High-speed Laryngeal Imaging Analysis of Vocal Fold Dynamics

Yuling Yan, Xin Chen, Kartini Ahmad+ and Diane Bless+*

* Corresponding Author
Department of Mechanical Engineering
University of Hawaii – Manoa
yyan@hawaii.edu

+ Department of Surgery
University of Wisconsin - Madison

Abstract

High-speed laryngeal Imaging (HSLI) provides a direct means to observe the actual vibration of the vocal folds and allows for comprehensive analyses of vocal fold dynamics in normal and abnormal voice productions. Glottal edge detection is a key operation for tracing vibrations of the vocal folds at spatial locations and obtaining glottal area waveform (GAW) from the HSLI data. We have developed methods for glottal edge detection that delineates the vocal folds on a frame-by-frame basis. Here we present analyses of representative clinical HSLI recordings of normal and pathological voices during sustained vowel phonations. A new approach for the analysis of vocal fold vibratory patterns and dynamic characteristics, described in Yan et. al. (2004), is applied to the GAW extracted from the HSLI data. For comparison, the same approach is applied to the analysis of acoustic data that are simultaneously acquired with the HSLI recording.

1. Introduction

Laryngeal imaging provides a direct means to observe the vocal fold vibration - high-speed digital imaging technique has made it possible to resolve the actual vibrations of the vocal folds (*Honda et al. 1987; Hirose et. al. 1988; Kiritani, 2000*) and therefore presents an opportunity for comprehensive analyses of glottal source dynamics based on direct imaging of the vibrating vocal folds (*Yan et al. 2004*). Since the mechanism and properties of voice production are ultimately dependent on the dynamics of the vocal fold (VF) vibration, analysis of high temporal and spatial resolution images of vibrating vocal folds may provide new information on the mechanism of voice production and how specific differences in the vibratory characteristics relate to voice pathologies.

Glottal edge detection is a key operation for tracing VF vibrations at spatial locations and obtaining glottal area waveform (GAW) or glottal width function etc. A challenge in HSLI is to effectively manage and process the large image data files (4000 image frames for a 2-second HSLI recording). In this paper, we propose a simple and practical method to facilitate fast glottal edge detection on a frame-by-frame basis. The method involves selection of a region of interest and applying thresholding method to segment the VF opening region either interactively or automatically.

Representative clinical HSLI recordings of normal and abnormal voices including VF stiffness from recurrent respiratory papillomatosis (RRP) and muscle tension diplophonia are analyzed. Since acoustics is the most widely studied of all voice measurements (*Baken & Orlikoff*), we also present the analysis of the simultaneously acquired acoustic data to compare with and evaluate the HSLI based analysis.

2. HSLI analysis of VF vibrations

A diagram summarizing the HSLI processing and the VF vibration analysis procedure are shown in Fig.1**.** The HSLI system (Kay Elemetrics) acquires images at a rate of 2000 frames/second with a spatial resolution of 160x140 pixels.

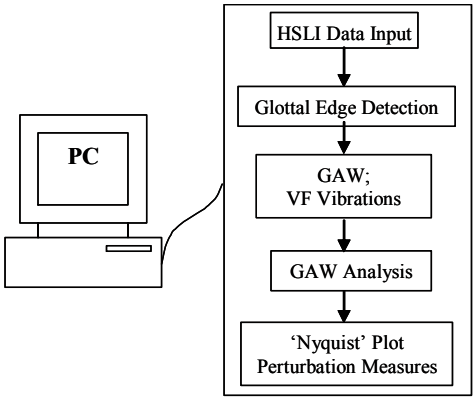

Figure 1 *HSLI analysis procedure*

2.1 Glottal edge detection

We developed an image segmentation method to delineate the vocal folds from HSLI data to obtain the GAW and VF vibrations at specific locations. Image segmentation techniques can be categorized into three classes: 1), characteristic feature thresholding; 2) edge detection; and 3), region exaction. The large size of HSLI files dictates the use of the simplest thresholding method for rapid segmentation of the region of VF opening. However in many laryngeal images, the gray value distributions of the object (region of the VF opening) and the background are indistinct; as a result, most existing methods such as Otsu method (*Otsu, 1979*) for automatic threshold value determination are unsuitable for our purposes. The method we developed adaptively determines the threshold value for each image frame based on the prior knowledge that the lowest gray level value pixel should lie within the region of the VF opening. Additionally, to optimize the computational efficiency and the accuracy of segmentation, we restrict a region of interest (RI) for the thresholding operation. The RI can be manually selected or automatically determined based on an evaluation of the VF motion from the difference image sequence obtained from original image sequence and a selected reference image.

This approach is applied to the analysis of representative HSLI data – an example of the results of segmentation and the outlined glottal edge is shown in Fig. 2.

