Empirical Distribution of Glottal Edges (EDGE): A Statistical Assessment of Vocal Fold Kinematics Using High-Speed Videoendoscopy

Emiro J. Ibarra, Gabriel E. Galindo, Gabriel A. Alzamendi, Juan P. Cortes, Christian Castro, Rodrigo Manríquez, Alba Testart, and Matías Zañartu, *Senior Member, IEEE*

***Abstract*—Although laryngeal high-speed videoendoscopy (HSV) is crucial for studying vocal fold vibrations, its translation to clinical practice has been hindered by the large volume of data it produces and the difficulty in interpreting current analysis methods. Although image processing techniques have been developed to map spatial-temporal data into two-dimensional representations, they alter the geometrical construction of the glottis and do not provide standard quantitative features, thus challenging clinical interpretation. In response, we propose a new visualization and analysis framework for assessing the dynamics of vocal folds based on the empirical distribution of the glottal edge using HSV. This procedure analyzes vocal fold oscillations by preserving the shape of the glottis and quantifying the asymmetry between right and left vocal fold displacements along the anterior-posterior axis. This method was evaluated on four groups of participants: ten with normal voices, ten with vocal fold nodules, ten with muscle tension dysphonia, and two with unilateral vocal fold paralysis. The proposed method produces distinct representations for normal and pathological vocal fold vibratory behaviors and derived features based on amplitude and phase asymmetry metrics that show statistically significant differences between normal and pathological groups. Comparative analysis with state-of-the-art techniques indicates that our proposed method can complement the assessment of vocal fold vibration and enhance the clinical translation of HSV.**

***Index Terms*—Glottal segmentation, high-speed videolaryngoscopy, vocal folds, voice diseases.**

Received 30 October 2023; revised 14 May 2024 and 12 August 2024; accepted 7 September 2024. Date of publication 17 September 2024; date of current version 6 February 2025. This work was supported in part by the National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders under Grant P50 DC015446, in part by the Universidad Técnica Federico Santa María under Grant DPP PIIC N° 020/2021, in part by ANID under Grant BASAL FB0008, Grant FONDECYT and Grant 1230828, in part by Beca de Doctorado Nacional under Grant 21190074, in part by STIC AmSud under Grant ASPMLM-Voice 21-STIC-05. G. A. A, in part by the Universidad Nacional de Entre Ríos, and in part by MINCyT under Grant PICT-2021-I-INVI-00122 and Grant PICT-2020-SERIEA-01865. *(Corresponding author: Matías Zañartu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Comité de Bioética of the Universidad de Valparaíso (CEC-UV) under Application No. CB057, project Fondecyt 1151077, and performed in line with the the Declaration of Helsinki.

Emiro J. Ibarra, Juan P. Cortes, and Matías Zañartu are with the Department of Electronic Engineering and the Advanced Center of Electrical and Electronic Engineering (AC3E), Universidad Técnica Federico Santa María, Valparaíso 2390123, Chile (e-mail: emiro.ibarra@sansano.usm.cl; juan.cortess@usm.cl; matias.zanartu@usm.cl).

Gabriel E. Galindo and Rodrigo Manríquez are with the Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso 2390123, Chile (e-mail: galindof@gmail.com; rodrigo.manriquezp@usm.cl).

Gabriel A. Alzamendi is with the Institute for Research and Development on Bioengineering and Bioinformatics, Consejo Nacional de Investigaciones Científicas y Técnicas - Universidad Nacional de Entre Ríos, Oro Verde 3100, Argentina (e-mail: galzamendi@ingenieria.uner.edu.ar).

Christian Castro is with the Faculty of Rehabilitation Sciences, School of Speech Therapy, Exercise and Rehabilitation Sciences Institute, Universidad Andres Bello, Santiago 7591538, Chile, and also with the PhD Program in Health Sciences and Engineering, Universidad de Valparaíso, Valparaíso 2362905, Chile (e-mail: christian.castro@unab.cl).

Alba Testart is with the Department of Speech and Language Pathology, Universidad de Playa Ancha, Valparaíso 2360003, Chile (e-mail: albatest@yahoo.fr).

This article has supplementary material provided by the authors available at https://github.com/Emiroji/Supplementary-material-for-Empirical-Distribution-of-Glottal-Edges-EDGE-Paper.

Digital Object Identifier 10.1109/JBHI.2024.3462632

I. Introduction

ENDOSCOPIC laryngeal image procedures offer the possibility of observing relevant details of vocal fold (VF) vibration, making them appealing resources in clinical practice and research [1], [2]. Laryngeal videostroboscopy and high-speed videoendoscopy are the most widely used imaging modalities for assessing the anatomy and function of vocal pathophysiology. However, HSV provides a more reliable dynamic description of glottal function due to its high sampling rates, typically over 4000 frames per second. HSV captures high-quality monochrome or color images of laryngeal structures in motion with high temporal resolution, providing a detailed intra-cycle visualization of the VFs during vibration [3], [4].

The first efforts using laryngeal HSV were reported more than 60 years ago [5]. This initial study involved separate amplitude measurements for each VF at three glottal locations: immediately anterior to the vocal processes, the anteroposterior midpoint, and the midpoint of the anterior half of the VF. Amplitude variations over time at these three points were tracked. By using parameters like the open quotient and speed quotient, the research explored VF vibrations throughout the opening and closing phases. The methodology required manual analysis of

2168-2194 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Universidad Tecnica Federico Santa Maria. Downloaded on March 17,2025 at 16:07:38 UTC from IEEE Xplore. Restrictions apply.

each frame of the glottal cycle, a process that was labor-intensive and remains so today [6]. The large number of frames in HSV containing biomechanical information about VF behavior makes visual inspection a challenging and time-consuming task.

Numerous approaches have since emerged to simplify and enhance data visualization for clinical assessments based on HSV [6], [7], [8], [9]. Most techniques focus on depicting the temporal progression of a crucial aspect of VF dynamics: the VF edges extracted from HSV data containing the glottis. Most efforts effectively illustrate specific aspects of VF movement in a spatial-temporal visual format. Some approaches focus on evaluating glottal dynamics by analyzing a single row transverse to the glottal midline, assessing the position [10], [11], [12], [13], trajectories [14], velocities [15], and motion [16], [17] of the vocal folds. Others examine the overall glottal dynamics across the entire anterior-to-posterior length of the VF, analyzing the glottal area [1], [18], Fourier transform [19], [20], Empirical Orthogonal Function [21], VF trajectories [7], [8], and principal component analysis (PCA) [9].

Among the methods that analyze VF dynamics in specific HSV lines, Videokymography (VKG) [10], [11] and its HSV-based version, Digital Kymography (DKG) [12], [13], are widely used to study both normal and pathological conditions. These methods directly track the spatial positions of the VF from a selected horizontal line, enabling the extraction of dynamic VF features such as open, close, and speed quotients, fundamental frequency, as well as amplitude and phase asymmetries between the left and right VFs [11], [13]. However, this approach is limited by focusing the analysis on a single line of the glottis without considering the entire glottal length.

Another most common representation is the glottal area waveform (GAW), which displays the frame-by-frame segmented glottal area as a function of time. This representation also facilitates the description of the fundamental frequency and the identification of the opening and closing phases of the glottis [1], [18]. Furthermore, recent studies [22], [23] have utilized this time series to extract several features related to glottal dynamics, mechanical measures, and symmetry assessments. However, the principal limitation of the GAW is the lack of detailed representation of the anatomical geometry and localization of the vocal folds.

Another approach has analyzed HSV data using the Fourier transform (FT) [19]. This method extracts light-intensity time sequences from consecutive video frames and transforms them to reveal spectral data, offering four distinctive visualization modes to highlight the amplitude and phase of FT of vocal fold dynamics. This method allows for a condensed and enriched visual representation of oscillatory frequencies across all frames, enhances the visibility of oscillatory patterns and detects covibrations in structures like the ventricular folds [19], and allowing the study of diplophonic intervals in VF vibration [20]. However, relying on light-intensity fluctuations can lead to nonlinear responses that exaggerate amplitudes at image edges and glare spots, while flat surfaces show minimal variation [19]. Moreover, the procedure suffers from low specificity, mainly due to mucosal wave artifacts [20].

Empirical Orthogonal Function (EOF) analysis [21] has been applied to HSV data to study the spatial-temporal patterns of VF edge displacements around the glottal midline. This method involves decomposing the multilinear time series into orthogonal uncorrelated components. This allows for a detailed examination of vibration patterns and the measurement of desynchronization between anterior and posterior modes. In an experiment involving a patient with left recurrent nerve paralysis, the authors showed that EOF analysis can suitable detect biphonation–characterized by two fundamental frequencies–arising from left-right asymmetry of the vocal folds. In another study case involving a patient with functional dysphonia, oscillation patterns with a prominent phase shift between the anterior and posterior portions of the vocal folds were observed. These studies discuss the utility of EOF for revealing complex vocal fold dynamics. However, the method loses anatomical information from HSV, and is less intuitive to interpret.

In contrast, the phonovibrogram (PVG) method provides a color-coded image that emphasizes the medial-lateral VF displacement over several cycles, i.e., the orthogonal distance from the glottal center-line to the contours of the left and right vocal folds along the antero-posterior axis of the glottis [7]. This method enables the comparison of left and right VF dynamics qualitatively and quantitatively through the analysis of PVG contour angles and symmetry. The PVG has been applied to investigate spatio-temporal differences in VF vibration between prepubertal children and adults [24], differentiate healthy and pathological (organic-based) male voices [25], discriminate early stages of malignant and precancerous vocal fold lesions [26], and analyze sustained and dynamic vocal fold vibrations [27]. One issue associated with PVG is its dependence on the glottal center-line axis, which is difficult to identify, especially in pathological cases. As an alternative, researchers proposed the Glottovibrogram (GVG) technique [8], that considers the distances between the vocal-fold contours themselves instead of the distance to the glottal center-line axis. Despite the GVG avoiding sensitivity with the middle glottal axis, the resultant visualization loses the vibratory differences between the left and right VFs.

Another significant HSV-derived visualization tool is the glottaltopogram (GTG), which condenses the complete vibratory cycle from an HSV recording into three images using principal component analysis (PCA) of the temporal brightness sequence in each pixel [9]. This method reveals the overall synchronization of VF vibratory patterns, providing advantages such as full spatial resolution across the entire laryngeal area and eliminating the need for glottal area segmentation. However, the limitations of this method include the loss of temporal information and the requirement of manual adjustments for brightness equalization.

In general, the techniques described above have been valuable for investigating both normal and pathological VF vibrations, particularly for identifying left/right asymmetries. However, some visualization modalities involve mathematical transformations [9], [19], [21], or alter the visualization of the vocal folds in HSV, as in PVG where the left VF is rotated posteriorly relative to the right. While these adjustments help standardize images for analysis, they also change the perceived spatial and temporal dynamics of the vocal folds. Altering the visualization of the glottis configuration by HSV-derived techniques could hinder a comprehensive functional assessment of phonation

Authorized licensed use limited to: Universidad Tecnica Federico Santa Maria. Downloaded on March 17,2025 at 16:07:38 UTC from IEEE Xplore. Restrictions apply.

pathophysiology. This poses challenges for clinicians in terms of interpretation [6].

