Chapter 4
Kymographic analysis of high speed videoendoscopy in normal adult population

Wenli Chen & Thomas Murry

Abstract

Dynamic vibratory patterns of normal adults during steady phonation obtained with high-speed digital kymography (DKG) are presented. Steady phonation consists of quasi-periodic harmonics configured in a tilted shape with robust peaks and no spectral smearing. A range of asymmetry between the left and right vocal folds (VF) vibration is also seen in normal voices.

Keywords: kymography, DKG, high-speed imaging, vocal onset gestures, steady phonation

Introduction

The early work of von Leden and Moore ushered in a new era of studying laryngeal physiology [1]. They demonstrated that high-speed visualization of the VF, also known as high-speed videoendoscopy (HSV), provides a comprehensive understanding of the complete motion of the VF and is essential to the management of hoarseness. Current ultra-fast capturing rate using digital technology allows visualization of the entire vibratory cycle to obtain detailed information regarding the vibratory motions of the VF, but it generates a massive database and a need for efficient analysis [2-3]. Additionally, the relative scarcity of normative data relating to the onset of phonation and steady state phonation further challenges the use of HSV, specifically in the clinical setting. Until we understand the characteristics of normal phonation, comparisons to pathological VF vibration remains limited.

Kymography

Kymography is an approach to visualize the entire cycle movement of comprehensive movements of the VF (see Figure 1). Much like HSV, kymography quantifies cycle-to-cycle movement of each VF from a single or multiple horizontal lines of a laryngeal image across time [4]. Three different kymographic techniques are currently available. These are: 1) Videokymography (VKG), which uses a VKG dedicated video camera to capture high-speed (up to 8000 f/s) or standard (50 f/s) images in real-time directly during examination; 2) Digital Kymography (DKG), which uses a software application to extract kymograms from previously obtained HSV images; and 3) Strobosvideokymography (SVKG), which, like DKG, also uses external software to extract kymograms from previously obtained videoendoscopic recordings. A comprehensive overview of kymographic imagining can be found in Švec and Schutte [5].
Figure 1. Digital kymogram of an adult during modal phonation task. Image rotated 90° to the left.

VKG or DKG is ideally suited to observe left-right asymmetry, amplitude of vibration, open quotient, propagation of mucosal waves, and movement of the upper and lower margins of the VF [5]. While recent studies of kymography have focused on visual perceptual rating [6-8], direct quantification of VF movement is important to provide objective cycle-to-cycle movements of the VF. Both DKG and HSV present methods to obtain direct quantitative measures of the dynamic properties of VF vibration. Until we understand the high-speed characteristics of normal phonation, comparisons to pathological VF vibration remain limited.

Minor irregularities in normal phonation have been identified using HSV and kymography. Bonilha and Deliyski [6] used kymographic analysis and reported period irregularities were less apparent than glottal width irregularities in normal speakers. Phase asymmetry in normal individuals was from 0-10% [8]. VF movements during onset and offset of phonation have also been examined in normal speakers. Shaw and Deliyski [9] used HSV and reported the presence of atypical magnitude and asymmetry of mucosal wave during different phonation tasks in subjects with normal voices. Kunduk et al. [10] also used HSV and demonstrated negligible left-right phase differences in 14 female subjects with normal voices.

Recent efforts in laryngeal imaging have focused on characterizing the vibratory motion of the normal voice. Interestingly minor irregularities in the normal voice in voice onset/offset [10] and steady phonation [6-9] have been identified. In steady phonation, normal voices exhibited period and glottal width irregularities, phase asymmetry, and atypical magnitude and asymmetry of mucosal wave. In onset and offset of phonation, minor left-right phase differences were observed in normal female voices [10]. Taken together, these studies demonstrate the presence of irregularities of VF vibration in healthy voices. More importantly, these studies highlight the need to investigate normal VF vibration using direct visualization methods in order for the data to be interpreted appropriately.

In this chapter, we explore the dynamic vibratory pattern of normal adult phonation during voice initiation and steady phonation obtained from direct visualization methods using DKG. Portions of this chapter are based on the previous work of Chen, Woo and Murry [11]. The methodologies used to obtain DKG, quantitative analysis to obtain parameters of VF vibration, and results obtained from normal speakers are presented.

