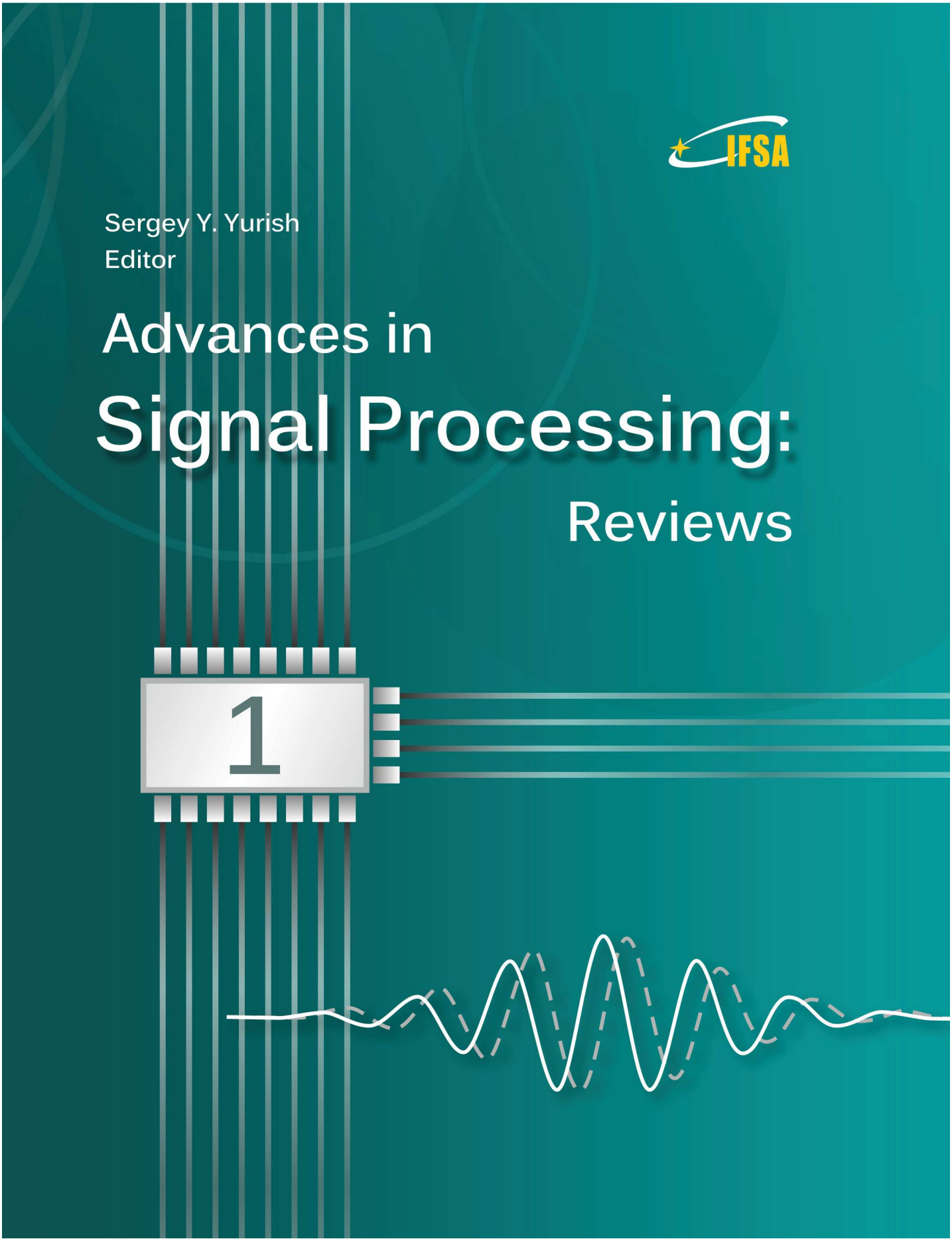
Vocal Folds Dynamics by Means of Optical Flow Techniques: A Review of the Methods

Gustavo Andrade-Miranda, Nathalie Henrich Bernardoni, Juan Ignacio Godino-Llorente, Henry Cruz


IFSA
Sergey Y. Yurish
Editor
Advances in
Signal Processing:
Reviews
1

Sergey Y. Yurish
Editor

Advances in Signal Processing: Reviews

Book Series, Volume 1

International Frequency Sensor Association Publishing

Chapter 13
Vocal Folds Dynamics by Means of Optical Flow Techniques: A Review of the Methods