Herein, we propose a novel HSV-derived visualization technique based on the Empirical Distribution of Glottal Edge (EDGE) to emphasize and visually describe distinct patterns of VF vibration while preserving the geometrical characteristics of the glottis as visualized in HSV. In this context, EDGE provides a single two-dimensional image condensing the entire HSV video, as well as a sequence of images (i.e., a video) portraying the most likely vibratory cycle. From these two visual representations, EDGE quantifies both amplitude and phase asymmetry along the anterior-posterior glottal axis, providing new insights into vocal fold kinematics. We describe visual and quantitative EDGE-based features that highlight distinct VF vibratory patterns and aid in differentiating between typical and pathological vocal fold behaviors, thereby opening new avenues for vocal fold research.

The resulting EDGE images show distinctive visual patterns between typical phonation and vocal pathologies such as vocal fold nodules (VFNs), muscle tension dysphonia (MTD), and unilateral vocal fold paralysis (UVFP). Moreover, the method generates features that show the potential to differentiate between typical and atypical vocal fold dynamics, complementing well-established techniques such as DKG, PVG, GVG, FT, and GTG. A key advantage of EDGE is its preservation of the glottal edge form, as observed in the HSV, which we argue could facilitate clinical interpretation. Additionally, it provides statistics of amplitude and phase asymmetries along the anterior-posterior glottal axis. However, EDGE lacks an explicit time-domain representation, in contrast to DKG, PVG, or GVG, and does not assess oscillatory frequencies, unlike the FT method [19]. EDGE visualizations also depend significantly on the segmentation of the glottal edge, a limitation not shared by GTG or FT.

II. Material and Methods

A. HSV Recordings of Sustained Phonation

In this study, HSV recordings included ten subjects with normal voices (9 females and 1 male), ten patients with MTD (7 females and 3 males), ten patients with VFNs (9 females and 1 male), and two female patients with UVFP. During the procedure, participants produced a sustained vowel /i/ at a comfortable pitch and loudness. HSV data acquisition was performed following the Institutional Review Board-approved protocol CB057-15, with informed consent from all subjects. Before video recording, a laryngologist and a speech-language pathologist provided a clinical diagnosis for all participants based on a complete clinical evaluation and a clinician-administered Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) assessment. Table I lists the age, occupation, and reported CAPE-V scores for all participants.

The HSV recording facilities include a high-performance, high-speed color camera system (SA-X2, Photron, Tokyo, Japan) connected to a rigid endoscope (9106, KayPentax, Montvale, NJ) using a 35-mm C-mount adapter and a 300 W constant xenon light source (7152B, KayPentax, Montvale, NJ). The sampling frame rate was set to 8 and 10 kilo-frames-per-second (kFPs), and the sample length was less than 0.6 seconds.

TABLE I
Demographics and CAPE-V Scores of Participants by Group

Group	Code	Age	Job	CAPE-V
Normal	FN01	23	Student	-
	FN02	21	Student	-
	FN09	23	Student	-
	FN10	23	Student	-
	FN11	24	Teacher	-
	FN12	32	Teacher	-
	FN13	28	Actress	-
	FN14	30	Teacher	-
	FN15	21	Student	-
	MN16	25	Student	-
MTD	FP03	23	Student	24
	FP04	23	Student	31
	FP17	28	Teacher	27
	FP18	32	Actress	32
	FP19	23	Student	25
	FP20	35	Teacher	27
	FP21	27	Student	27
	MP22	23	Student	21
	MP23	25	Student	25
	MP24	27	Teacher	27
VFNs	FP05	21	Student	35
	FP06	24	Student	42
	FP25	23	Student	35
	FP26	21	Student	41
	FP27	27	Singer	35
	FP28	23	Teacher	32
	FP29	30	Lawyer	35
	FP30	30	Teacher	48
	FP31	28	Actress	38
	MP32	25	Student	41
UVFP	FP07	45	Teacher	25
	FP08	53	Vendor	65

Codes: N for Normal, P for Pathological, F for Female, and M for Male.

B. Empirical Distribution of Glottal Edge (EDGE)

The EDGE method is a visualization technique designed to succinctly represent the dynamic information captured in HSV recordings of VF vibrations. It achieves this by calculating the empirical distribution of glottal edges from HSV data, which it then utilizes to characterize the principal features of VF vibration. The method condenses the HSV data into a two-dimensional image, either representing the empirical distributions individually for the right and left VF edges or globally as a probability mass function of glottal edges. Furthermore, the EDGE method identifies the most probable vibratory cycle in the HSV data. The empirical image modalities allow for the quantification of amplitude and phase asymmetry along the anterior-posterior glottal axis. This section outlines the main steps involved in the EDGE method.

1) Pre-Processing of HSV Video: The pre-processing stage involves rotation and displacement compensations to eliminate undesirable effects caused by endoscope movement and accommodation of surrounding laryngeal tissue. Furthermore, it includes the delimitation of the region of interest (ROI) for frame-by-frame enhancement of the glottis area visualization.

Successive glottal cycles in the HSV data are roughly estimated from the fluctuation of total pixel intensity, with the valleys indicating glottal opening [17]. For each glottal cycle, the frame with the maximum opening is segmented using the semi-automatic software, Glottal Image Explorer (GIE) [28], an

Authorized licensed use limited to: Universidad Tecnica Federico Santa Maria. Downloaded on March 17,2025 at 16:07:38 UTC from IEEE Xplore. Restrictions apply.

open-source tool. The segmentation technique employed by this software is seeded region growing. Thus, the glottal area and VF edge contour for every maximum glottal opening in the HSV are obtained.

The anterior-posterior glottal axis is estimated as the first principal component obtained via PCA from contour points. The average tilt angle concerning the vertical image axis is applied to rotate all frames from one cycle in the opposite direction. Additionally, the centroid, ventral point, and dorsal point in the glottis contours are tracked along the successive maximum glottal openings, and the relative displacement between cycles is used to translate the frames in the opposite direction [14]. As a result, HSV data is reoriented and stabilized so that the sagittal glottal axis is vertically aligned with the anterior VF commissure pointing downward and the left/right VF edges on each side of the glottal midline axis.

Finally, the ROI in the enhanced HSV data is manually selected, ensuring that glottis is completely visible and the sagittal glottal axis is aligned centrally. The resultant frames are resampled to a size of 256×256 pixels.

The pre-processing results in a uniform VF vibratory behavior display across the successive HSV frames, regardless of the anatomical differences (dimension and shape) in the laryngeal structures and the endoscope orientation (distance and tilt angle) during the recordings. The compensated and resized glottal ROI is gathered in a three-dimensional array $I(x, y, k)$, for Cartesian indices $x, y \in \{1, 2, \ldots, 256\}$, and times $t_k = kT_s$, $k \in \{1, 2, \ldots, N\}$, where $f_s = 1/T_s$ is the sampling frame rate and N is the number of frames.

2) Glottal Segmentation and Edge Reconstruction: The next step involves glottal area segmentation and the high-quality reconstruction of the left/right VF medial edges from the $I(x, y, k)$ array. For all frames with an appreciable glottal opening, the glottis is semi-automatically segmented using GIE once again, and the VF edge contours are outlined: C^{l} and C^{r} are binary images, i.e., they take values of 0 or 1, outlining the left and right VF edges, respectively.

In HSV, the vibrating VF regions are sometimes hidden by anatomical structures of the larynx, such as the arytenoids in the posterior VF area or the anterior commissure behind the epiglottis or the epiglottic tubercle. Thus, we manually selected the anterior-posterior length, L, so that the vibrating portions of the VF edges remain visible in the open phase throughout the entire video recording (see Fig. 1).

During glottal closure, the glottis significantly reduces in size or may even disappear entirely, presenting a challenge for tracking the VF edges. We apply an edge tracking method based on a Kalman filter to predict the left/right edges $C^{\mathrm{l,r}}(x, y, k)$ during the VF collision [29]. While interpolation may be sufficient to predict VF edges during the closed phase in typical VF dynamics, the Kalman filter offers a more robust estimation in pathological cases. The Kalman-based tracking approach is similar to the one proposed by [29], which employed a polynomial approximation to delineate both VF edges and used the values and rate of change of these parameters as process observations. In contrast, for our approach, we take the detected left/right vocal fold edges as single elements (pixels) per row, and each one is tracked using position and velocity as observations for the Kalman filter.

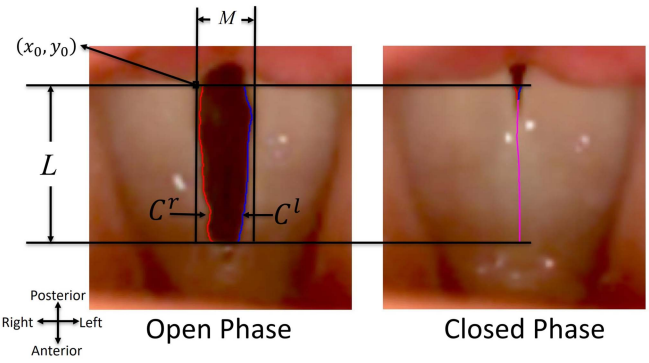

Fig. 1. Reconstruction of the left/right VF edges in the open phase (left image) and closed phase (right image). Edge points missed during the closed phase are estimated using a Kalman filter.

As a result, left/right VF edges are obtained for all frames in the HSV data, including the closed phase. Fig. 1 illustrates the results of edge detection during both the open phase (left image) and the closed phase (right image). In the open phase frame, the edges are computed using the segmentation algorithm, with the left glottal contour $C^{\mathrm{l}}(x, y, k)$ represented in blue and the right contour $C^{\mathrm{r}}(x, y, k)$ in red. The closed phase frame shows the estimated edges from the Kalman filter, with both edges overlapped in magenta. It is important to note that a third-order polynomial curve is fitted to the estimated edges in the closed phase to smooth the contours and obtain accurate estimates of both VF edges.

3) Probabilistic Modeling of VF Edges: The EDGE method aims to provide a visual representation (i.e., a picture) of the main temporal-spatial characteristics of VF vibration. It involves computing the empirical frequency distribution of glottal edges by processing an HSV recording frame-by-frame. In this method, wide-sense stationarity is assumed, requiring sustained vowel phonation tasks for the analysis, where the properties of VF vibration remain relatively constant over time. Thus, this approach is intended for the study of sustained phonation rather than transient gestures, but allows for merging of multiple repetitions of separate sustained phonatory gestures. It is important to note that, unlike stroboscopy, which requires periodicity in VF vibration, the EDGE method does not depend on this condition, thereby making it particularly suitable for studying pathological cases.

Let $\mathcal{D} = \{(x_i, y_j) : x_i = x_0 + i,\ y_j = y_0 + j,\ i = 1, 2, \ldots, M,\ j = 1, 2, \ldots, L\}$ be the rectangular image domain where VF vibrations are described, with (x_0, y_0) the top-left corner in $\mathcal{D}$ (see Fig. 1).