Methods

HSV requires a capturing system, recording system, and analysis software in order to be completely useful for research and clinical purposes. KayPentax High-Speed Digital Imaging (HSDI) system (KayPentax Photonmotion) offers a system, which consists of a 90° rigid endoscope (Model 9100) coupled with a 300 W Xenon light source. This HSDI system acquires gray-scale images at a rate of 2000 f/s with a spatial resolution of 256 x 120, rotated to a vertical position for capture. A microphone added to the system allows one to acquire and monitor both frequency and intensity of phonation.
Kymography image extraction

Kymograph analysis of the VF motions is critically dependent on accurate delineation of the edge of the VF from HSV [12]. Therefore, HSV samples are pre-processed using video-editing software. The levels of the image are adjusted to enhance brightness and contrast between the glottis and VF. Image rotation function is implemented in such a way that the image is vertically aligned. Video segments that demonstrate a full view of the VF with minimal movement of the subject are extracted from the HSV samples. Kay Image Processing Software (KIPS, Model 9181) is used to generate the kymogram by placing a transverse line across the glottis at the mid-membranous portion of the VF where VF contact is greatest (see Figure 2).

![Figure 2](image)

**Figure 2.** Extraction of digital kymogram of an adult using high-speed videoendoscopy. Kymographic data is extracted by placing a transverse line across the glottis to sample VF movements at the mid-membranous portion of the VF. Image rotated 90° to the left.

VF edge and frequency analysis

Once the kymographic image has been obtained, two types of analysis can be performed. Kymographic edge analysis (KEA) traces the movements of the left and right VF, resulting in a timeline of the coordinates of the left and right VF across time. This is particularly sensitive to pathological voice as it visualizes any robust peculiarities in the movements of the VF. KEA is also ideally suited to investigate voice onset gesture as it displays the initial pattern of VF adduction during pre-phonatory phase.

Spectral analysis of the vibratory waveform is ideally suited to characterize the vibratory behavior of the VF during steady phonation. Spectral analysis transforms the movement of the VF edge from a time domain into a frequency domain, also known as a spectrum. This allows an understanding of the frequency, power, periodicity, envelope, and symmetry of the steadily vibrating VF.

Once the kymographic image has been extracted, kymographic edge analysis is subsequently applied to identify and trace the VF edges (see Figure 3). If automatic tracing is imprecise, manual corrections (e.g., edge restrict, brightening, darkening, erase function) can be used to ensure accuracy in delineation of the VF edges. When the tracing approximates the VF edges, KEA is applied to the kymogram. The values are kymographic edge data, which present the coordinate values of the left and right edges of the VF movement across time (see Figure 4). Subsequently, spectral analysis can be obtained using the Fast Fourier Transform (FFT) function applied to the kymographic edge data. This FFT function is capable of transforming frequencies up to 1000 Hz. The resulting spectrum, therefore, displays frequencies from 0-1000 Hz. The vibratory power of the left and right edges of the VF is shown in Figure 5.
Figure 3. Kymogram of an adult during modal phonation task. The image is a kymogram created with the sampling line selected at the point shown in Figure 2 with edge detection. Image rotated 90° to the left.

Figure 4. Kymographic edge data of an adult during modal phonation task. Kymographic edge data visually displays the movements of VF across time. The upper tracing and the lower tracings represent right and left VF, respectively.

Figure 5. Digital kymographic spectrum of the left (black) and right (red) VF of an adult during modal phonation task. The first peak represents F0 while the subsequent peaks represent harmonic energies.
Kymographic analysis for voice onset in normal adult voices

Voice onset gesture is the pre-phonatory oscillation of the VF. Voice onset gesture is suggested to be important for normal VF vibration as less well-defined symmetry has been suggested to interfere with normal modes of vibration and may disrupt continuing vibratory motion in otherwise normal VF [13]. DKG provides a direct quantitative method to study voice onset gestures as it has the capability to delineate the exact coordinates of each VF motion across time.

