Phonovibrography: The Fingerprint of Vocal Fold Vibrations

Michael Doellinger, Ulrich Eysholdt, Joerg Lohscheller

University Hospital Erlangen, Medical School, Department of Phoniatrics and Pediatric Audiology, Bohlenplatz 21, 91054 Erlangen, Germany

ABSTRACT

A new approach for quantitative and objective analysis of vibration patterns of the vocal folds is presented. Vocal fold vibrations are recorded during sustained phonation in real time using a digital high-speed system coupled to a rigid endoscope. Within these high-speed movies, the edges of the oscillating vocal folds are detected by an adapted region-growing algorithm. The used segmentation algorithm is evaluated regarding its stability and correctness within 372 recordings proofing its clinical applicability. From the segmented vocal fold edges, a two-dimensional color encoded matrix is performed, visualizing the oscillations of the entire vocal fold edges over time. These created so called Phonovibrograms allow a direct and objective view on the occurring laryngeal dynamics. The Phonovibrograms can be used for assessing therapy progress and as a basis for further classification of different types of dysphonia. Phonovibrograms have the potential to serve as the first standard in endoscopic high-speed evaluation of vocal fold dynamics.

Index Terms— Biomedical Image Processing, Biomedical Signal Processing, Biomedical Recording, Medical diagnosis

1. INTRODUCTION

The analysis of vocal fold dynamics is an important issue in the field of laryngeal high-speed imaging. Since properties of vocal fold dynamics can not be quantitatively obtained by visual inspection, computer-aided systems are required. Multiple approaches are proposed for the extraction of laryngeal dynamics which focus mainly on the description of the glottal area function, the types of glottal closure [1,2], or the trajectories of vocal folds [3,4]. However, up to now there is no common method for the investigation of vocal fold dynamics enabling a full analysis of the entire two-dimensional vocal fold dynamics.

Within this work we present an analysis technique - Phonovibrography – which comprises an image processing procedure and a multidimensional visualization strategy. To guarantee that the vocal fold characteristics are effectively visualized, the outcomes of the image processing are evaluated within an extensive study. The results of the visualization strategy expose that the two-dimensional illustration of vocal fold vibrations (Phonovibrogram) contains the entire dynamic information inevitable for further investigations.

2. METHODS

The analysis of the dynamic properties of vocal fold oscillations depends essentially on the outcome of image processing procedures. These algorithms aim to extract the relevant movement information captured within laryngeal high-speed movies. All subsequent analytical methods are highly dependent on the accurate results of the image processing algorithms [5,6].

2.1. Recording of the vibrating vocal folds

Vocal fold recordings were performed during sustained phonation. Each subject was instructed to phonate a vowel /ae/ as in ‘hat’ in a comfortable way. To record the vocal fold vibrations of up to 250 Hz in real time, a digital high-speed camera (High-Speed Endocam, Wolf Corp., Knittlingen, Germany) coupled to a rigid endoscope was used, Figure 1. In our department, the high-speed camera system allows recordings of up to 4000 frames per second and a grayscale resolution of 256x256 pixel. The illumination of the larynx was facilitated using a 250 W xenon light source. In Figure 1, a single frame within a recording is exemplarily shown exhibiting the glottis and both vocal folds.

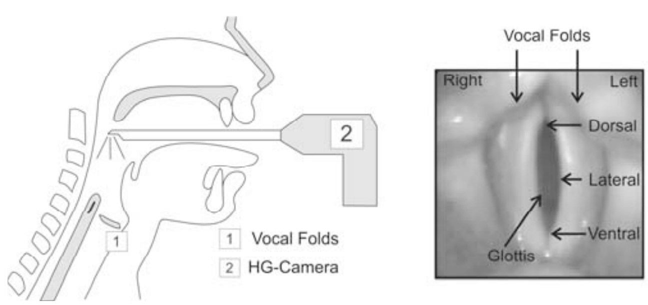

Figure 1: Scheme of the endoscopic high-speed recording procedure of the oscillating vocal folds (left) and one frame of the recorded vocal folds (right).