Even though the VF vibration is a complex biomechanical process, for simplicity, we assumed that both left/right VF edges are composed of independent, randomly distributed discrete contour points. We denote $p_{ij|\mathcal{S}} = \Pr[x = x_i, y = y_j|\mathcal{S}]$ the probability of seeing in $(x_i, y_j) \in \mathcal{D}$ an edge point from the $\mathcal{S}$-side VF, $\mathcal{S} \in \{\mathtt{l}, \mathtt{r}\}$, with $0 \leq p_{ij|\mathcal{S}} \leq 1$ and $\sum_{i=1}^{M} \sum_{j=1}^{L} p_{ij|\mathcal{S}} = 1$. The probability mass functions

Authorized licensed use limited to: Universidad Tecnica Federico Santa Maria. Downloaded on March 17,2025 at 16:07:38 UTC from IEEE Xplore. Restrictions apply.

$p_{ij|\mathcal{S}}$summarize the statistical information in $\mathcal{D}$ about the edge shapes, medial-lateral excursions, and left/right asymmetries; unfortunately, they are in general latent and unknown.

Here, we propose a histogram-like, non-parametric method to obtain empirical approximations, $F_{ij}^{\mathcal{S}}$, of the probability mass functions from HSV-derived data [30]. Let $\mathbb{I}^{\mathcal{S}}_{(x=x_u,y=y_v)}$ indicate whether an edge point in the $\mathcal{S}$-side VF locates in $(x_u, y_v) \in \mathcal{D}$, where $\mathbb{I}^{\mathcal{S}}_{(x=x_u,y=y_v)} = 1$ if $(x, y) = (x_u, y_v)$, and 0 otherwise. We hypothesize that the edge points are independent discrete random variables:

$$\mathbb{E}\left[\mathbb{I}^{\mathcal{S}}_{(x=x_u,y=y_v)}\right] = \sum_{i=1}^{M}\sum_{j=1}^{L} \mathbb{I}^{\mathcal{S}}_{(x_i=x_u,y_j=y_v)}\, p_{ij|\mathcal{S}} = p_{uv|\mathcal{S}}. \tag{1}$$

VF vibration along the successive cycles is portrayed in the edge contour $C^{\mathcal{S}}(x_i, y_j, k)$, $(x_i, y_j) \in \mathcal{D}$, $k \in \{1, 2, \ldots, N\}$. Since the $\mathcal{S}$-side VF contour was defined in Section II-B2 assuming one edge point per row for every frame, there is available a set of $L \times N$ edge point positions in the processed HSV data. Assuming that $C^{\mathcal{S}}(x, y, k)$ is a mean-ergodic random process, the mean value in the left-hand side of (1) can be approximated by the finite-sample estimator:

$$\bar{\mu}_{ij}^{\mathcal{S}} = \frac{1}{LN}\sum_{k=1}^{N} C^{\mathcal{S}}(x_i, y_j, k), \tag{2}$$

where $0 \leq \bar{\mu}_{ij}^{\mathcal{S}} \leq 1$ and $\sum_{i=1}^{L}\sum_{j=1}^{M} \bar{\mu}_{ij}^{\mathcal{S}} = 1$, by construction. By the weak law of large numbers, $\bar{\mu}_{ij}^{\mathcal{S}} \to p_{ij|\mathcal{S}}$ in probability $\forall (x_i, y_j) \in \mathcal{D}$, as $N \to \infty$ [31]. Thereby, we propose to empirically approximate the probability mass functions of the left/right VF edges, $\forall (x_i, y_j) \in \mathcal{D}$, from the HSV data, as follows:

$$F_{ij}^{\mathcal{S}} = \bar{\mu}_{ij}^{\mathcal{S}} = \frac{1}{NL}\sum_{k=1}^{N} C^{\mathcal{S}}(x_i, y_j, k), \quad \mathcal{S} \in \{\mathtt{l}, \mathtt{r}\}. \tag{3}$$

The empirical frequency distribution $F_{ij}^{\mathcal{S}}$ provides a new visual representation of the VF configuration in a single image, which highlights the most likely spatial locations for the left/right glottal edges, and the vibration asymmetries along the glottis longitude. Fig. 2(a) shows an example of the distributions $F_{ij}^{\mathtt{r}}$ and $F_{ij}^{\mathtt{l}}$ obtained from HSV data for a female subject without voice disorders. The color map ranges from low (blue) to high (red) frequency distribution, with blank space indicating zero frequency. Note the similar distribution observed in the left and right estimates illustrating regular and symmetric VF vibrations.

A global empirical probability mass function concurrently describing both edges can additionally be estimated from HSV data. By the sum rule [30], the probability of having an edge point in $(x_i, y_j) \in \mathcal{D}$ is given by:

$$\begin{aligned} p_{ij} &= \Pr[x = x_i, y = y_j] \\ &= \sum_{\mathcal{S}\in\{\mathtt{l},\mathtt{r}\}} \Pr[x = x_i, y = y_j|\mathcal{S}]\Pr[\mathcal{S}]. \end{aligned} \tag{4}$$

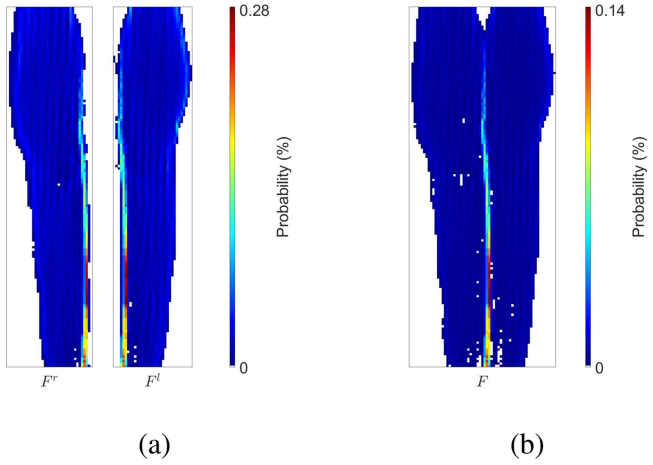

Fig. 2. Visualization based on the empirical frequency distribution obtained from HSV data for a female subject (FN010) with typical phonation. (a) Images displaying empirical probabilities $F_{ij}^{\mathtt{r}}$ and $F_{ij}^{\mathtt{l}}$ for the right and left VF edges, respectively. (b) Global probability mass function F_{ij}.

Following the development described above, and assuming $\Pr[\mathcal{S}] = \frac{1}{2}$ for $\mathcal{S} \in \{\mathtt{l}, \mathtt{r}\}$, we approximate the empirical probability mass functions as follows:

$$\begin{aligned} F_{ij} &= \frac{1}{2}\sum_{\mathcal{S}\in\{\mathtt{l},\mathtt{r}\}} F_{ij}^{\mathcal{S}} \\ &= \frac{1}{2NL}\sum_{\mathcal{S}\in\{\mathtt{l},\mathtt{r}\}}\sum_{k=1}^{N} C^{\mathtt{s}}(x_i, y_j, k), \end{aligned} \tag{5}$$

where $\sum_{i=1}^{L}\sum_{j=1}^{M} F_{ij} = 1$, by construction

The global empirical distribution (5) represents the likelihood of an element in $\mathcal{D}$ belonging to the glottal edge, regardless of whether it corresponds to the left or right VF. An example of F_{ij} is illustrated in Fig. 2(b), obtained from the same HSV data for a female subject with typical phonation as before. Although the individual distributions for the left/right edge displacements are not distinguishable, the overall shape of the vibrating glottal tissue is preserved in the representation. Note the appreciable residual posterior chink (blank region at the top center position), and the high probability in the medial region evidencing VF cycles with long closed phases.

4) Cycle-Based Conditional Analysis: Features regarding the glottis area or glottal phase (e.g., minimum/maximum area, open/closure instants, opening/closing/closed phases) provide relevant clinical insights about the regularity of VF vibrations [13], [18]. We extend the method introduced in Section II-B3 to address the estimation of phase-conditioned empirical frequency distributions of the glottal edges.

In regular VF oscillatory patterns, the glottal shape repeats in successive cycles. Let $\mathcal{T}$ be the time instant within the glottal cycle at which a glottal event of interest appears. For example, $\mathcal{T}$ could indicate when the glottis achieves maximum glottal area, glottal closure, or any intermediate configuration. We denote $p_{ij}|\mathcal{S}, \mathcal{T} = \Pr[x = x_i, y = y_j|\mathcal{S}, \mathcal{T}]$ the probability of

Authorized licensed use limited to: Universidad Tecnica Federico Santa Maria. Downloaded on March 17,2025 at 16:07:38 UTC from IEEE Xplore. Restrictions apply.

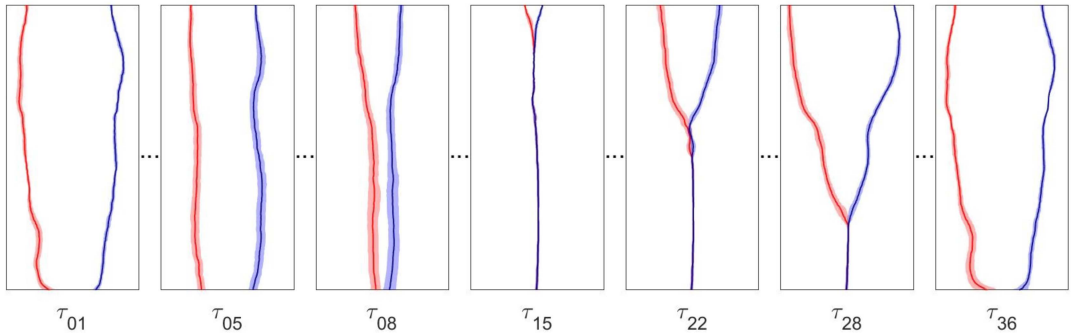

Fig. 3. Reconstructed cycle of vocal fold vibration from expected value (solid line) and variance (shaded regions), computed from the time conditional empirical distribution for the subject FN02 with no clinical signs of voice disorders.

having an edge point in $(x_i, y_j) \in \mathcal{D}$ conditioned on integrating the $\mathcal{S}$-side VF in the instant $\mathcal{T}$, with $0 \leq p_{ij|\mathcal{S},\mathcal{T}} \leq 1$ and $\sum_{i=1}^{M}\sum_{j=1}^{L} p_{ij|\mathcal{S},\mathcal{T}} = 1$.

The time-conditioned probability distributions $p_{ij|\mathcal{S},\mathcal{T}}$, $\mathcal{S} \in \{\mathtt{l}, \mathtt{r}\}$, provide statistical information about VF edges specific for the instant $\mathcal{T}$ in the glottal cycle. As these probabilities are unknown, we therefore introduce an approach for computing empirical approximations $F_{ij}^{\mathcal{S},\mathcal{T}}$, built on the HSV data.