Murry et al. [14] examined kymographic images derived from HSV in normal female subjects (see Figure 6). Samples were obtained from three healthy adult female subjects with no history of voice disorders. Subjects were asked to produce tokens at a comfortable pitch and loudness. Four time-measure values were obtained to examine voice initiation behavior of the VFs: 1) Mean time from onset of phonatory gesture to vibratory motion, 2) mean time from onset of vibratory motion to maximum vibration, 3) mean left/right difference in onset of vibratory motion, 4) mean time from onset of phonatory gesture to vibratory symmetry. They found female adult voice onset gestures characterized by a mild left-right asymmetry. The left-right voice onset time differences influence the time to vibratory stability. In addition, Murry et al. [14] found that a long pre-phonatory gesture does not necessarily disrupt left-right symmetry.

![Figure 6. DKG image (top). Image rotated 90° to the left. Digital kymographic edge data (bottom) of voice onset gestures of an adult during modal phonation task. The upper tracing and the lower tracings represent right and left VF, respectively.](image)

Voice onset characteristics have been examined by the seminal work of Wittenberg et al. [15] using visual perceptual measurements of digital kymography. They examined imitation of hard phonation and breathy phonation voice onset gestures. In hard phonation, they found prolonged pre-phonatory standstill with false VF obscuring the true VF at onset. Conversely, they found short pre-phonatory standstill, no glottal closure, and a lack of involvement of the false vocal folds (FVF) during breathy phonation. Tigges et
al. [16] also examined hard onset and normal onset phonation using kymography and reported the pre-phonatory standstill for the hard onset to be longer than in normal initiation.

Voice onset gestures have also been investigated using glottal area waveform (GAW). Kunduk et al. [17] compared voice onset gestures of two subjects varying in age (23-year-old and 78-year-old). In the older subject, voice onset gestures was characterized by a slow increase in glottal opening and an alternative decreasing and increasing pattern in the rate of opening/closing of the glottis before reaching the 90% of the maximum glottal area. They suggested that the pattern was an irregular rather than a steady and continuous increase in the glottal opening. The younger subject, however, demonstrated a sharp increase in GAW during voice initiation.

**Steady phonation in normal adults**

Vibration of the VF is critical to voice production. Disturbances in the regularity of vibration during steady phonation result in distinct voice disturbances. DKG is capable of capturing the cycle-to-cycle movement of the VF. Therefore, multiple parameters of VF vibration can be obtained. In the study by Chen et al. [11], kymographic images from HSV were examined for spectral configuration and symmetry in adult males and females. Samples were obtained from seven subjects (3 males, 4 females) with no history of voice disorders. Subjects were asked to produce tokens at modal, low, and high frequency at comfortable loudness and modal frequency with increased loudness. Several parameters of the digital kymography spectrum in normal adult phonation were observed.

**Spectral Configuration**

Chen et al. [11] found that normal adult VF vibration spectrum consisted of robust peaks with frequency of the harmonics positioned at approximate multiples of the fundamental frequency. Furthermore, the spectrum is characterized by a large and robust peak, followed by a medium-to-small second spectral peak, followed by a third small peak with no spectral smearing (see Figure 7). Although the overall tilted spectral shape of the spectrum in normal adult phonation is present with variation of normal voice, frequency and intensity of phonation may contribute to the specific power distribution of the harmonics. See Figure 8 for an example of changes in DKG spectrum with increases in intensity and frequency of phonation.

**Figure 7.** Spectral shape of normal adult VF vibration in modal, low, high, and modal loud phonation task. F0 (H1), 2nd harmonic (H2) & 3rd harmonic (H3) are tabulated across subjects to determine spectral shape of normal adult phonation. Healthy adult voices exhibit a tilted spectral shape with greater power in the H1 followed by attenuation of higher harmonics.
Figure 8. DKG spectrum for an adult in modal, low, high, and modal loud phonation task. DKG spectra for an adult male across the four phonatory tasks. Spectral configuration changes are seen with changes in frequency (Modal, Low, High) and intensity (Modal Loud).

The study by Chen et al. [11] is the first study to examine direct measures of VF vibration using spectral analysis from digital kymography. Indirect measures using spectral analysis have been extensively reported in acoustical studies [18-19]. Similar to DKG spectrum, the acoustic spectrum also consisted of a tilted shape with no spectral smearing [20]. The spectral slope per octave of normal voice consists of approximately 12 dB, 6 dB for pressed voice, and 18 dB for breathy voice [13]. Gauffin and Sundberg [21] reported that energy of the first harmonic is associated with degrees of VF excursion, whereas energies of higher harmonics can be associated with discontinuity occurring with VF impact. Therefore, differences in amplitude of H1 and H2 of the voice spectrum are useful for quantification of the degree of glottal adduction [22-23].

