Objective Measures of Laryngeal Imaging: What Have We Learned Since Dr. Paul Moore

Peak Woo, *New York, New York*

Summary: Dr. Paul Moore pioneered the use of high-speed cinematography for observation of normal and abnormal vocal fold vibrations during phonation. His analysis of the glottal area waveform, opening and closing speed index, and open quotient from the high-speed films were labor intensive but relevant today. With advances in digital image capture and automated image extraction techniques, stroboscopy and high-speed images of vocal fold vibration may be analyzed with objective measures. Digital high-speed image capture in color is now clinically practical at high resolution. Digital kymography now allows analysis of the vibratory waveform from each vocal fold. Serial capture and comparison can document changes in vibratory function with treatment. Quantification of vocal fold vibration using such techniques is now practical. This is a review of vocal fold vibration capture and analysis techniques since Dr. Moore.
Key Words: G. Paul Moore–Imaging–Dysphonic–Vocal fold vibration–Subglottic pressure–Ultra high-speed photography–Stroboscopy–Video stroboscopy.

INTRODUCTION

There is often a significant discrepancy between the subjective complaints and objective findings in dysphonic patients. Many patients have perceptually dysphonic voices but lack visible pathology by mirror examination. More in-depth examination of vocal fold vibratory function may reveal pathologic evidence to explain the abnormal acoustic manifestations. Direct observation of the complex vibratory patterns of the vocal folds during phonation offers the clinician one of the best indications of laryngeal function. Normal vocal fold vibration produces a complex and intricate movement pattern, which depends on a combination of anatomical and physiological factors. These factors include the size and rheology of the vibratory membrane, adjustments of flow, subglottic pressure, and muscle tension. Because vocal fold oscillations occur at rates of 100–250 Hz during normal phonation, special methods are necessary to facilitate observations of vocal fold vibration. The two techniques are ultra high-speed photography[1] and stroboscopy.[2] With these techniques, the research investigator and the clinician alike can appreciate the complex details of vocal fold vibration. Attempts to understand and study this fascinating vibratory behavior have been the subject of great interest in the past and this interest continues today.

The principle of using flashing lights for examination of rapidly moving but periodic oscillation with stroboscopy has been well known for over a century. Today, videostroboscopy is one of the standard methods used to examine moving vocal folds. Laryngovideostroboscopy is used extensively for the analysis of vocal folds vibration to diagnose voice disorders. Since the initial book on videostroboscopy by Schönharl,[3] the technique and interpretation of laryngeal videostroboscopy have become well formalized. Stroboscopic signs associated with benign mucosal lesions can also be systematically rated and characterized.[4] Laryngovideostroboscopy can reveal abnormalities of laryngeal structure, absence of vibration, and vibratory asymmetry. The pliability of the vocal fold after surgical intervention can be assessed qualitatively by observation of the vibratory characteristics using stroboscopy.[5]

Objective measurements of the vocal fold vibration pattern date back to the initial high-speed cinematography recordings of vocal fold vibration from the works of Timcke, von Leden, and Moore.[6–9] Although the concept of objective measurements of vocal fold vibrations is attractive, the actual clinical application of objective measurements has not yet been achieved. This is because of the factors related to the complex vocal fold vibratory image. The use of high-speed image by cinematography was already published by 1936.[1,10] Dr. Moore had his first impact on vocal fold imaging in two publications in 1937 and 1938[11,12] and continued through publications that were based on collaboration with clinicians. It consisted of laborious work of frame-by-frame analysis of the high-speed cinematography image.

Despite the unique nature of these publications, there were few other investigators who embraced high-speed imaging (HSI) and analysis using high-speed cinematography. Unfortunately, the high cost of high-speed cinematography or the intensive labor associated with its analysis may have been the reasons why investigators were reluctant to adopt these new methods. Synchronized electronic flash stroboscopy became available and was embraced as a clinical tool.[13] After the publications by Moore and von Leden, HSI was not mentioned again for at least another decade. In 1976, Childers et al[14] pointed to the possible role of computers in image quantification to overcome the deficiencies associated with the manual mapping of the vocal fold image.

With the availability of digital image processing, automated image extraction and quantification are now possible. Honda and Hirose were the first to use a computer to capture digital images of vocal fold vibration with a charge-coupled device (CCD). They cited the practicalities and advantages of a digital approach in their analysis.[15] In a series of publications, Eysholdt[16,17] pointed to the advantages of high-speed motion

Accepted for publication February 6, 2013.
This study was presented at The Voice Foundation Annual Meeting; June 2, 2012; Philadelphia, Pennsylvania.
From the Department of Otolaryngology, Mount Sinai School of Medicine, New York, New York.
Address correspondence and reprint requests to Peak Woo, MD, Department of Otolaryngology, Mount Sinai School of Medicine, 300 Central Park West, 1-H, New York, NY 10024. E-mail: peakwoo@peakwoo.com
Journal of Voice, Vol. 28, No. 1, pp. 69-81
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2013.02.001