Let $\mathcal{K} \subset \{1, 2, \ldots, N\}$ denotes the HSV frames (without replacement) assigned to instant $\mathcal{T}$ in the successive glottal cycles. Inspired by (2), we introduce the finite-sample time-conditioned estimator [30]:

$$\bar{\mu}_{ij}^{\mathcal{S}\mathcal{T}} = \frac{1}{L\,|\mathcal{K}|}\sum_{k\in\mathcal{K}} C^{\mathcal{S}}(x_i, y_j, k), \tag{6}$$

where $|\mathcal{K}|$ is the cardinality of set $\mathcal{K}$. By construction, $0 \leq \bar{\mu}_{ij|\mathcal{T}}^{\mathcal{S}} \leq 1$ and $\sum_{i=1}^{L}\sum_{j=1}^{M} \bar{\mu}_{ij|\mathcal{T}}^{\mathcal{S}} = 1$. As in previous section, the *theoretical* time-conditioned probability functions, $\forall \mathcal{S} \in \{\mathtt{l}, \mathtt{r}\}$, are approximated empirically from the HSV data [30]:

$$p_{ij|\mathcal{S},\mathcal{T}} \approx \bar{\mu}_{ij}^{\mathcal{S}\mathcal{T}} = \frac{1}{L\,|\mathcal{K}|}\sum_{k\in\mathcal{K}} C^{\mathcal{S}}(x_i, y_j, k), \quad (x_i, y_j) \in \mathcal{D}, \tag{7}$$

on the assumption that $\bar{\mu}_{ij}^{\mathcal{S}\mathcal{T}} \to p_{ij|\mathcal{S},\mathcal{T}}$ in probability as $|\mathcal{K}| \to \infty$, by the weak law of large numbers [31]. Computation of the approximated conditional probability functions in (7) thus combines the HSV data and cycle information obtained in the pre-processing stage (as described in Section II-B1).

Based on approximation (7) and Bayes' rule, the probability distributions jointly conditioned on vertical position y_j and instant $\mathcal{T}$ can be approximated by:

$$F_{ij}^{\mathcal{S},\mathcal{T}} = p_{i|j,\mathcal{S},\mathcal{T}} = \frac{\bar{\mu}_{ij}^{\mathcal{S}\mathcal{T}}}{\sum_{m=1}^{M}\bar{\mu}_{mj}^{\mathcal{S}\mathcal{T}}}, \quad (x_i, y_j) \in \mathcal{D}, \tag{8}$$

where we assume that the approximation fulfills in the limit $|\mathcal{K}| \to \infty$. For every y_j, the empirical time-conditioned distribution (8) provides a description of the medial-lateral position for an element of the $\mathcal{S}$-side VF in an instant $\mathcal{T}$ within the glottal cycle. Conditional distributions for both the left and right VFs can be gathered in a sequence of images, one for every phase instant $\mathcal{T}$, that portrays VF edge displacements across the glottal cycle.

Fig. 3 shows a sequence of empirical time-conditioned distributions in (8) across an average glottal cycle, ranging from the maximum opening ($\mathcal{T}_{01}$) to closure ($\mathcal{T}_{15}$) to subsequent maximum opening ($\mathcal{T}_{36}$) glottal instants. Instants $\mathcal{T}_{01} - \mathcal{T}_{05} - \mathcal{T}_{08} - \mathcal{T}_{15}$ span the glottal closing phase, whereas $\mathcal{T}_{15} - \mathcal{T}_{22} - \mathcal{T}_{28} - \mathcal{T}_{36}$ correspond to the opening phase. The visualizations are from the same HSV data considered in the previous examples. Solid lines depict the most likely (mean) positions for left/right VF edges. The sequence of images shows details of the glottal dynamics, for example, the gradual zipper-like VF opening/closing [32], the membranous glottis shape at the maximum opening, and the residual posterior chink in the closed phase. Furthermore, the shaded areas describe the cycle-to-cycle dispersion across VF oscillations, where narrow shaded areas are evidence of a regular vibration, a sign of typical (no disordered) vocal function.

5) Assessment of Vibratory Asymmetry: Measures based on the empirical distributions $F_{ij}^{\mathcal{S}}$ for the left/right VF edges are utilized to assess the amplitude and phase asymmetries in VF vibration. We define the 'effective displacement range' for each row y_j, as the interval between the 5th and 95th percentiles of these distributions, which encapsulates the central 90% of detected VF displacements. This range, denoted as $A_j^{\mathcal{S}}$ is assumed to be the effective range of displacement for left/right edge elements. In Fig. 4, the dotted lines indicate these percentile limits, while the double arrow illustrates the computed amplitude displacement for the left/right edges within this effective displacement range. Following the method described in [11] and [13], the amplitude asymmetry AA_j is computed as follows:

$$AA_j = \frac{A_j^{\mathtt{l}} - A_j^{\mathtt{r}}}{A_j^{\mathtt{l}} + A_j^{\mathtt{r}}}, \tag{9}$$

for every row y_j along the anterior-posterior glottal axis. The range of AA_j is $[-1, 1]$. Thus, an absolute value close to zero (one) suggests symmetric (strongly asymmetric) VF behavior, whereas a positive (negative) result indicates an edge displacement biased to the left (right).

Phase asymmetry quantifies the phase delay between the left and right VF edges. The visualizations obtained from the

Authorized licensed use limited to: Universidad Tecnica Federico Santa Maria. Downloaded on March 17,2025 at 16:07:38 UTC from IEEE Xplore. Restrictions apply.

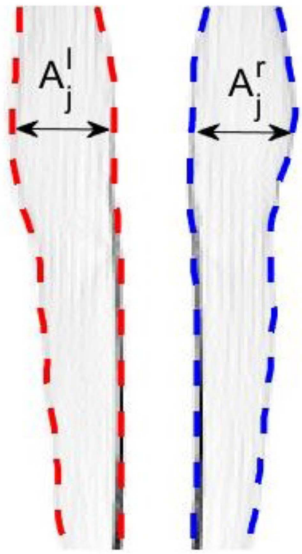

Fig. 4. Effective range of displacements for the left and right VF edges. Dotted lines represent the percentile range from the 5th to the 95th of the empirical distributions, from which horizontal amplitude displacements $A_j^{l,r}$ at row y_j for left (blue) and right (red) vocal folds are obtained.

time-conditional analysis introduced above (see Fig. 3) allow for the computation of phase asymmetry for every row y_j as follows:

$$PA_j = \frac{t_j^{\mathrm{l}} - t_j^{\mathrm{r}}}{T_0}, \tag{10}$$

where t_j^{l} and t_j^{r} are the time indices at the maximum lateral displacement of the left/right VF edges, respectively, and T_0 is the fundamental period [11]. Similar to (9), a value close to zero indicates in-phase, symmetric vibratory behavior. Additionally, a positive (negative) number implies that the left VF edge shows an advance (delay) in phase relative to the right one.

III. Results

In this section, the proposed probabilistic modeling of VF edges is applied and shown in detail to analyze the phonatory cycle in HSV recordings from the thirty-two participants. For comparison, EDGE visualizations are later contrasted with DKG, PVG, GVG, FT, and GTG techniques, as these methods are the current state-of-the-art visualizations of VF vibrations. Additionally, the amplitude and phase asymmetries estimated with the EDGE method were compared with those obtained using the DKG technique.

A. Empirical Distribution of Glottal Edges in Clinical Cases

Fig. 5 displays the HSV-based visualizations provided by the global empirical probability mass function, (5), for eight cases of shoes in Table I that include healthy and pathological female sustained phonation. These two-dimensional representations illustrate the empirical distribution for the left and right glottal edges, where the color map represents the probability distribution.

The visualizations for typical phonation without voice disorders –cases FN01 and FN02– exhibit similar probability distributions. Two features are worth noting. On the one hand, the broad and mirrored symmetric probability regions for left/right edge displacements during the open phase are evident. On the other hand, effective glottal closure is characterized by edge superposition with high probability along the glottal midline due to VF collision.

The HSV data for cases with MTD –panels FP03 and FP04– produced visualizations featuring reduced motility in the posterior glottal portion, manifested as narrow left/right medial regions with high probability and a significant gap between the left/right edges. These visualizations allow for the distinction of the most likely size of the membranous posterior glottal opening throughout the entire phonatory cycle. Besides, the differences observed in the left/right probability regions provide evidence of asymmetric VF vibration.

The results for the HSV data regarding phonotraumatic phonation due to VFNs -cases FP05 and FP06– illustrate a recurrent hourglass glottal shape and an incomplete glottal closure. The high-probability region in the center results from the restricted edge movements due to the VF nodules. Additionally, an asymmetric dispersion in the left/right edge displacements is evident, especially in the central and anterior portions.

The empirical distributions for cases with UVFP –panels FP07 and FP08– exhibit the most extreme disparities in left/right edge displacements. Case FP07 demonstrates significant asymmetry due to a motile right VF, which displays a wide range of displacement, contrasted by a non-vibrating left VF. This asymmetry is evident from the discernible high-probability region biased toward the left edge. Conversely, in case FP08, the displacement in the anterior and central VF portions is notably more symmetric, with residual irregular vibration observed in the posterior portion. This case pertains to a female patient with VF paralysis who recovered VF mobility following clinical treatment.

In general, the EDGE for healthy individuals tends to be mirror-symmetric, with a high probability region along the glottal midline, describing the regular vibratory behavior of the VF and the nature of glottal closure. In pathological cases, this characteristic structure is lost, especially when the pathology produces asymmetrical vibration or reduces VF mobility, resulting in other visual patterns such as areas with cyan and green colors. To quantify these differences in the color patterns of the derived EDGE images, the mean and standard deviation (STD) of intensity pixels were computed in the indexed representation (gray image). The mean results for all subjects were 20.23, 25.50, 43.99, and 38.95 for Normal, MTD, VFNs, and UVFP, respectively (individual results are shown in the supplementary material Table II in the global EDGE column).

A one-way Analysis of Variance (ANOVA) was conducted to identify statistically significant differences between the groups for this feature. The analysis, excluding the UVFP group due to limited samples (only two cases), yielded an F-value of 23.50 with $p < 0.001$, indicating significant differences among

Authorized licensed use limited to: Universidad Tecnica Federico Santa Maria. Downloaded on March 17,2025 at 16:07:38 UTC from IEEE Xplore. Restrictions apply.