2.2. Clinical Evaluation of Image Processing

In our department we developed an image processing procedure containing an adapted region-growing approach [7]. This allows to identify the relevant structures during phonation, i.e. the oscillating edges of the vocal folds and therefore the alternating glottis. Only minimal user interaction is required to extract the vocal fold contours from high-speed movies even for poor image quality.
The outcome of this procedure was evaluated within an extensive study comprising 372 high-speed movies each composed of 500 succeeding frames. These recordings were randomly chosen from our data pool of several thousands of movies. For segmentation a performance of 100 images per second could be achieved by using a Pentium®4 (3.00 GHz) personal computer. From all recordings the segmentation quality was assessed by a committee of ten clinical and voice experts. In a total of 600 single raw images the experts marked manually the lateral, dorsal and ventral contours of the vocal folds, Figure 2. The resulting 24.000 points from the experts were set into relation to the computed outcomes of the segmentation procedure.

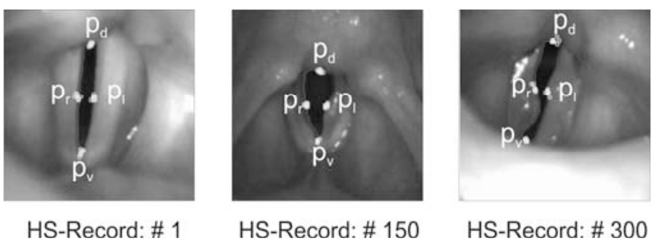

Figure 2: Expert labels superimposed on image processing results shown for three different high-speed recordings.

2.3. Phonovibrogram: A two-dimensional visualization of vocal fold vibrations

By superimposing the high-speed movies with the detected contours of the vocal folds their movement can be visualized. Hover, this kind of visualization enables only a successive inspection of the oscillations. In the following a multidimensional method is presented to visualize the segmented vocal fold contours of an entire high-speed movie within one single image, a so-called Phonovibrogram (PVG):

1. For each image the distances of the segmented vocal folds to the glottal main axis are computed (Figure 3, left).
2. The distances are virtually turned around the posterior point P (Figure 3, left) and stored within a column vector. The entries of this vector are then color coded (Figure 3, middle). The intensity of the color - normally red - encodes the magnitude of a positive distance to the glottal axis while the color blue encodes a negative distance, which occurs if a vocal fold crosses the glottal axis. Black encodes zero distance to the glottis axis.
3. The obtained column vectors of all images of the high-speed movie are fused to a data matrix D (Figure 3, right) which is called Phonovibrogram.

This procedure was applied to all segmented high-speed recordings within the study.

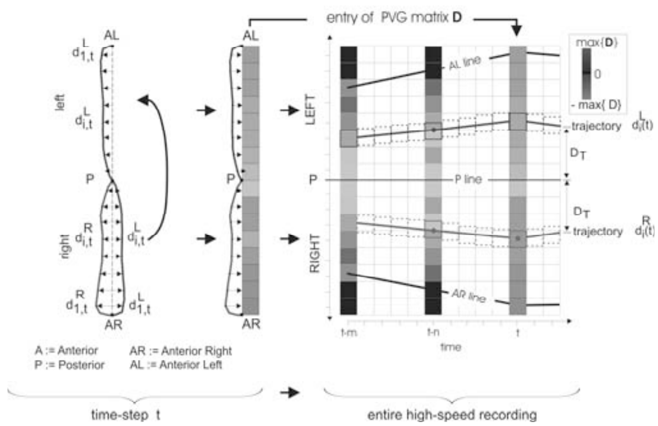

Figure 3: Deriving a Phonovibrogram (PVG) from a segmented high-speed movie of the vocal folds.

3. RESULTS AND DISCUSSION

The analysis of the data revealed that the achieved accuracy of the expert group was 3.2 +/- 1.5 pixel. In contrast the accuracy of the region-growing algorithm for segmentation of the vocal fold edges in reference to the expert points was only 2.5 +/- 2.1 pixel which confirms the performance and applicability of the algorithm.

