Moment based kymographic analysis of voice pathologies resulting in severe hoarseness

By

***R. Patel, +L. Liu, *D. Bless, and +N. Galatsanos**

*Department of Surgery
University of Wisconsin
Madison, WI 53705

+Dept. of Electrical and Computer Engineering
Illinois Institute of Technology
Chicago, IL 60616

ABSTRACT

Introduction

Currently there is no assessment tool that can quantitatively differential diagnose organic vocal fold pathologies resulting from severe vibratory disturbances. Wittenberg et al (1997) has documented that functional disorders can be diagnosed using high speed digital imaging. The aim of this project is to gain a better understanding of vocal fold vibrations associated with severe voice quality disturbances; resulting from various organic voice pathologies. In this study severe disturbance in voice quality is defined as an overall voice quality that receives a rating of type III or type IV voice disturbance (Titze, 1995). The above aim will be accomplished in this study by using kymographic analysis of data obtained from high speed digital imaging to differentially diagnose vocal fold pathologies based on the resulting vibratory disturbances using the classical Body-Cover theory (Hirano, 1981) of vocal fold vibrations. The *cover*, consisting of squamous epithelium and superficial and intermediate layers of lamina propria, is very pliable and can propagate a wave but has no contractile properties. The *body*, consisting of the deep layer of lamina propria, and vocalis muscle contributes to vocal stiffness by active contraction. The body cover relationship can be observed using high speed motion imaging and has important implications in understanding the vibratory characteristics of vocal folds in cases with severe voice quality disturbances.

A kymogram is a spatio-temporal image that shows a fixed horizontal line in the vocal fold image over time. Since the successive line images are presented in real time, the system makes it possible to observe left-right asymmetries, open quotient, propagation of mucosal waves, and anterioposterior modes of vocal fold vibration. Tigges et al (1999) used multiplane kymography to reveal the anterior posterior modes of vocal fold vibration and also to investigate different types of phonatory onsets.

We have used enhanced image processing and moment based features, which are rotational and translational invariant to compare across types of pathology. Moment based features have been used extensively for image recognition applications (Hu, 1962), but its application to study aperiodic vocal fold vibrations is unique to this study. In speech moment based features have been used to characterize the aperiodic noise during fricative production (Forrest et al 1988). In another similar study we have used moment based features across age, gender and tasks to detect the vibrational characteristics across subjects with normal and disordered voice qualities.

Methods and Analysis

In this preliminary analysis data was obtained from 35 pathology subjects. Subjects, who were rated as having and overall grade of type III or type IV voice quality, by a certified speech language pathologist, were included in this study. Based on the body-cover theory of vocal fold vibration, 35 pathology subjects were further classified into 5 subjects with mass lesions, 7 subjects with pathology affecting the body / deep layer of the vocal fold (vocal fold paralysis), 5 subjects with stiffness resulting from superficial layer of vocal fold (sulcus vocalis), and 18 subjects with stiffness resulting from injury to the superficial and intermediate layers of vocal fold (vocal fold papilloma).

After a thorough case history, high speed endoscopic visualization of subject's vocal fold was performed using a rigid seventy degrees endoscope. Kay Elemetrics high speed video system, model 9700, was used. It records images at 2000f/p/s for two seconds (i.e. 4096 frames) at 384 Mb/sec into very high speed video RAM. Simultaneous acoustic and EGG signals were also recorded. Audio signal was sampled at 50kHz. The gray scale image resolution was 160 x 140 and a 300-Watt xenon light source was used. Recordings were obtained under three phonatory conditions: normal pitch normal loudness (NPNL), high pitch normal loudness (HPNL) and normal pitch loud (NPL) for each subject. For this preliminary analysis phonatory condition NPNL was used for further analyses.

These high speed video recordings were used to create kymographs. Kymographs were created based on a 'single' point, whose location the experimenter was able to select. For the purposes of this analysis, the midportion of the membranous vocal folds was selected for the location of the horizontal scan line for the generation of the kymographs. This point is used to determine the 'y' value of the scan line. As a first step processing, central horizontal lines from each video frame were extracted. The extracted lines are then concatenated. A few hundred subsequent scan lines visualize the vocal fold dynamics. The resulting gray scale arrays (kymographs) showed the change in distance between edge points on left and right vocal folds during phonation.

From the kymographs seven moment based features were extracted based on the combinations of U (n, m) where $m + n = 3$. Moments represent statistical features of the intensities (gray levels) of the image (kymographs). These moment based features are rotational and translation invariant. Because of these invariant properties these features are used to describe and characterize different shapes in images. Analysis was performed on 200 frames for each sample. Since the invariant moment based features were infinitely small, natural logarithm based features were used for comparisons among the various categories.

Eigen vectors in two dimensions were generated based on moment1 and moment 2 values, as 89% of the variance was accounted by the above two moments. The first three moment values accounted for 99% of variance. Graphical representation of the various pathology groups was similar for the two- and three-dimensional figures. For ease of visualization a two dimensional representation is shown in Figure 1. Eigen vector 1 (ordinate) and eigen vector 2 (abscissa) generated from the kymographs of 35 subjects with severe disturbances in voice quality are depicted in Figure 1. In order to depict the various groups, circle fits based on the means and standard deviations for each group were made. Subgroups among severe voice pathologies are clearly visible. These can be attributed to the resulting vibration disturbances among pathologies. Using the Body – Cover theory of classification of various pathologies, high speed digital imaging can be used to gain additional insights into the vibratory disturbances among pathologies affecting the layered structure of the vocal folds.

References

Hirano M (1981). Clinical examination of voice, Edn 5, Berlin: Springer-Verlag.

Hu M. K. (1962). Visual Pattern Recognition by Moment Invariants, *IRE Trans. Info. Theory*, 1, 179-187.

Forrest K, Weismer G, Milenkovic P, Dougall R N (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of Acoustical Analysis of America*, 84 (1), 115-123.

Tigges M, Wittenberg T, Mergell P, Eysholdt U. (1999). Imaging of vocal fold vibration by digital multi-plane kymography. *Computerized Medical Imaging and Graphics*, 23, 323-330.

Titze I. R. (1995). Workshop on acoustic voice analysis, summary statement. Published by NCVS, Wendell Johnson Speech and Hearing Center, The University of Iowa, Iowa, 52242.

Wittenberg T, Mergell P, Tigges M, Eysholdt U. (1997). Quantitative characterization of functional voice disorders using motion analysis of high speed video and modeling. *IEEE Transactions*, 10, 1663-66.

Fig 1: Types of pathology