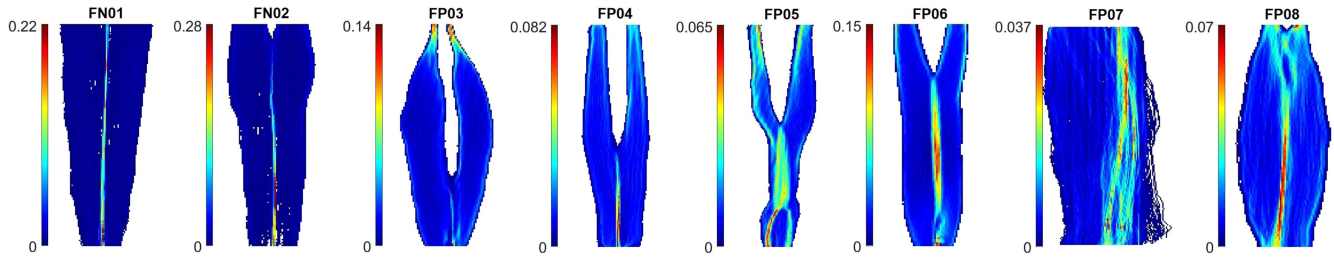

Fig. 5. Empirical distribution of glottal edges for subjects with normal voices (FN01 and FN02), MTD (FP03 and FP04), VFNs (FP05 and FP06), and UVFP (FP07 and FP08). The color bar represents the probabilities in percentage. EDGE revealed a homogeneous distribution in normal voices, a wide posterior gap in MTD, diminished displacement in the areas affected by nodules in VFNs, and restricted movement in VF affected by UVFP.

the groups. Subsequent post-hoc Tukey's Honestly Significant Difference (HSD) and Cohen's d analyses (results presented in Table III of supplementary material) suggest that the differences between the VFNs and Normal groups are statistically significant with large effect sizes, as well as between VFNs and MTD. However, insufficient statistical evidence exists to conclude that the means between MTD and Normal groups are significantly different. This lack of differentiation with MTD is attributed to the complexity and inter-subject variability of MTD cases. For example, patients with MTD do not always present with a posterior glottal gap or reduced vocal fold mobility, which are more consistently observed features in patients with VFNs.

B. Conditional Distribution of Glottal Edges in Clinical Cases

The empirical distribution of glottal edges conditioned on consecutive time events along the glottal cycle (see Section II-B4) provides a summarized visualization of the most likely vibratory VF behavior in a HSV recording. Fig. 6 shows the resulting visualization conditioned on five glottal events –maximum opening, intermediate closing phase, closed phase, intermediate opening phase, and adjoining maximum opening– for HSV recordings of cases FN01, FP04, FP06, and FP07 listed in Table I. Interactive versions of the visualizations shown in Fig. 6 are available in the videos of the supplementary files.

In the case of a female without voice disorders (see Fig. 6(a)), the visualizations illustrate the characteristic triangular dorsal-ventral shape during both the closing and opening phases. This pattern results from a persistent opening at the posterior cartilaginous portion of the glottis and the complete closure of the glottis in the closed phase. The sequence of images lucidly portrays the gradual and symmetric, zipper-like glottal closing/opening.

For those cases corresponding to impaired phonations (Fig. 6(b)–(d)), the visualization sequences derived from the empirical conditional distribution seem to provide further visual and clinical evidence about the pathophysiology of VF behavior. Notably, the increased left/right edge asymmetries, the incomplete closure at various points along the VF length during the closed phase, alterations in the kinematics of VF edges during opening and closing phases, and edge lateralization due to uneven VF collision are all apparent.

Furthermore, the shaded areas delineating cycle-to-cycle variability provide additional clinical evidence. This ranges from a narrow dispersion for typical periodic vibration (Fig. 6(a)), to discernible variability regions unevenly distributed across the VF length for hyperfunctional phonations (Fig. 6(b) and (c)), to extensively increased shaded areas for chaotic VF vibration caused by unilateral paralysis (Fig. 6(d)).

The results suggest that the time-conditional analysis proposed here contributes to clinical information regarding VF behavior, furnishing further evidence concerning healthy or pathological phonation that complements the description offered by the HSV-derived empirical frequency distribution of glottal edges. To the best of the author's knowledge, no other visualization method offers a succinct illustrative description of VF kinematics while preserving the same spatial and geometric depiction of the glottis area as observed in HSV, similar to the method proposed here.

C. Assessment of Vibratory Asymmetry in Clinical Cases

Fig. 7 illustrates the results of vibratory asymmetry for eight subjects listed in Table I. The remaining subjects are shown in the figures of the supplementary material. In each panel, the computed AA (depicted by the blue line) and PA (depicted by the orange line) measures along the anterior-posterior axis in the glottis are shown. Additionally, visualizations based on the effective displacement range for the empirical probability distributions for the left and right edges are plotted separately. Asymmetry values and anterior-posterior positions are reported as percentages on the horizontal and vertical axes, respectively.

For the cases featuring healthy periodic VF vibration (Fig. 7(a) and (b)), both AA and PA measures are close to zero (magnitudes less than 10%) throughout the VF length, aligning with the observed mirrored symmetry in the empirical distributions of the left/right edges. The MTD patients (Fig. 7(c) and (d)) show increased magnitudes of AA (maximum values in the range of 10–30%), mainly in the anterior and posterior extremes, along with less changes in the PA. For the cases of VFNs (Fig. 7(e) and (f)), the magnitudes of AA and PA measures increased, but the localization and significance manifest differently along the entire length.

Authorized licensed use limited to: Universidad Tecnica Federico Santa Maria. Downloaded on March 17,2025 at 16:07:38 UTC from IEEE Xplore. Restrictions apply.

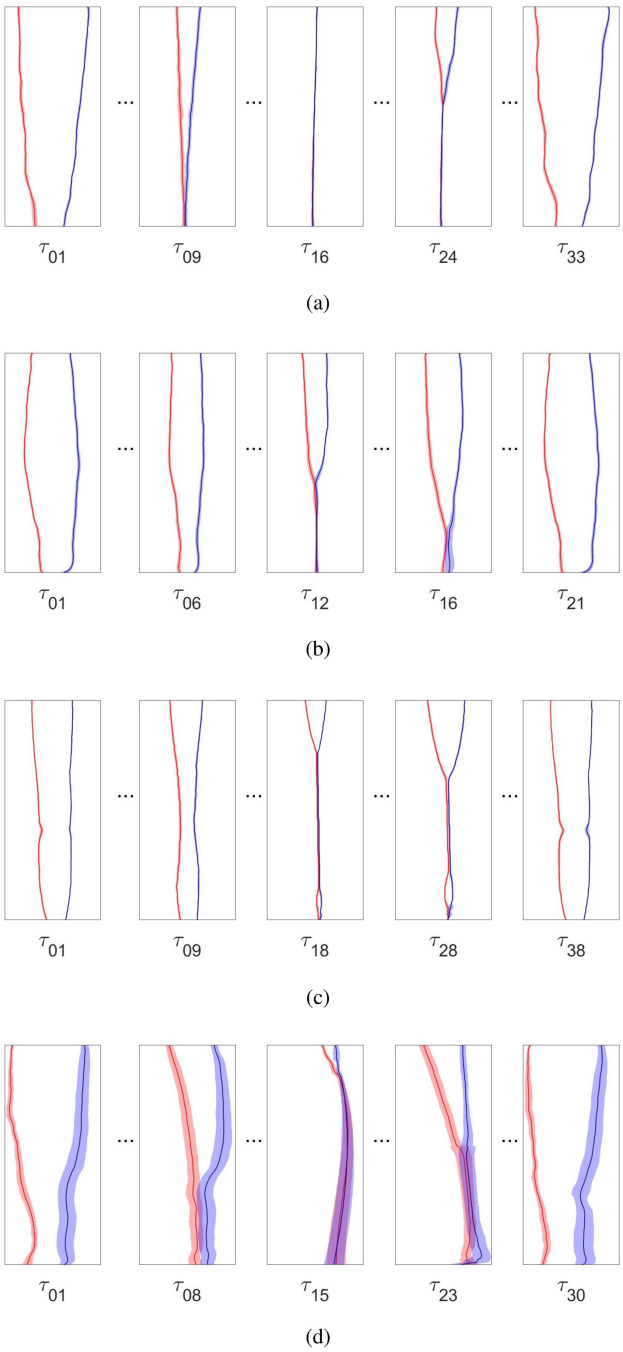

Fig. 6. Empirical distribution conditioned on five glottal events (from left to right: maximum opening, intermediate closing phase, closed phase, intermediate opening phase, and adjoining maximum opening). (a) FN01: Normal subject, exhibiting a zipper-like glottal closing/opening pattern; (b) FP04: A patient with MTD, showing incomplete glottal closing in the posterior area; (c) FP06: A patient with VFNs, where an hourglass-shaped pattern is evident; (d) FP07: A patient with UVFP, noted for the lack of movement in the left VF.

Recordings for patients with unilateral paralysis (Fig. 7(g) and (h)) exhibit significant increases in the magnitudes of both AA and PA measures across the entire VF length. For the FP07 case, negative AA values with magnitudes ranging from 20 to 50% indicate wider edge displacements in the right VF contrasting with the non-motile left VF, whereas PA shows varied edge behavior in the posterior and anterior VF regions (positive and negative PA values, respectively). The treated unilateral paralysis (case FP08) displays symmetric VF vibration, producing AA magnitudes close to zero over most of the VF length. However, the observable increase in PA magnitudes provides evidence of residual perturbations in VF vibration. This phenomenon of residual perturbations is visually evident in the reconstructed cycle vibration (see video of FP08in supplementary material).

These examples demonstrate how asymmetry measures can further describe the lack of regularity in disordered vibrations. As expected, AA measures effectively emphasize the VF positions where the ranges of edge displacement differ the most, as illustrated by the left/right probability distributions. Additionally, the color pattern of derived left and right EDGE can provide valuable information. In this context, the mean pixel intensity in the indexed version of the left/right EDGE was quantified. The mean values for all subjects in each group were 39.70/40.62, 50.18/45.49, 73.09/72.19, and 57.65/50.70 for Normal, MTD, VFNs, and UVFP, respectively (results for each subject are shown in the supplementary material, Table II in the columns Right EDGE and Left EDGE). Similar to the global EDGE mean, the significant difference from these features occurs between VFNs and the other groups (summarized in Table III of the supplementary material). It is important to highlight that the largest difference between the mean of left and right EDGE was in the MTD group, suggesting that the difference between this pathology and normal cases lies in the asymmetric behavior of VF vibration.

To compare asymmetry metrics between the studied groups, the statistics (absolute mean and STD) were obtained by averaging the asymmetry measures along the glottal length. The mean values for AA in different groups are 4.1%, 10.8%, 12.6%, and 22.5% for Normal, MTD, VFNs, and UVFP, respectively, while for PA they are 2.5%, 7.6%, 10.3%, and 26.9% (individual results are reported in Table II of the supplementary material, in the columns AA and PA). A one-way ANOVA excluding UVFP yielded F-values of 6.91 and 8.44 for AA and PA metrics, respectively, with $p < 0.01$ for both, indicating significant differences among the groups for these metrics. The post-hoc Tukey's HSD and Cohen's d analyses (presented in Table IV of the supplementary material) suggest that the differences in AA and PA between the MTD and Normal groups are statistically significant with large effect sizes, as well as between VFNs and Normal. However, for these features, there is insufficient statistical evidence to conclude that the means of AA or PA differ between MTD and VFNs.

On the other hand, the mean amplitude and phase asymmetry metrics obtained with EDGE are consistent with those reported in previous studies. For instance, [13] using medial DKG, found amplitude and phase asymmetries of 5% and 7%, respectively, for seven normal subjects and 11% and 20% for forty-six subjects with pathological voice disorders. Additionally, [11], utilizing medial VKG, observed amplitude and phase asymmetries of 5% and 2%, respectively, for three normal subjects, 33% and 7% for a patient with UVFP, and an average of 16.5% and 2.5% for two patients with polyps on two folds. Furthermore, another type of asymmetric metric based on EOFs, as reported

Authorized licensed use limited to: Universidad Tecnica Federico Santa Maria. Downloaded on March 17,2025 at 16:07:38 UTC from IEEE Xplore. Restrictions apply.