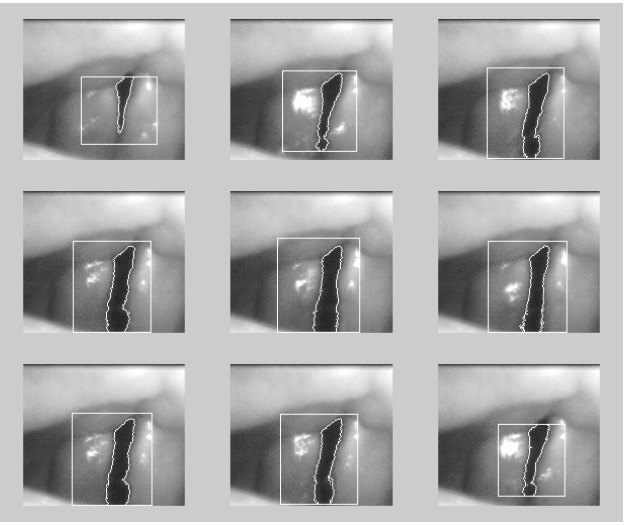

Figure 2 *Results of glottal edge detection; the rectangle outlines the automatically selected RI*

2.2 VF vibration analysis

Our new approach to VF vibration analysis involves constructing an analytic signal through the Hilbert transform and obtaining the analytic phase plot (Yan et. al, 2004), which is referred to hereafter as the 'Nyquist' plot. The 'Nyquist' plot reveals a comprehensive, standardized VF vibratory pattern and can be used to quantify dynamic characteristics of the VF while providing representative 'voice signatures' that are specific to voice quality and health.

2.2.1 Analytic signal and 'Nyquist' plot

An analytic signal is a complex function that has only positive frequency components. The Hilbert transform is ideally suited to construct an analytic signal (*Hahn, 1996*). In this approach, if a real function is assumed to be $x(t)$, a complex function is then constructed as: $\phi(t) = x(t) + jy(t)$; whose real part is $x(t)$ and the imaginary part, $y(t)$, is the Hilbert transform (HT) of $x(t)$, defined by: $y(t) = H\{x(t)\} = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x(\tau)}{\tau - t}d\tau$

Clearly, $y(t)$ is the convolution of $x(t)$ and a time function ($-\frac{1}{\pi t}$), which is a quadrature filtering processing that phase-shifts the input $x(t)$ by +/- 90 degree. For example, the HT of $\cos\phi(t)$ is $\sin\phi(t)$, and the analytic signal is a complex sinusoid $e^{j\phi(t)}$, and the analytic phase trace plot, or 'Nyquist' plot, is therefore a unit circle.

The complex analytic signal and the analytic phase information are exploited to map an instantaneous trace representing VF dynamics evolving in time; as the analytic phase varies 360 degree, the real signal completes an oscillatory cycle in the time course. If the periodicity is reached after this oscillation cycle, the trace pattern repeats itself from period to period. The trace pattern reveals the intra-period waveform property; a round circle corresponds to sinusoidal waveform; while the degree of shape distortion from a circle indicates the harmonic distortion. If a time signal is quasi-periodic (nearly periodic), the inter-cycle variations will show as slight scattering which can be quantified to provide amplitude and frequency variability of the VF oscillations (see Yan et. al, 2004 for details).

Overall, the 'Nyquist' plot provides a comprehensive display of the vocal system's dynamic behavior, and the plot reveals the regularity and synchronization of the vocal system oscillations.

3. Results of Analyses from Clinical Voice Samples

Clinical HSLI recordings representing normal and pathological voices are analyzed to demonstrate the effectiveness of our analytical approach and methods for the HSDI image processing and VF vibration analysis

represented by GAW analysis. The voice pathologies included vocal fold stiffness from recurrent respiratory papillomatosis (RRP) and muscle tension diplophonia. The samples were selected from the midportion of the HSLI recording during the phonation of a sustained vowel /i/ to avoid any vocal onset or termination effects. The Kay high-speed videoendoscope system was used for simultaneous recording of laryngeal images and acoustic data. Comparisons between the HSLI based GAW analyses and acoustic analyses are presented in the form of Nyquist plots.

3.1. VF Vibration in normal voices

Figure 3 shows the GAW extracted from HSLI recording of a normal voice during a sustained phonation of /i/ at normal pitch normal loudness (npnl). For comparisons, the simultaneously acquired acoustic signal is shown in Fig. 4 – the acoustic signal contains information on both glottal source dynamics as revealed in GAW and vocal tract transfer function.