analysis of vocal fold vibration in detecting asymmetries, transients, breaks, and irregularities. The use of a line scanning camera by Schutte and Svec[18–21] allowed the clinician to query the vocal fold image at a single line at high speeds and it further added to the understanding of the need for in-depth, real-time analysis of the vocal vibratory function. The development of the digital kymography (DKG) line analysis, although not very intuitive for the clinician, did spur the growth of a clinical commercial system based on high-speed image acquisition. Such a system was already available in 2000. Today's commercial systems produce color, high-speed images of 2000 frames per second at 512×512 resolutions for 8 seconds. Such resolutions were unheard of just 20 years ago! HSI of vocal fold vibration and its clinical applications are now readily accessible within the time and cost constraints of a clinical practice.[22,23] Today, DKG from high-speed video images can be routinely captured, analyzed, and compared. Such rapid capture and analysis of the vibratory capability of vocal fold vibration was only a dream just a decade ago. Today's tools allow easy quantification of the vocal fold vibration image for study of normal laryngeal physiology as well as pathologic laryngeal states. They also offer objective measures of surgical and therapy techniques.

The goal of this paper was to summarize the objective measures that one can obtain from stroboscopy and high-speed video. I will reintroduce the concept of the glottal area waveform (GAW) that Dr. Timcke and his coauthors first described.[7] The role of automatic edge detection for automated numeric quantification will be presented. From the DKG tracing of the high-speed video, I will introduce the concept of kymography waveform analysis. Finally, I will give some clinical examples of how such analysis may be used in the laryngological care of patients with voice disturbances.

MEASUREMENTS FROM THE VIDEOSTROBOSCOPY

Videostroboscopy is a well-established technique in the clinical evaluation of the dysphonic patient. In the voice clinic, it is most useful for identifying small lesions and mass lesions and for verification of stiffness.[24] Laryngeal asymmetry during vocal fold oscillation is one of the most obvious abnormalities that one can identify during sustained phonation and points to asymmetric rheological changes in the vocal fold cover. Often, these asymmetrical changes are indicators for recommending surgical intervention.[25] After surgical intervention, videostroboscopy examination often shows improvement in the vocal fold edge, configuration, phase closure, and return of amplitude and mucosal wave.[26] Objective measures of vibratory capability based on videostroboscopy examination are more challenging. The single-flash-timing laryngeal videostroboscopy is difficult to standardize from examination to examination. Control of the patient's phonatory volume, frequency, and the size of the laryngeal image must be standardized if one wishes to compare the vibratory pattern before and after the treatment.[27] Some authors have recommended overlaying images from the prior examination over the current examination using a transparency tracing to standardize the distance of the endoscope from the vocal folds and the size of the laryngeal image.[28] Such an approach is most likely reserved for the research laboratory where specific information is obtained. Until more superior imaging techniques becomes available through the fiberscope, imaging and analysis of vocal fold vibration is largely limited to per oral endoscopy and image capture. This means that the subject must be able to tolerate per oral endoscopy with a limited amount of phonation token recorded with the tongue out and the rigid endoscope in.

Not all stroboscopy examinations are useful for imaging analysis as the stroboscopic image is a montage of many glottal cycles. One is assuming that the multiple frames that are analyzed from the video are produced at the same fundamental frequency and loudness. Using the standard stroboscopic flash rate of 1.5 Hz above the fundamental frequency, a video frame rate of 30 frames per second will result in a complete glottal cycle in 20 video frames. If the patient can hold a steady phonation for 2 full seconds at the same fundamental frequency and loudness, then a montage of video frames that is representative of the sustained phonation for that token can be acquired. The stroboscopic image from the assembled three glottal cycles is a montage of the vibratory pattern for the vocal folds for those 2 seconds. For a male, it would be approximately 250 glottal cycles, whereas for a female, it would be approximately 500 glottal cycles! Provided the waveforms are quasiperiodic and repeatable from cycle-to-cycle on the visual inspection of the stroboscopy video, this sequence can then be subject to analysis. Figure 1 is a montage of a glottal cycle obtained by capturing every frame of the video cycle and limiting the frame of interest to the vibratory margin. This is assembled as a single image made of many video frames with the area of interest. Even in the presence of pathology, when one compares this to other glottal cycles from the same stroboscopy token and finds that the glottal cycles look similar, one can then assume that the stroboscopic montage is representative of the actual glottal cycle.

If the glottal cycle is not evenly illuminated, a frame dropout will occur. This will result in a glottal cycle that is not suitable for automated image extraction. An inherent limitation of videostroboscopy is that not all patients can have the entire vocal fold margin visualized during videoendoscopy. This is because some patients will have a tilting of the epiglottis that obscures the anterior commissure. Some patients will have arytenoid hooding that prevents the posterior vocal fold from being visualized. In patients with mass lesions of the vocal folds, the vibratory pattern may not be delineated because of mass effect or ventricular hyperfunction. For patients with severe dysphonia who cannot sustain phonation for greater than 2 seconds at a steady fundamental frequency, it is best not to attempt objective analysis using videostroboscopy methodology.

Although objective evaluation is difficult because of the montage nature of video recordings, some authors have tried to use the strobe image for quantification.[29] The mucosal wave propagation across the superior surface of the vocal fold can be identified for tracking and some information regarding mucosal pliability can be estimated. In a comparison of 162 patients examined by videostroboscopy and high-speed glottography, perceptive evaluations of vibratory movements of vocal folds revealed a higher variability than assumed. These findings