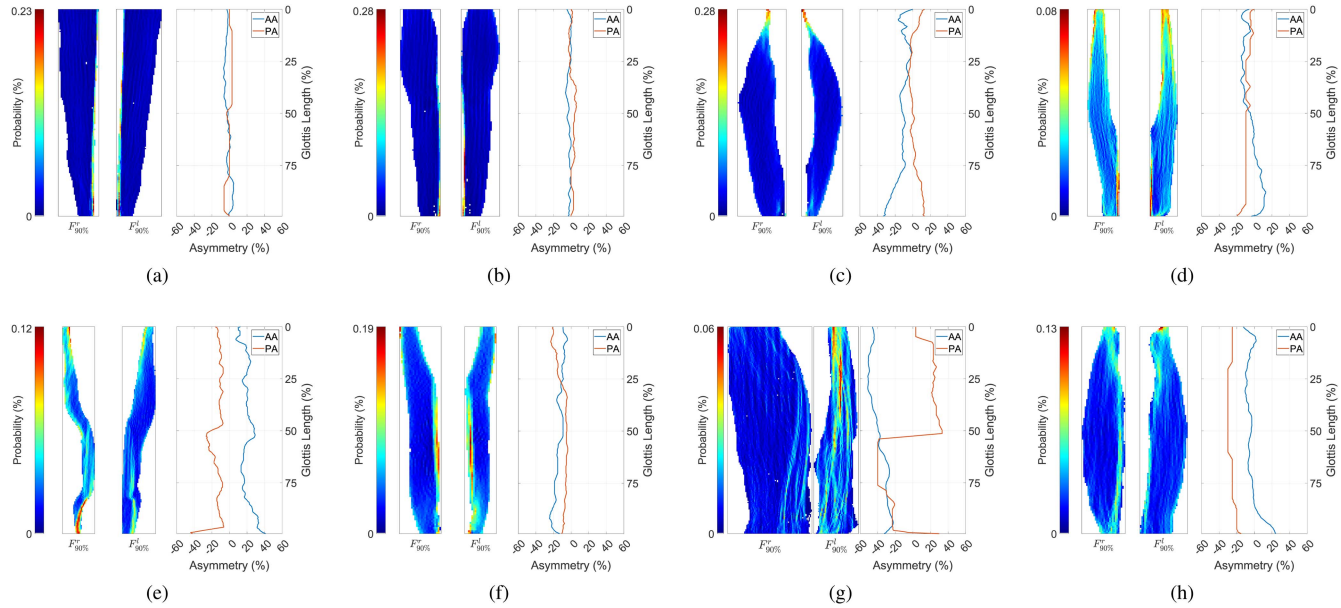

Fig. 7. [Results for Assessment of vibratory assymetry] Assessment of vibratory asymmetry in clinical cases. The panels depict a effective displacement range of the individual empirical distribution for the right ($F^{\mathrm{r}}(x,y)$) and left ($F^{\mathrm{l}}(x,y)$) edges, alongside the characterization of Amplitude Asymmetry (AA) and Phase Asymmetry (PA) along the anterior-posterior dimension. The cases represented are: (a) FN01, (b) FN02, (c) FP03, (d) FP04, (e) FP05, (f) FP06, (g) FP07, and (h) FP08. The color maps represent the empirical probability as a percentage.

in [21], revealed significant left-right asymmetries in patients with UVFP compared to those with normal and MTD conditions.

These metrics of VF vibration asymmetry, along with the features obtained from the EDGE images, enable differentiation between normal and pathological cases. Although classification is not the primary focus of this work due to the scarcity of data analyzed, we still pursued an exploratory experiment to study the potential of the derived EDGE features alone to differentiate among groups. This experiment involved selecting a vector with eight features that show statistical differences among the groups (Normal, MTD, and VFNs), including the mean and STD of global EDGE, the mean of left and right EDGE, and the mean and STD of AA and PA. We then used Linear Discriminant Analysis (LDA) for dimensionality reduction, and the resultant first and second canonical variables were clustered using K-means, as shown in Fig. 8. The analysis reveals that all normal subjects are grouped in the same cluster. For pathological cases, in the MTD group, two cases are mixed with normal cases. This occurred for two subjects with no evidence of VF asymmetry recorded in the HSV (see patients FP018 and MP24 in Table II of the supplementary material). In the VFNs group, two cases are misclassified into different clusters. Even with nothing but kinematic data, these pilot results indicate that 86.67% of simple EDGE-based features were clustered correctly, supporting the potential of this method to quantitatively study VF kinematics.

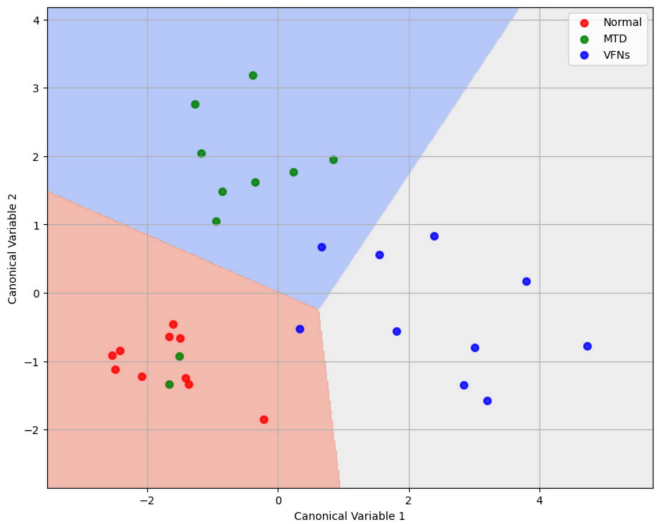

Fig. 8. First vs. second canonical variables of Linear Discriminant Analysis (LDA) from EDGE-based features for Normal: Red, MTD: Green, and VFNs: Blue. Colored areas represent K-means clustering.

D. Subjective Comparison With Other Methods

Fig. 9 displays the DKG, PVG, GVG, FT, and GTG for HSV recordings corresponding to cases FN02, FP04, FP06, and FP07. The DGK was applied to the three rows highlighted in white in the reference image [12]. For PVG, the color code denotes the distance from the vocal fold edge to the middle glottal line [7], while in the GVG it denotes the distance between the left/right VFs [8]. In FT, colors represent the amplitude of oscillation in dB [19]. For GTG, the images from left to right represent the 1st and 2nd projections on the principal components and the associated error term [9].

For case FN02 with healthy vocal folds, all methods display a regular pattern in VF vibration, and the symmetry between the left and right VFs can be identified in DKG, PVG, FT, and GTG.

Authorized licensed use limited to: Universidad Tecnica Federico Santa Maria. Downloaded on March 17,2025 at 16:07:38 UTC from IEEE Xplore. Restrictions apply.

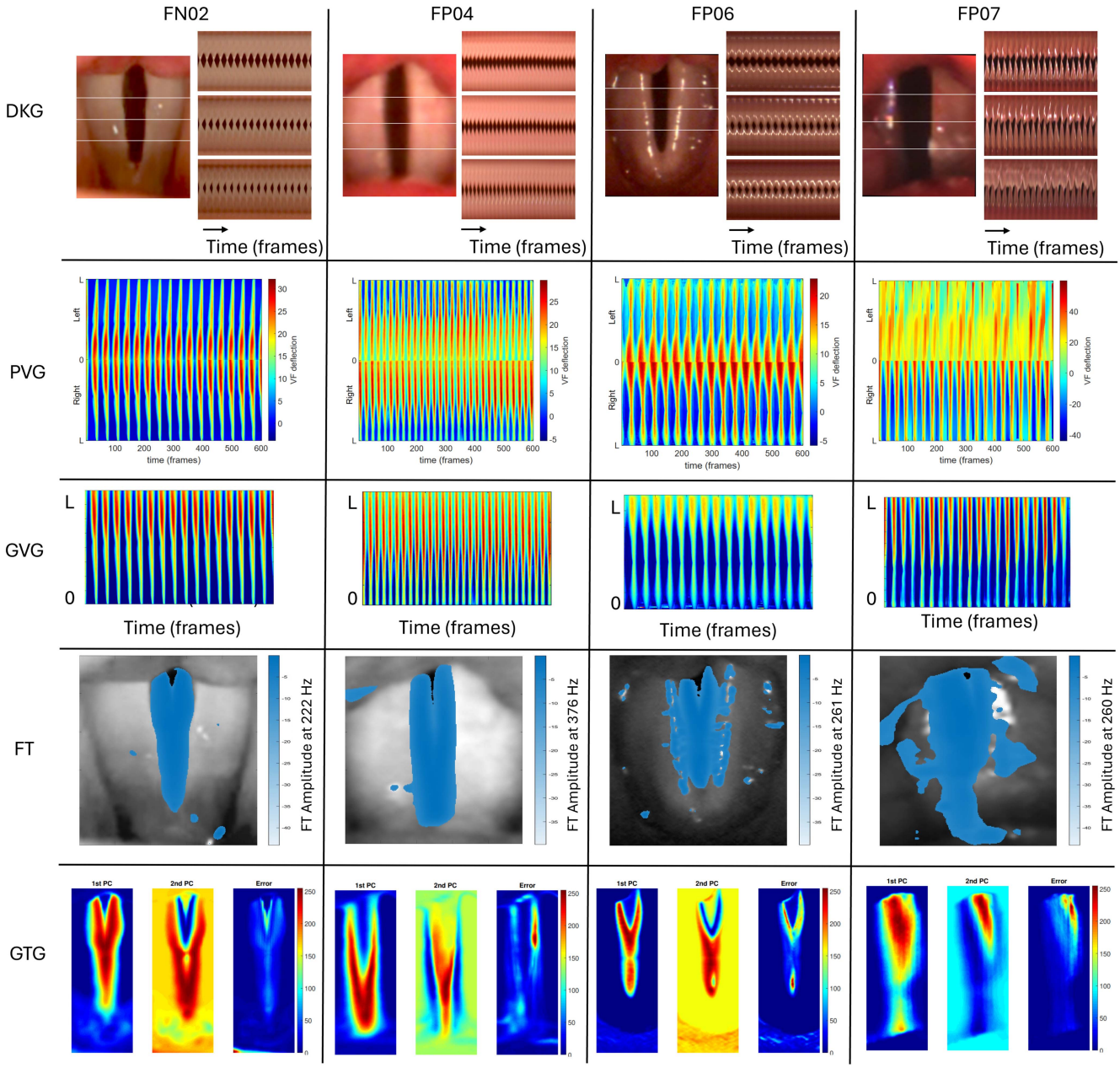

Fig. 9. Digital Kymography (DKG), Phonovibrogram (PVG), Glottovibrogram (GVG), Fourier transform (FT), glottaltopogram (GTG) for cases FN02, FP04, FP06 and FP07.

This symmetric pattern is also reflected by EDGE in (Figs. 3 and 5), as well as in the asymmetry assessment (Fig. 7(b)).