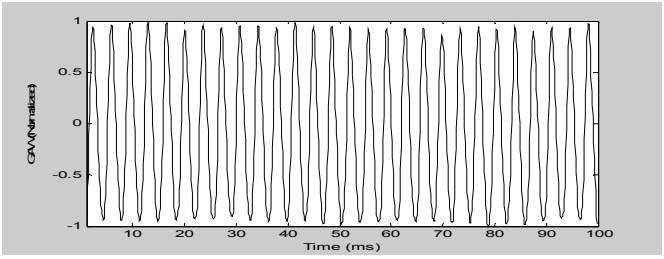

Figure 3 *GAW (normalized) extracted from HSLI recording (100 ms) of a normal voice*

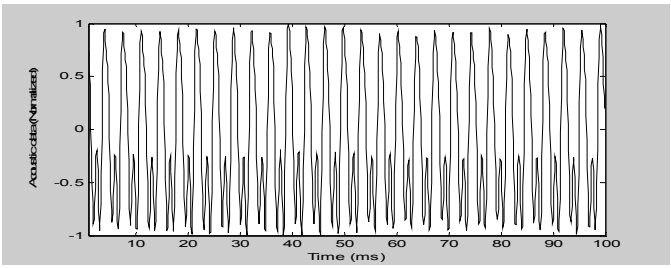

Figure 4 *Acoustic data (normalized) simultaneously acquired from the normal voice*

The Nyquist plots obtained from GAW and acoustic analyses are shown in Fig. 5, both indicate slight cycle-to-cycle scatter, or near periodicity of the VF vibration. In all analyses, the GAW (sampling frequency of 2kHz) and acoustic data (sampling frequency of 50kHz) were resampled and anti-aliasing filtering (low-pass at 5kHz) was applied to acoustic data prior to the analyses.

3.2. VF Vibrations in pathological voices

Figure 6 (upper) shows the GAW extracted from HSLI recording from a RRP patient exhibiting VF stiffness during a sustained phonation of /i/ (npnl). The simultaneously acquired acoustic signal is shown in Fig. 6 (lower), which correlate well with the GAW extracted from the HSLI. Both signals show irregularity of the VF vibratory behavior. This is nicely revealed in the Nyquist plots obtained from GAW and acoustic analyses (Fig. 7).

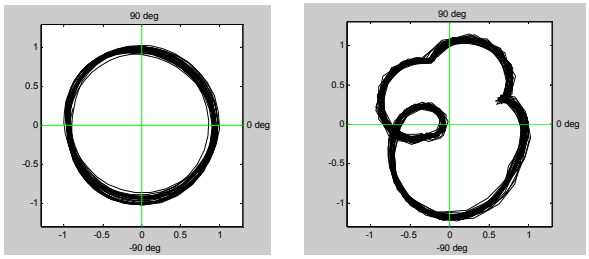

Figure 5 *Nyquist plot of the normal voice; Left: GAW analysis; Right: acoustic analysis*

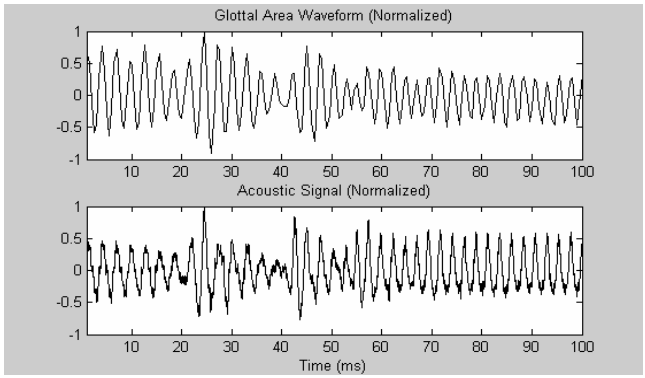

Figure 6 *Normalized GAW (upper) and acoustic data (lower) from the RRP patient*

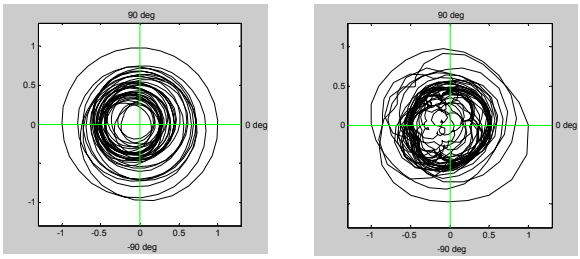

Figure 7 *Nyquist plots from the VF stiffness (RRP); Left: GAW analysis; Right: acoustic analysis*

Finally an HSLI recording of a non-modal phonation of sustained /i/ from a speaker with perceived diplophonia was analyzed. Fig. 8 shows the GAW (upper) extracted from HSLI and acoustic data (lower) of a segment of recording of 100 milliseconds (200 image frames). These vocal signals apparently show a near supra-periodicity. Three consecutive data sets each contains 100 ms recording are analyzed and the Nyquist plots from GAW (Fig. 9) and from acoustic data (Fig. 10) clearly show a transition from normal to diplophonia (from left to right).