Asymmetry

DKG has been suggested to be ideally suited for examining asymmetry of VF vibration [24]. Previous studies using digital kymography adopted a visual perceptual rating to judge VF phase and glottal asymmetry. Shaw and Deliyski [9] examined healthy subjects using stroboscopy, high-speed videoendoscopy, and digital kymography. They found mild asymmetry of VF vibration in the majority of normophonic speakers during both habitual and pressed phonation. Furthermore, anterior-posterior asymmetry was noted slightly greater than left-right asymmetry [7]. Bonilha and Deliyski [6] examined glottal width and period irregularity in 52 normal speakers using videostroboscopy and DKG. They reported that period irregularities in VF vibration are less frequent in normophonic
speakers than glottal width irregularities. In the study of Chen et al. [11], asymmetry is examined quantitatively by comparing spectral peak differences between the left and the right VF spectrum. All subjects demonstrated a range of asymmetry (range: 3-26%, $\bar{x} = 13.42\%$, standard error = 6.8%) and the asymmetry was similar across variations in normal phonation.

**Future applications**

The value of HSV continues to gain importance in the clinical setting. With the development of automated analysis techniques such as those by Yan et al. [25], clinical evaluation and diagnostic decision-making will ultimately become possible on a routine basis. For some voice disorders, standard endoscopy may be suitable for diagnostic purposes, but for others (e.g., spasmodic dysphonia, VF paresis, muscle tension dysphonia), HSV may resolve differential diagnoses. Moreover, in cases such as muscle tension dysphonia, HSV before and after behavioral treatments may help to quantify the differences and provide additional qualitative data to support the need of voice therapy in such cases.

**Summary of DKG findings**

VF vibration critically determines the quality of voice production. Diagnosis and treatment of voice disorders are heavily dependent on the evaluation of the structure and function of the VF. The current gold standard in the clinical evaluation is laryngeal video-stroboscopy (LVS). However, due to low sampling rate and limited applications to irregular VF vibration [26], LVS is unable to resolve cycle-to-cycle VF vibration of all healthy and pathological voices. HSV and its ultra-fast capturing rates visualize the entire cycle-to-cycle movement of the VF in normal and pathological voices and across age and gender, adding to the understanding of VF physiology.

Previous approaches to objectively quantify cycle-to-cycle VF movement focused on examining the glottal area during VF vibration [25, 27-29]. Kymographic edge analysis and spectral analysis of kymography provides additional information on the vibratory behavior of each VF movement for the entire cycle. Recent studies using HSV and DKG have focused on a limited number of subjects.

By combining spectral analysis and edge analysis, we can quantify several parameters (e.g., voice initiation delay, symmetry, power, frequency, envelop). The knowledge of the presence and variability of endoscopic parameters in normal voices will enhance the significance of these parameters in vocal disorders. Only after a good understanding of what constitutes normal VF vibration has been established will we be able to accurately compare and characterize the abnormal voice.

In normal adult voices, steady phonation consists of quasi-periodic harmonics configured in a tilted shape with robust peaks and no spectral smearing. A range of asymmetry between the left and right VF vibration is also seen in normal voices.

Quantitative analysis of digital kymography in the clinical population is in its infancy. Švec, Sram, and Schutte [30] reported on several relevant features to look for in kymographic images in voice pathology. The ten categories of visual features of voice disorders include absence of VF vibration, interference of surround, cycle variability, closure duration, left-right asymmetry, lateral peaks, mucosal waves, opening vs. closing duration, medial peaks, and cycle aberration.

Future use of kymographic edge analysis will allow objective assessments of the differences in symmetry of vibration, mean time from onset of phonatory gesture to vibratory symmetry, and mean left/right differences in onset of vibratory motion. We may also expect to precisely document parameters of voice disorders (e.g., diplophonia or vocal paralysis) during steady phonation. Spectral analysis of kymography obtained from HSV
will provide assessment of asymmetry of amplitude, larger envelop of peaks, and aperiodicity. This information in turn may help to determine the most salient VF features that contribute to perceived voice quality.

References