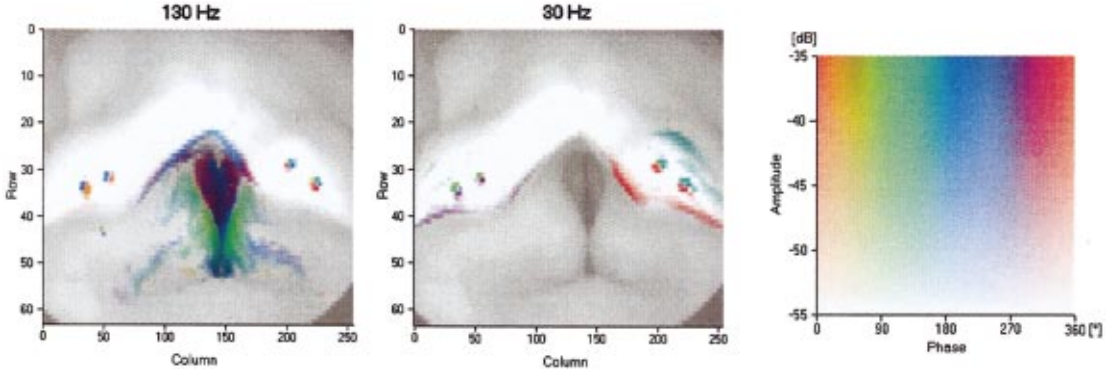

FIG. 6. Representations of intensity oscillations with different phases and covibrations. Phase relations are represented by color hue, and amplitude by color saturation. Left image: oscillations at 130 Hz. Oscillations in the vocal folds and in the ventricular folds are shown in different colors due to opposite phase in the intensity variations. Right image: oscillations at 30 Hz. The coloring reveals covibration in the mucosa covering the arytenoid cartilages.

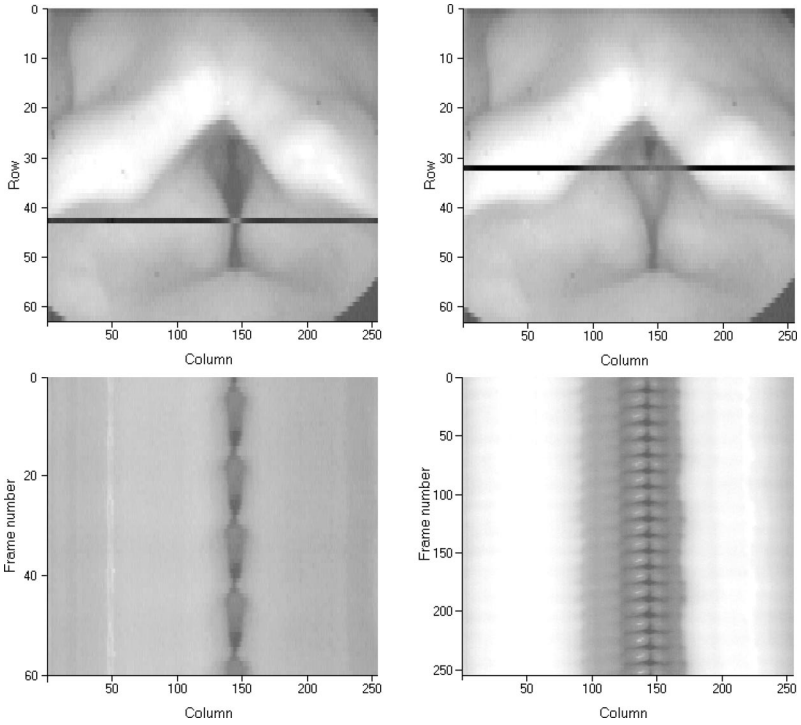

FIG. 7. Kymograms of the recording in Fig. 6. The left panel shows a kymogram at the ventricular folds. It can be seen that the closed phase of the glottis was synchronous with the maximum separation of the ventricular folds. The right panel shows a kymogram at the level of the arytenoid cartilages. On the right side of the kymogram, a low-frequency oscillation can be seen, which is consistent with the findings in the right laryngeal image in Fig. 6.