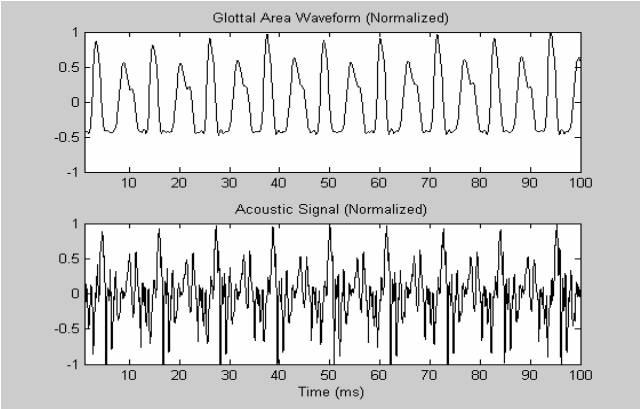

Figure 8 *Normalized GAW (upper) and acoustic data (lower) from a diplophonic voice*

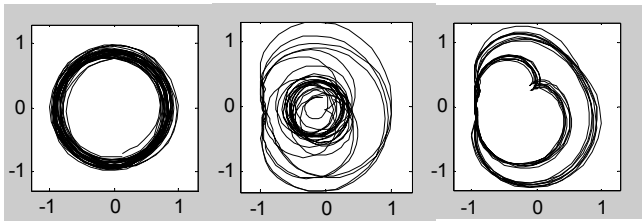

Figure 9 *Nyquist plots obtained from the GAWs* of the diplophonic voice recording *(*each contains 100 ms recording*)*

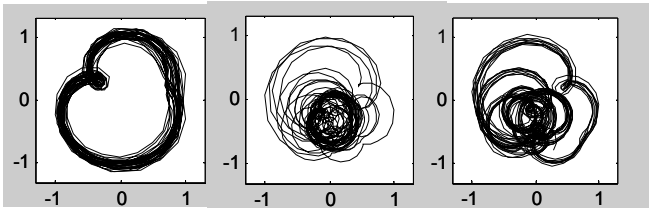

Figure 10 Nyquist plots obtained from the acoustic recording of diplophonic voice (each contains 100 ms recording)

4. Conclusions

We developed an adaptive image segmentation and thresholding method for the glottal edge detection in HSLI. From the extracted glottal edge, we can trace the GAW and VF vibrations at specific locations. In this study GAW is used to analyze the vocal dynamics as it correlates with voice quality. Representative clinical HSLI recordings of normal and abnormal voices exhibiting VF stiffness and muscle tension diplophonia were processed and the GAWs extracted from the HSLI were analyzed using the analytic phase trace approach ('Nyquist' plot) that we have developed (Yan et al. 2004). Analyses of the acoustic signals that are simultaneously acquired with the HSLI were also presented for comparisons and evaluations of our HSLI analysis procedure. The analyses based on imaging and acoustic measurement lead to well correlated results and both yield useful clinical information.

This study is significant because it shows that HSLI with our analytical approach generates comprehensive information on voice quality and vocal pathology – this information could help establish clinical protocols and measurement parameters for the differential diagnosis of voice disorders. Since HSLI provides direct images of the vibrating vocal folds it may help to explain variations obtained from the indirect acoustic measures and complement the acoustic analysis procedure.

5. Acknowledgment

This work is partly supported by a grant from National Science Foundation (BES 0402439) awarded to Yan.

6. References

[1] Honda K, Kiritani S, Iwagawa H, Hirose H., 1987. High-speed digital recording of vocal fold vibrations using a solid-state image sensor. In: *Laryngeal Function in Phonation and Respiration*, Baer T, Sasaki C, Harris KS, (eds.). Boston: Little, Brown and Co., 485-491.

[2] Hirose H, Kiritani S, Imagawa H., 1988. High-speed digital image analysis of laryngeal behavior in running speech. In: *Vocal Physiology: Voice Production*, Fukimura O. (ed.). New York: Raven Press Ltd.

[3] Kiritani S., 2000. High-speed digital image recording for observing vocal fold vibration. In: *Voice Quality Measurement,* Kent R D & Ball M. (eds.). Singular Publishing.

[4] Yan Y., Ahmad K., Kundak M., Bless D., 2004. Analysis of vocal-fold vibrations from high-speed laryngeal images using a Hilbert transform based methodology. *J. of Voice.* In press.

[5] Otsu N., 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9(1), 62-66.

[6] *Hahn, S L., 1996. Hilbert Transforms in Signal Processing.* Artech House.

[7] Baken R.J, Orlikoff R.F., 2000. *Clinical Measurement of Speech and Voice*. 2nd ed. Singular.