High-speed Imaging of Vocal Fold Vibration Onset Delay: Normal *Versus* Abnormal

Peak Woo, *New York, New York*

Summary: Objectives. Vocal fold vibration onset delay (VFVOD) is heard frequently in spasmodic dysphonia and in muscle tension dysphonia. VFVOD changes due to other vocal pathologies have not been investigated. VFVOD during sustained vowel production was estimated with high-speed video in 10 normal and 40 pathologic subjects (scars, vocal fold paralysis, vocal fold nodules, and polyps). Analysis of high-speed video was done using digital kymography.
Results. VFVOD can be divided into two portions. Pre-phonation delay (PPD) is the duration when the vocal folds are nearly approximated to the time of first observed oscillation. Steady state delay (SSD) is the time when vocal folds are observed to come into oscillation until steady state of oscillation is observed. Normal subjects have almost zero PPD with vocal fold oscillation observed before full vocal fold adduction. Pathologic cases showed prolonged PPD because of (1) false cord adduction, (2) prolonged true vocal fold adduction, and (3) delay to onset of vocal fold vibration. Normal subjects have SSD of three to five cycles before steady state. Pathologic states result in increased SSD. Causes for increased SSD include (1) slow ramping up to steady state, (2) partial vibration of vocal folds, and (3) diplophonia with alternating beats before achieving steady state. There are significant differences between normal and pathology groups in both PPD and SSD.
Conclusion. VFVOD is elevated in pathologic states. This can be due to increase in PPD or SSD. VFVOD is an under-recognized phenomenon that may contribute to complaints of vocal fatigue and dysphonia.
Key Words: High-speed video–Voice delay–Dysphonia–Video laryngoscopy–Speech disorder.

INTRODUCTION

High-speed cinematography was first reported in 1937.[1] With the recent availability of commercial high-speed laryngeal video (HSV) imaging systems, more information has been gained regarding details of vocal fold vibration not observable by stroboscopy.[2] Schutte et al were the first to present videokymography as a way to investigate all vibration abnormalities. They concluded that all vocal fold vibrations, including those with rough, breathy, hoarse, or diplophonic voice productions, can be observed.[3] Recently, there has been an increase in reporting of objective measurements of vocal fold vibratory function with the development of phonovibrograms.[4] Others have used high-speed laryngoscopy to study the vocal fold vibratory patterns using glottal ratio indices that include the full opened glottal area, glottal width, and glottal length.[5] All these measures seek to look at the cycle-to-cycle differences during sustained phonation and attempt quantitative studies. With image enhancement and edge extraction, it is possible to perform frequency analysis of the vibratory function from the HSV and obtain measures before and after treatment.[6]

These studies have focused on the steady state of vocal fold oscillation. One of the advantages of kymography and of HSV is their ability to study the vocal folds during voice onset. During initiation of phonation, the vocal folds are adducted toward midline, and vibration is typically observed before full glottis closure. There follows a series of glottal cycles until the vocal folds appear to vibrate in a quasi-periodic manner. The duration from when the vocal folds are brought into near approximation to steady state of vocal fold oscillation can be termed vocal fold vibration onset delay (VFVOD). VFVOD abnormalities have been noted in conditions of spasmodic dysphonia[7] and in patients with muscle tension dysphonia.[8] Excessive muscle activity of the extra-laryngeal muscles and intrinsic laryngeal muscles is thought to result in excessive laryngeal tension and voice delay. In our experience in doing HSV studies in patients with laryngeal pathologies, we have noted abnormalities in voice onset time. We postulate that patients with complaints of vocal fatigue and dysphonia may have difficulty in achieving steady state of target phonation because of abnormal VFVOD. This may be an important pathophysiology component of the dysphonic state that has not been previously studied.

In the past, HSVs of vocal fold vibration have focused on evaluation of the vibratory characteristics of the steady state. By recording the vibration of each glottal cycle, the differences between glottal cycles can be better differentiated. The main advantage is the visualization of aperiodic movements that otherwise would not be visible with stroboscopy.[9] Such an approach is critical in the understanding of period doubling voice disorders such as diplophonia.[10] The ability to look at cycle-to-cycle variations with HSV allows quantitative methods to characterize the stability of the steady state. Objective data derivations from HSV have been accomplished.[11]

Although evaluation of the steady state is valuable, the dynamics of the vocal folds during voicing is extremely versatile and worthy of study by HSV. Until recently, voice gestural changes of the vocal mechanism have largely not been evaluated by HSV. McDonnell et al used HSV examination to evaluate voicing onset of a small number of singers with different vocal gestures.[12] Such an approach has not been used to evaluate

Accepted for publication August 30, 2016.
This paper was presented as an oral presentation at the Voice Foundation Annual Meeting, Philadelphia, PA, June 5, 2016.
From the The Department of Otolaryngology Head and Neck Surgery, Icahn School of Medicine, New York, New York.
Address correspondence and reprint requests to Peak Woo, The Department of Otolaryngology Head and Neck Surgery, Icahn School of Medicine, 300 Central Park West 1-H, New York, NY 10024. E-mail: peakwoo@peakwoo.com
Journal of Voice, Vol. 31, No. 3, pp. 307–312
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.08.020

pathologic states. Voice onset abnormalities can be an area of importance in clinical laryngology and in understanding patients with complaints of dysphonia.