In the pathological cases, DKG, PVG, FT, and GTG illustrate asymmetry patterns in VF vibration. The amplitude asymmetries are visually evident in DKG and PVG; however, in FT and GTG, the type of asymmetry (amplitude or phase) is difficult to discern. In contrast, in EDGE, the amplitude asymmetries are visualized by comparing the empirical distribution of the left/right VFs and along with the quantified from anterior to posterior glottal length, as shown in Fig. 7.

The incomplete glottal closure observed in cases FP04 and FP06 is discernible in the posterior DKG, GTG, and FT. This characteristic is also observed in EDGE (Fig. 5), with the advantage that the posterior gap in EDGE coincides with the size observed in the HSV.

In bilateral nodules (case FP06), DKG and FT do not present a distinguishable pattern. In contrast, PVG and GVG nodules are distinguished by a concave shape [7], located in an area with reduced deflection (blue color) in the middle of the VF length. This is contrasted by a lighter region (green color) in the first PCA GTG coefficient. In our probabilistic method, the typically hourglass-shaped pattern and high probability region directly allow for the identification and localization of the injured VF area (see Figs. 5 and 6(c)).

Authorized licensed use limited to: Universidad Tecnica Federico Santa Maria. Downloaded on March 17,2025 at 16:07:38 UTC from IEEE Xplore. Restrictions apply.

For the paralysis case FP07, the wide deflection observed in the empirical distribution of the right edges compared to the left edges (see Fig. 7(g)) enables intuitive identification and localization of the pathology. In the phonovibrogram, the irregular pattern in the PVG of the affected vocal fold indicates paralysis [7], [33]. In DKG, this anomaly is also identified by the reduced mobility of the paralyzed VF. In other methods, the characterization of paralysis is less obvious. For example, in GTG, the dark blue area in the second PCA component may represent the difference in left/right vocal fold deflection, and the anterior-left VF section in the error image could correlate with the paralyzed VF. However, these color patterns in GTG can vary depending on the manual adjustment of brightness.

Note that DKG, PVG, and GVG preserve the temporal representation of the vibratory cycle. This allows for the detection of changes in the periodicity of the oscillatory pattern of the VF over time. For example, in FP07, these three methods exhibit variability in the vibratory register over time, such as changes in the color pattern in GVG and upper PVG, or in the patterns of the DKGs. This variability is not observable in EDGE, FT, and GTG due to their limitations in temporal representation.

E. Cycle-to-Cycle Estimates of Amplitude and Phase Asymmetries

In previous work, [12] proposed an automated measurement of vocal fold vibratory asymmetry from HSV. This method obtains the lateral displacement waveforms of the left and right vocal folds by tracking the edges defined by the boundary between the glottis and vocal fold tissue regions using a DKG. Consequently, the technique yields cycle-to-cycle left-right asymmetry measurements at multiple positions along the anterior-posterior length of the glottis. In this work, we adapted this methodology for comparison with our asymmetry measurements. Instead of obtaining lateral displacement waveforms from the edges defined by the DKG, we utilized our edge reconstruction method.

Fig. 10 displays the amplitude asymmetry and phase asymmetry as assessed by our proposed method alongside the mean values with 95% confidence bounds obtained from cycle-to-cycle estimation. The resulting graphics illustrate that the amplitude and phase asymmetries from our estimations closely align with the mean values derived from cycle-to-cycle measurements across the entire length of the glottis. Based on these results, we conclude that the AA and PA derived from our proposed method offer a reliable approximation for quantifying vocal fold vibratory asymmetry.

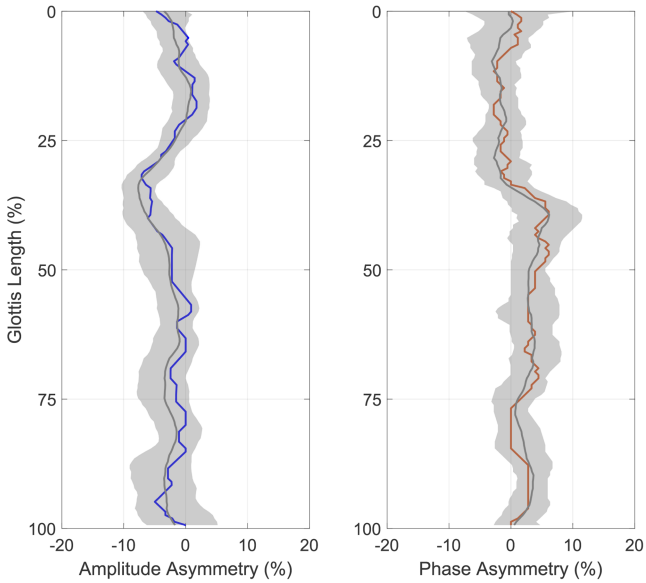

Fig. 10. Amplitude (blue solid line) and phase (orange solid line) asymmetry obtained from the probabilistic method compared to the mean (gray solid line) and 95% confidence intervals (shaded regions) from cycle-to-cycle asymmetry estimations.

IV. Discussion

This study introduces EDGE, a probabilistic analysis of vocal fold vibratory behavior, leveraging glottal edge detection from high-speed videoendoscopy. This method generates a single two-dimensional image that encapsulates the behavior of numerous vibratory cycles, a sequence of images that reconstructs the most likely oscillatory patterns, and measures of vibratory asymmetry of the vocal folds along the length of the anterior-posterior glottis. The results obtained from the cases studied suggest that the proposed analytical tools may support the clinical interpretation of both standard and aberrant vibratory patterns through EDGE-based features, including those derived from color patterns and amplitude and phase asymmetry behaviors.

EDGE shows different visual patterns with the potential to characterize subjects with healthy and pathological voices. For instance, individuals with healthy voices typically display VF vibrations with symmetrical horizontal displacements and minimal irregularities, producing a homogeneous distribution of probabilities. This results in an image where the left and right vocal folds exhibit similar color and shape. In contrast, our analysis revealed amplitude and phase asymmetric behavior in patients with vocal fold pathologies, where certain irregularities, imperceptible in the video, become discernible in the shape and color of the empirical distribution of glottal edges.

When patients with MTD present an incomplete closure featuring posterior and membranous gaps, this is revealed by the EDGE method as a posterior area without the probability of obtaining a glottal edge (posterior blank area in EDGE). However, this pattern is not always present in MTD patients, with amplitude and phase asymmetry being the most common irregularity features in their vibratory patterns. These findings are consistent with clinical notions outlined in [34] and provide preliminary evidence for the expected asymmetric behavior in these patients. Nevertheless, it is important to highlight that in some MTD patients, there are no evident anomalies in VF vibration, resulting in EDGE features similar to those obtained in normal subjects. For these cases, acoustic and aerodynamic-based features could be the key to understanding the pathology's effects.

Authorized licensed use limited to: Universidad Tecnica Federico Santa Maria. Downloaded on March 17,2025 at 16:07:38 UTC from IEEE Xplore. Restrictions apply.

For patients with vocal nodules, EDGE effectively differentiates two separate vibratory zones, with a region in between exhibiting no vibration and displaying heightened contact–attributes characteristic of the lesion region. Accurately identifying the site of the lesion is pivotal for approximating the nodule's size and is essential for considering treatment strategies. Furthermore, the absence of vibration within a specific region could offer valuable insights into the progression of the injury, as newly formed nodules may correspond to a smaller non-vibratory area compared to older nodules, which may have a larger non-vibratory area.

The EDGE analysis for patients with vocal fold paralysis reveals a more pronounced irregularity in vibration compared to other participants in this study. The afflicted VF tends to demonstrate increased stiffness and diminished mobility compared to the healthy VF, resulting in a distinguishable contrast in vibratory patterns. These discernible differences aid in accurately pinpointing the specific area of the paralyzed section of the VF. However, more UVFP cases need to be analyzed with EDGE to ensure that these features represent a common pattern.

A. Significance of EDGE

In contrast with the visualization methods reported in the state-of-the-art, EDGE is the only method that provides a compact visualization of vocal fold oscillatory behavior while maintaining the form of the glottal edge, without any mathematical transformation. This enables the mapping of left-right VF trajectories and allows for the visual interpretation of asymmetric behavior, the identification of reduced motility, and the localization of posterior and anterior gaps and lesions such as nodules. EDGE overcomes the limitations of other approaches such as DKG, which biases the VF dynamic analysis towards lines of HSV. PVG rotates the left VF to the upper part, which was found, in personal communications, to be less intuitive for clinicians in our group to interpret. Additionally, this method requires identifying the middle axis of the glottal area, which could represent a challenge, especially in pathological cases. In GVG, the lack of individual left-right presentation limits its use for studying asymmetrical vibratory behaviors between the left and right VFs. The methods of FT and GTG are the most difficult to interpret in pathological cases. They also exhibit variations in visualization associated with the selection of the frequency to display in the case of FT or the adjustment of brightness in GTG.

Furthermore, VF vibratory asymmetry derived from EDGE represents a pioneering attempt to study anterior-posterior amplitude and phase asymmetry. This analysis enables the identification of the vocal fold region exhibiting the highest degree of asymmetry, offering an advantage over methods documented in the literature that typically measure asymmetries at the middle of the glottal length [11], [12], [13]. For instance, in subjects FP03, FP04, and FP06, the middle asymmetry (at 50% of glottal length) is lower compared to the asymmetries in the posterior and anterior regions. Consequently, an analysis focused solely on local asymmetries could bias the study of the VF vibration. We propose that the mean absolute values of AA and PA along the length of the vocal fold offer the robustness required for exploring asymmetric VF behavior.

B. Limitations of EDGE

One limitation that EDGE shares with PVG and GVG is the need for glottal segmentation. Segmentation methods face challenges such as high inter-subject variability in VF vibration, and measuring noise and artifacts. In this work, we have implemented semi-automatic software to ensure accuracy in glottal segmentation. However, automating EDGE is desirable in future developments and deep learning algorithms will be a viable solution as proposed in [22], [35].

Another limitation of EDGE is its lack of temporal representation, a constraint that became apparent during the analysis of the UVFP case FP07. Unlike EDGE, other methods such as DKG, PVG, and GVG can identify temporal pattern variability within this case. The anterior-posterior phase asymmetry shown by EDGE in this case could be related to the higher dispersion observed in conditional EDGE (Fig. 6(d)), which reduces the reliability of this metric. However, further study of cases with similar behavior is needed to draw definitive conclusions.

The EDGE method relies on assumptions of stationary (or steady state) conditions in VF vibration. Therefore, the discussion and results presented in this work are constrained to the analysis of sustained vowel phonation. Running speech, in contrast, involves rapid changes in pitch, volume, and voice quality, leading to non-stationary VF vibrations. Consequently, the EDGE method is only intended for the study of sustained phonation rather than transient gestures, but it does allow for merging multiple repetitions of separate sustained phonatory gestures.

V. Conclusion

The probabilistic analysis method proposed in this study shows great potential for describing complex vibratory behavior through visual analysis. The results obtained from the normal and pathological groups suggest that EDGE-based features could differentiate between individuals with typical and pathological voices, particularly through a novel and comprehensive analysis of amplitude and phase asymmetries along the anterior-posterior glottal axis.