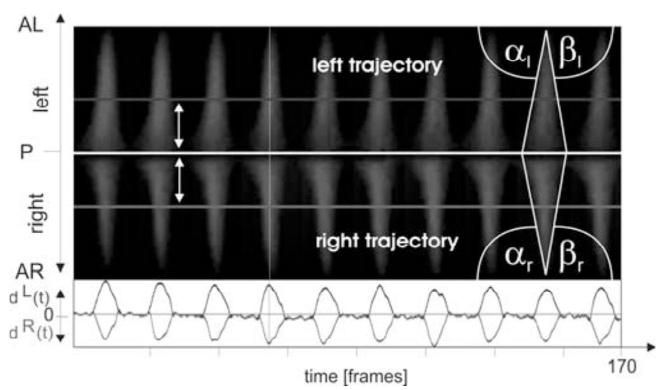

Figure 4: Phonovibrogram of a female subject, 22 years, with no clinical signs of voice disorders.

The upper part of Figure 4 exemplarily shows a PVG computed from vocal fold vibrations of a female subject, 22

years, with no clinical signs of voice disorders. Here, a section of 170 frames (10 oscillation cycles) is presented which corresponds to 42.5 ms. The fundamental frequency of the oscillations was 235 Hz.
In the given grayscale representation, the black area within PVGs denotes glottal closure. The transition from black to bright represents the opening of the glottis, i.e. the vocal folds detach from each other. The bright area corresponds to an open glottis. The following transition from bright to black is the closing of the glottis, i.e. the vocal folds get in contact again and the phonatory cycle restarts. The angles α (opening), β (closing) reflect the phase delay between anterior and posterior part of the vocal fold vibrations. A 90° angle corresponds to a simultaneous opening or closing from anterior to posterior.

The pattern of vocal fold oscillations within this specific PVG (Figure 4) possess a regular Triangular-Shape as marked by the white contours. The PVG exhibits symmetric (left - right) vocal fold oscillations confirming the diagnosed healthy voice. The opening starts posterior and continues in a linear phase delay to the anterior part. This can be identified by the temporal earlier transition from black to gray at posterior compared to the anterior part. This behavior is described by the angles $\alpha_l, \alpha_r < 90°$. The closing process is vice versa, which begins anterior and continues again with a linear phase delay towards posterior. This is described by the angels $\beta_l, \beta_r < 90°$. Within the posterior parts of the vocal folds the highest oscillation (bright) amplitudes occur. With increasing distance to the dorsal ending the vocal fold amplitudes diminish. The closing phase (dark) in the anterior area is much longer compared to the posterior position. At posterior, the vocal folds barely close (hardly black), which is often seen in female voices.
Also, the trajectories of the vocal folds can be extracted from the computed PVGs, since each line within the PVG corresponds to a certain point on the vocal fold edge. The lower graph within Figure 4 reveals vocal fold trajectories at medial position - indicated by the white arrows - of the PVG. The trajectories confirm the symmetry of the oscillations.

The structure of the PVGs instantly reflects the following dynamic behavior:

1. The pixel intensity within the PVG corresponds to the deflections in transversal direction.
2. The angel between anterior and posterior of the transition conditions from open to closed (β_l, β_r) and vice versa (α_l, α_r) give insight into the longitudinal movements and phase delays of the vocal fold.
3. The left-right symmetry of the oscillations can be seen by comparing the upper part (left vocal fold) of the PVG with the lower part (right vocal fold).
4. The regularity can be verified along the time axis for succeeding periods.

For healthy voice, five different basic types of PVGs for men and women could be identified within the presented study. The PVGs exhibited the shown Triangle- (Figure 4), V-, Rectangular-, Convex-, and a Concave-shaped geometry. V-Shape means, that the opening starts anterior and the closing starts posterior which is in contrast to the behavior of the Triangle-Shape. A Rectangular-Shape is given, when the opening and closing start simultaneous along the entire vocal fold edge. Convex means, that vocal folds mainly oscillate in the medial part. The Concave-Shape represents an oscillation, which is predominantly at dorsal and medial position.
Asymmetries and irregularities within PVGs mirror dysphonic subjects when the vocal fold oscillations are disturbed. Then, the symmetry of the geometric pattern for the left and right vocal fold is lost. Phase differences between the left and right oscillations occur.