The purpose of this study is to estimate the VFVOD of 40 subjects with disparate pathologies seen in a general laryngology clinic and compare them with normal subjects. We hypothesize that VFVOD can be calculated objectively by reading the digital kymography (DKG) tracing and counting the number of video frames from the time the vocal folds come into adduction until the vocal folds have achieved full steady oscillation. Because the sampling time of the HSV is known, the number of video frames between vocal fold adduction to full steady oscillation can be calculated easily by frame counting. In this way, analysis can be done simply by reading the DKG plot.

MATERIALS AND METHODS

Forty subjects with voice pathologies were studied by HSV. Ten healthy subjects without pathology served as controls.

In the pathology group, there were 11 patients with unilateral paralysis or paresis; 9 with polyps; 11 with scar and sulcus; 5 with vocal nodules; and 6 with nodules and 3 other cases with one spasmodic dysphonia, one muscle tension dysphonia, and one presbyphonia.

The high-speed recordings were done using a 70 rigid endoscope (model 9106, PENTAX Medical, Montvale, NJ). The endoscope image is coupled to a black and white high-speed camera (Kay Elemetrics High-Speed Digital Imaging (HSDI) system, PENTAX Medical; Photron Motion, San Diego, CA). The HSV AVI video was analyzed by *Kay's Image Processing Software* (PENTAX Medical, Montvale, NJ). The software was used to generate the digital kymogram by placing a transverse line across the glottis at the mid-membranous portion of the vocal folds.

All subjects were examined by rigid laryngoscopy. Standard local lidocaine anesthesia (2%) was used. The patients were asked to phonate at modal phonation. In the modal task, subjects produced /i/ at a comfortable pitch and loudness. The patients were asked to sustain phonation for 2 seconds then inhale. During the repeated sustained phonation gesture, the high-speed camera was triggered to capture 8 seconds of data that include the onset, the steady phonation, and the end of the phonation token for at least two tokens.

HSV was acquired using Kay Elemetrics HSDI system (PENTAX Medical, Montvale, NJ). The laryngeal image was used to fill the screen in a set manner so as to standardize the scope to vocal fold distance. The scope was coupled with a 300-watt Xenon light source (PENTAX Medical, Montvale, NJ). The HSDI system acquired grayscale images at a rate of 2000 frames per second, with a spatial resolution of 256 × 120 pixels. The raw AVI file was saved for later analysis. The best tokens with a clear and full view of the larynx were saved onto the hard drive for analysis. All subjects tolerated the examination without any difficulties.

Data analysis

Figure 1 is a plot of the DKG tracing in a normal subject showing the onset of vocal fold vibration. The voice onset is characterized by continued vocal fold adduction as the folds are approximated. This is followed by just noticeable vocal fold vibration. After several

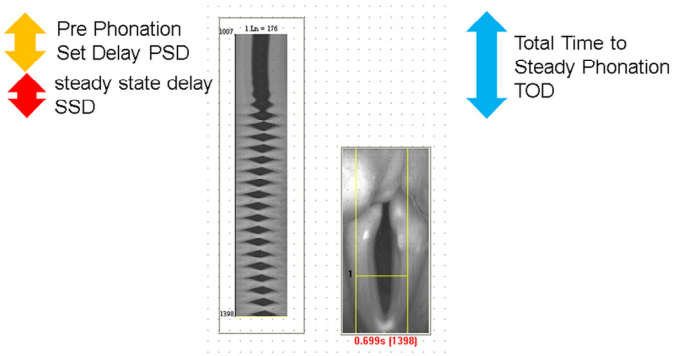

FIGURE 1. Schematic of calculation of PPD, SSD, and TOD from the DKG plot.

cycles, the steady state of vocal fold oscillation has been achieved. The frame number on the video when adduction approaches the width of one vocal fold is defined as the beginning of the frame for counting of the voice onset time. This typically corresponds to one grid mark on the DKG plot. The time from when the vocal folds are nearly adducted to just noticeable vocal fold vibration is termed the pre-phonation delay (PPD). The time from just noticeable vocal fold vibration to steady state of vocal fold oscillation is termed steady state delay (SSD). SSD time plus the PPD time was added to obtain the total onset delay (TOD) time. The TOD is the calculated time from DKG to correspond to VFVOD. For each subject, one or two phonation tokens were averaged to obtain the SSD, PPD, and TOD. The data were entered into a spreadsheet and analyzed by *SPSS* (Version 24, International Business Machines, Armonk, New York) statistical software. Comparison was made between two categories: pathology and normal. Additional analysis was performed for SSD, PPD, and TOD between pathology groups for each condition of normal, vocal fold paralysis, vocal fold polyp, vocal fold nodule, scar, and other.

RESULTS

There was no statistical difference between gender and age between the normal and the abnormal groups.

Figure 2 is the DKG plot of five of the normal subjects. In each subject, the vocal folds come into approximation, the vocal folds start to oscillate before full closure, and full vibration is achieved in three to five cycles after the first oscillation is noted. The only exception noted is the fourth figure with a prolonged PPD.

Figure 3 is a DKG plot from five of the abnormal subjects. Each graph is labeled with the pathologic condition. Notice that there is prolonged PPD. Unlike in the normal subjects, the prephonation time in the pathology group is often characterized by complete closure of the vocal folds with a delay from adduction of vocal folds to just noticeable vibration of the vocal folds. There is also the impression that full steady onset requires many more cycles to establish the steady state of quasi-periodic oscillation. This would result in an increase in the SSD. This results in a prolonged TOD time for the pathology group compared with the normal group (Figure 2 *vs.* Figure 3).

Whereas the PPD delay was not specific to the pathology, the SSD delay can show changes specific to the abnormal vibratory properties of the vocal folds. These abnormalities include