In this context, the new features generated by our proposal could provide researchers and clinicians with a valuable tool for quantifying irregularities in vocal fold dynamics and potentially yield intuitive glottal images that clinicians may find easier to interpret. However, future work should include comparative studies, user experience research, and clinical trials to validate the intuitive use of EDGE and its advantages over current state-of-the-art methods.

Additionally, the temporal independence exhibited by our methodology suggests potential for its application in videostroboscopy, which would provide a powerful clinical tool capable of enriching the assessment of vocal function for the larger clinical community. In turn, this could enhance the precision of clinical diagnoses, surgical procedures, and vocal therapy.

Authorized licensed use limited to: Universidad Tecnica Federico Santa Maria. Downloaded on March 17,2025 at 16:07:38 UTC from IEEE Xplore. Restrictions apply.

ACKNOWLEDGMENT

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

[1] D. D. Deliyski and R. E. Hillman, "State of the art laryngeal imaging: Research and clinical implications," *Curr. Opin. Otolaryngology Head Neck Surg.*, vol. 18, no. 3, pp. 147–152, Jun. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20463479http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2931794

[2] P. Woo, "Stroboscopy and high-speed video examination of the larynx," in *Laryngology* (Sataloff's Comprehensive Textbook of Otolaryngology Head & Neck Surgery), 1st ed., vol. 4, R. T. Sataloff, A. K. Lalwani, M. P. Fried, A. Tabaee, and M. S. Benninger, Eds. New Delhi, India: Jaypee Brothers Medical Publishers Ltd, 2016, ch. 18, pp. 193–234.

[3] D. D. Deliyski, M. E. Powell, S. R. Zacharias, T. T. Gerlach, and A. de Alarcon, "Experimental investigation on minimum frame rate requirements of high-speed videoendoscopy for clinical voice assessment," *Biomed. Signal Process. Control*, vol. 17, pp. 21–28, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1746809414001700

[4] M. E. Powell et al., "Efficacy of videostroboscopy and high-speed videoendoscopy to obtain functional outcomes from perioperative ratings in patients with vocal fold mass lesions," *J. Voice*, vol. 34, no. 5, pp. 769–782, 2019.

[5] P. Moore and H. Von Leden, "Dynamic variations of the vibratory pattern in the normal larynx," *Folia Phoniatrica et Logopaedica*, vol. 10, no. 4, pp. 205–238, 1958.

[6] G. Andrade-Miranda, Y. Stylianou, D. D. Deliyski, J. I. Godino-Llorente, and N. H. Bernardoni, "Laryngeal image processing of vocal folds motion," *Appl. Sci.*, vol. 10, no. 5, 2020, Art. no. 1556. [Online]. Available: https://www.mdpi.com/2076-3417/10/5/1556

[7] J. Lohscheller, U. Eysholdt, H. Toy, and M. Dollinger, "Phonovibrography: Mapping high-speed movies of vocal fold vibrations into 2-d diagrams for visualizing and analyzing the underlying laryngeal dynamics," *IEEE Trans. Med. Imag.*, vol. 27, no. 3, pp. 300–309, Mar. 2008.

[8] S.-Z. Karakozoglou, N. Henrich, C. d'Alessandro, and Y. Stylianou, "Automatic glottal segmentation using local-based active contours and application to glottovibrography," *Speech Commun.*, vol. 54, no. 5, pp. 641–654, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167639311001282

[9] G. Chen, J. Kreiman, and A. Alwan, "The glottaltopogram: A method of analyzing high-speed images of the vocal folds," *Comput. Speech Lang.*, vol. 28, no. 5, pp. 1156–1169, Sep. 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/25170187

[10] J. G. Švec and H. K. Schutte, "Videokymography: High-speed line scanning of vocal fold vibration," *J. Voice*, vol. 10, no. 2, pp. 201–205, 1996. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0892199796800476

[11] Q. Qiu, H. Schutte, L. Gu, and Q. Yu, "An automatic method to quantify the vibration properties of human vocal folds via videokymography," *Folia phoniatrica et logopaedica, Official Organ Int. Assoc. Logopedics Phoniatrics*, vol. 55, pp. 128–36, May 2003.

[12] D. D. Mehta, D. D. Deliyski, T. F. Quatieri, and R. E. Hillman, "Automated measurement of vocal fold vibratory asymmetry from high-speed videoendoscopy recordings," *J. Speech, Language, Hear. Res.*, vol. 54, no. 1, pp. 47–54, 2011. [Online]. Available: https://pubs.asha.org/doi/abs/10.1044/1092-4388%282010/10-0026%29

[13] D. D. Mehta, M. Zañartu, T. F. Quatieri, D. D. Deliyski, and R. E. Hillman, "Investigating acoustic correlates of human vocal fold vibratory phase asymmetry through modeling and laryngeal high-speed videoendoscopy," *J. Acoustical Soc. America*, vol. 130, no. 6, pp. 3999–4009, Dec. 2011.

[14] J. Lohscheller, H. Toy, F. Rosanowski, U. Eysholdt, and M. Döllinger, "Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos," *Med. Image Anal.*, vol. 11, no. 4, pp. 400–413, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361841507000369

[15] D. D. Deliyski, P. P. Petrushev, H. S. Bonilha, T. T. Gerlach, B. Martin-Harris, and R. E. Hillman, "Clinical implementation of laryngeal high-speed videoendoscopy: Challenges and evolution," *Folia Phoniatrica et Logopaedica, Official Organ Int. Assoc. Logopedics Phoniatrics*, vol. 60, no. 1, pp. 33–44, 2008, doi: 10.1159/000111802.

[16] G. Andrade-Miranda, N. H. Bernardoni, and J. I. Godino-Llorente, "Synthesizing the motion of the vocal folds using optical flow based techniques," *Biomed. Signal Process. Control*, vol. 34, pp. 25–35, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809417300022

[17] H. Ammar, "Optical flow based waveform for the assessment of the vocal fold vibrations," *Australas. Phys. Eng. Sci. Med.*, vol. 42, no. 1, pp. 91–109, Mar. 2019, doi: 10.1007/s13246-018-0717-9.

[18] Y. Maryn et al., "Intersegmenter variability in high-speed laryngoscopy-based glottal area waveform measures," *Laryngoscope*, vol. 130, no. 11, pp. E654–E661, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/lary.28475

[19] S. Granqvist and P.-Å. Lindestad, "A method of applying Fourier analysis to high-speed laryngoscopy," *J. Acoustical Soc. America*, vol. 110, no. 6, pp. 3193–3197, 2001.

[20] P. Aichinger, "Diplophonic voice-definitions, models, and detection," Ph.D. dissertation, Division of Phoniatrics-Logopedics, Department of Otorhinolaryngology, Medical University of Vienna, 2015.

[21] J. Neubauer, P. Mergell, U. Eysholdt, and H. Herzel, "Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial modes," *J. Acoustical Soc. America*, vol. 110, no. 6, pp. 3179–3192, 2001.

[22] A. M. Kist et al., "A deep learning enhanced novel software tool for laryngeal dynamics analysis," *J. Speech, Language, Hear. Res.*, vol. 64, no. 6, pp. 1889–1903, 2021.

[23] A. M. Kist, S. Dürr, A. Schützenberger, and M. Döllinger, "Openhsv: An open platform for laryngeal high-speed videoendoscopy," *Sci. Rep.*, vol. 11, no. 1, 2021, Art. no. 13760.

[24] M. Döllinger, D. Dubrovskiy, and R. Patel, "Spatiotemporal analysis of vocal fold vibrations between children and adults," *Laryngoscope*, vol. 122, no. 11, pp. 2511–2518, 2012. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/lary.23568

[25] C. Bohr et al., "Spatiotemporal analysis of high-speed videolaryngoscopic imaging of organic pathologies in males," *J. Speech, Language, Hear. Res.*, vol. 57, no. 4, pp. 1148–1161, 2014. [Online]. Available: https://pubs.asha.org/doi/abs/10.1044/2014_JSLHR-S-12-0076

[26] J. Unger, J. Lohscheller, M. Reiter, K. Eder, C. S. Betz, and M. Schuster, "A noninvasive procedure for early-stage discrimination of malignant and precancerous vocal fold lesions based on laryngeal dynamics analysis," *Cancer Res.*, vol. 75, no. 1, pp. 31–39, 2015. [Online]. Available: https://cancerres.aacrjournals.org/content/75/1/31

[27] J. Unger, M. Schuster, D. J. Hecker, B. Schick, and J. Lohscheller, "A generalized procedure for analyzing sustained and dynamic vocal fold vibrations from laryngeal high-speed videos using phonovibrograms," *Artif. Intell. Med.*, vol. 66, pp. 15–28, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0933365715001402

[28] P. Birkholz, "Glottalimageexplorer–an open source tool for glottis segmentation in endoscopic high-speed videos of the vocal folds," in *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung*, O. Jokisch, Ed. Dresden, Germany: TUD Press, Dresden, 2016, pp. 39–44.

[29] M. E. Díaz-Cádiz et al., "Estimating vocal fold contact pressure from raw laryngeal high-speed videoendoscopy using a hertz contact model," *Appl. Sci.*, vol. 9, no. 11, Jun. 2019, Art. no. 2384. [Online]. Available: https://www.mdpi.com/2076-3417/9/11/2384

[30] K. P. Murphy, *Machine Learning: A Probabilistic Perspective* (Series Adaptive Computation and Machine Learning). Cambridge, MA, USA: The MIT Press, 2012.

[31] M. Capiński and P. E. Kopp, *Measure, Integral and Probability*, vol. 14. Berlin, Germany: Springer, 2004.

[32] G. A. Alzamendi, S. D. Peterson, B. D. Erath, R. E. Hillman, and M. Zañartu, "Triangular body-cover model of the vocal folds with coordinated activation of the five intrinsic laryngeal muscles," *J. Acoustical Soc. America*, vol. 151, no. 1, pp. 17–30, Jan. 2022, doi: 10.1121/10.0009169.

[33] J. Lohscheller and U. Eysholdt, "Phonovibrogram visualization of entire vocal fold dynamics," *Laryngoscope*, vol. 118, no. 4, pp. 753–758, Apr. 2008. [Online]. Available: http://doi.wiley.com/10.1097/MLG.0b013e318161f9e1

[34] R. E. Hillman, C. E. Stepp, J. H. V. Stan, M. Zañartu, and D. D. Mehta, "An updated theoretical framework for vocal hyperfunction," *Amer. J. Speech- Lang. Pathol.*, vol. 29, no. 4, pp. 2254–2260, 2020, doi: 10.1044/2020_AJSLP-20-00104.

[35] M. Döllinger et al., "Re-training of convolutional neural networks for glottis segmentation in endoscopic high-speed videos," *Appl. Sci.*, vol. 12, no. 19, 2022, Art. no. 9791. [Online]. Available: https://www.mdpi.com/2076-3417/12/19/9791

Authorized licensed use limited to: Universidad Tecnica Federico Santa Maria. Downloaded on March 17,2025 at 16:07:38 UTC from IEEE Xplore. Restrictions apply.