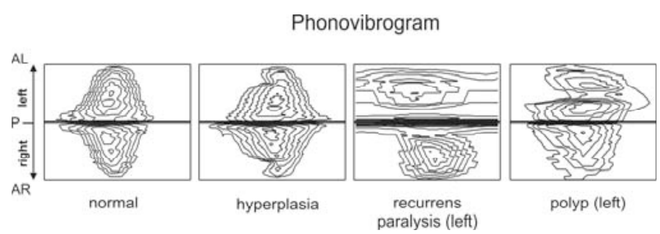

Figure 5: PVG-fingerprint: Isolines of a single PVG-oscillation cycle derived from four different subjects (normal voice, phonation hyperplasia, recurrens paralysis, and polyp).

The presented PVG visualization method summarizes within a two-dimensional image the characteristics of vocal fold vibrations and can be regarded as a finger-print of vocal fold vibrations.
Each image in Figure 5 shows a plot of isolines computed from a single oscillation cycle of four different PVGs representing a typical structure. The pattern of the PVG-cycles exhibits clearly the characteristics of vocal fold vibrations, see Figure 5. In normal voice the PVG is symmetric. In case of a hyperplasia a notch can be observed in the PVG for both vocal folds. The subject suffering from a recurrens paralysis shows on the healthy side (i.e., right vocal fold) a basis PVG type whereas the regularity is destroyed on the paralyzed side. In case of a polyp (i.e., left vocal fold) the PVG pattern is distorted where the polyp is located.

The presented stable and evaluated image processing technique in combination with the two dimensional visualization by PVGs will enable an objective evaluation of vocal fold dynamics in clinical practice and a documentation of diagnostics. The outcome of surgical or conventional voice therapy can be assessed by comparing the pre- and post- therapeutic PVGs.

In current projects, the geometric properties of the PVGs are exploited to classify vocal fold vibrations, e.g. using Hidden Markov-Models [8]. This will be used as objective diagnostic help to automatically differentiate between healthy and pathological voices and further to subdivide the different pathologies.

4. REFERENCES

[1] T. Wittenberg, M. Moser, M. Tigges, U. Eysholdt, "Recording, processing and analysis of digital highspeed sequences in glottography," *Mach Vision Appl*, vol. 8, no. 12, pp. 399-404, 1995.

[2] Q. Qiu, H.K. Schutte, L. Gu, Q. Yu, "An automatic method to quantify the vibration properties of human vocal folds via videokymography," *Folia Phoniatr Logop*, vol. 55, no. 3, pp. 128-136, 2003.

[3] R. Schwarz, U. Hoppe, M. Schuster, T. Wurzbacher, U. Eysholdt, J. Lohscheller, "Classification of Unilateral Vocal Fold Paralysis by Endoscopic Digital High-speed Recordings and Inversion of a Biomechanical Model," *IEEE Trans Biomed Eng.,* vol. 53, no. 6, pp. 1099-1108, 2006.

[4] M. Döllinger, U. Hoppe, F. Hettlich, J. Lohscheller, S. Schuberth, U. Eysholdt, "Vibration parameter extraction from endoscopic image series of the vocal folds," *IEEE Trans. Biomed. Eng*., vol. 49: pp. 773 – 781, 2002.

[5] J. Lohscheller, M. Döllinger, M. Schuster, U. Eysholdt, U. Hoppe, "Quantitative Investigation of the Vibration Pattern of the Substitute Voice Generator," *IEEE Trans Biomed Eng.,* vol. 51, no. 8, pp. 1394-1400, 2004.

[6] Y. Yan, K. Ahmad, M. Kunduk, D. Bless, "Analysis of vocal-fold vibrations from high-speed laryngeal images using a Hilbert transform-based methodology," *J Voice*, vol. 19, no. 2, pp. 161-175, 2005.

[7] J. Lohscheller, R. Schwarz, T. Wurzbacher, H. Toy, U. Eysholdt, "Objective Evaluation and quantitative analysis of high speed recordings in unilateral vocal fold paralysis," *Forum Acusticum*, Budapest, 2005, In Proceedings (CD).

[8] G. A. Fink, *Mustererkennung mit Markov-Modellen*, Teubner Verlag, Wiesbaden-Germany, 2003.