ACKNOWLEDGMENTS

We would like to thank Johan Sundberg and Britta Hammarberg for valuable input and editorial assistance. We would also like to thank Hans Larsson for valuable discussions and assistance with high-speed data handling. This work was supported by research grants from the Swedish Council for Work Life Research and the Swedish Research Council for Engineering Sciences.

Dunker, E., and Schlosshauer, B. (**1964**). ''Irregularities of the Laryngeal Vibratory Pattern in Healthy and Hoarse Persons,'' *Proceedings of Research Potentials in Voice Physiology*, edited by D. W. Brewer (State University of New York), pp. 151–184.

Eysholdt, U., Tigges, M., Wittenberg, T., and Proschel, U. (**1996**). ''Direct evaluation of high-speed recordings of vocal fold vibrations,'' Folia Phoniatr. Logop. **48**(4), 163–70.

Flanagan, J. L., and Landgraf, L. L. (**1968**). ''Self-oscillating source for vocal tract synthesizers,'' IEEE Trans **AU-16**, 57–64.

Hammarberg, B. (**1995**). ''High-Speed Observations of Diplophonic Phonation,'' in *Vocal Fold Physiology, Voice Quality Control*, edited by O. Fujimura and M. Hirano (Singular, San Diego), pp. 343–345.

Hess, W. (**1983**). *Pitch Determination of Speech Signals* (Springer, New York).

Kiritani, S., Imagawa, H., and Hirose, H. (**1988**). ''High-Speed Digital Image Recording for the Observation of Vocal Cord Vibration,'' in *Vocal Physiology: Voice Production, Mechanisms and Functions*, edited by O. Fujimura, pp. 261–269.

Kiritani, S. (**1995**). ''Recent Advances in High-Speed Digital Image Recording of Vocal Cord Vibration,'' Proceedings of International Congress of Phonetic Sciences **4**, 62–67.

Köster, O., Marx, B., Gemmar, P., Hess, M., and Künzel, H. J. (**1999**). ''Qualitative and quantitative analysis of voice onset by means of multidimensional voice analysis system (MVAS) using high-speed imaging,'' J. Voice **13**, 355–374.

Larsson, H., Hertegård, S., Lindestad, P.-Å., and Hammarberg, B. (**2000**). ''Vocal Fold Vibrations: High-Speed Imaging, Kymography and Acoustic Analysis,'' Laryngoscope **110**, 2117–2122.

Liljencrants, J. (**1991**). ''A translating and rotating mass model of the vocal folds,'' STL-QPSR, KTH, Stockholm 1/1991, 1–18.

Moore, P., White, F., and von Leden, H. (**1962**). ''Ultra-high speed photography in laryngeal physiology,'' J. Speech Hear Disord. **27**(2), 165–171.

Švec, J. G., and Schutte, H. K. (**1996**). ''Videokymography: High-speed line scanning of vocal fold vibration,'' J. Voice **10**, 201–205.

Švec, J. G. (**2000**). ''On Vibration Properties of Human Vocal Folds,'' Doctoral thesis, University of Groningen, The Netherlands.

Tigges, M., Wittenberg, T., Mergell, P., and Eysholdt, U. (**1999**). ''Imaging of vocal fold vibration by digital multiplane kymography,'' Comput. Med. Imaging Graph. **23**(6), 323–330.

Titze, I., and Liang, H. (**1993**). ''Comparison of *F*0 extraction methods for high-precision voice perturbation measurements,'' J. Speech Hear. Res. **36**, 1120–1133.

Titze, I. (**1973**). ''The human vocal cords: A mathematical model. I,'' Phonetica **28**, 129–170.

Titze, I. (**1974**). ''The human vocal cords: A mathematical model. II,'' Phonetica **29**, 1–21.

Wong, D., Ito, R. I., Cox, N. B., and Titze, I. R. (**1991**). ''Observation of perturbation in a lumped-element model of the vocal folds with application to some pathological cases,'' J. Acoust. Soc. Am. **89**, 383–394.