Gustavo Andrade-Miranda, Nathalie Henrich Bernardoni, Juan Ignacio Godino-Llorente and Henry Cruz

13.1. Introduction

13.1.1. Voice Assessment and Endoscopy

A healthy voice is crucial for people's daily life, especially for professional voice users. Voice assessment and diagnosis of potential disorders is typically carried out in the clinics by means of different objective tools, including acoustic analysis and vocal-folds visualization using videoendoscopic techniques, perceptual gradings, and self-evaluation questionnaires [1]. According to the American Academy of Otolaryngology-Head and Neck-Surgery, the basic protocol to evaluate a patient with a voice disorder has to include a rigorous clinical history, physical examination, and visualization of the larynx via laryngoscopy [2]. In comparison with a physical examination, or an acoustic analysis, only laryngoscopic techniques allow a direct visualization of vocal folds in motion and the determination of voice disorder's ethiology. Hence, improving the techniques for laryngeal functional examination has become a current challenge and the aims of advanced scientific research [3-5]. Fig. 13.1 schematically illustrates the laryngoscopic procedure to record vocal-folds dynamics using a rigid endoscope (90°).

The most common tool used by clinicians for laryngeal imaging is Laryngeal Videostroboscopy (LVS). It has been used to examine subtle abnormalities along the vocal-folds vibratory margin, such as cysts or scars, and to detect subtle problems such as mild inflammation, vocal-folds swelling, white patches or excessive mucus. However, its low recording frame rate of 25 frames per second (fps) does not enable to highlight all vibratory peculiarities of dysfunctional voices [6]. A higher video frame rate is often necessary to an in-depth assessment of vocal-folds vibratory features. It is made possible

Gustavo Andrade-Miranda
University of Guayaquil, Faculty of Industrial Engineering. Av. Las Aguas, Guayaquil, Ecuador

by Laryngeal High-Speed Videoendoscopy (LHSV), which addresses the limitations of LVS.

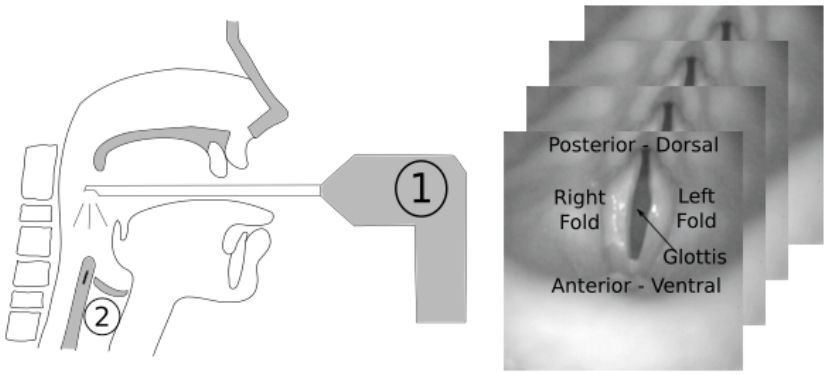


Fig. 13.1. Illustration of clinical video recording of vocal-folds dynamics. Left panel: recording situation including camera and rigid endoscope 90° (1). Right panel: laryngeal video depicting vocal-folds movements (2).

Nowadays, due to the fast-growth of imaging technology, it is possible to find high-speed cameras with frame rates up to 10000 fps. LHSV is currently regarded as a superior method with respect to LVS for the assessment of vocal-folds vibratory movements due to several reasons [7]:

i. LHSV is applicable to the assessment of unstable phonations such as transient, subharmonic, or aperiodic phonations. It is more useful for investigating vocal-folds pathologies;

ii. LHSV allows vocal assessment of male and female phonation in most of the clinical scenarios, such as phonation at normal pitch and loudness, onset and offset, high and low pitch in modal register, breathy and pressed phonation. It provides greater validity for assessment of intracycle and intercycle vibratory characteristics compared with LVS;

iii. LHSV data can be analyzed with a greater variety of methods than LVS data, enabling more interpretable and extensive evaluation on both a qualitative and a quantitative basis.

The applicability of LHSV to voice and its technological developments are active fields of research. It has provided valuable insight into the mechanisms of phonation both in normal and in voice disorders. The use of LHSV in combination with advanced image-processing techniques is the most promising approach to investigate vocal-folds vibration and laryngeal dynamics in speech and singing.

13.1.2. Scientific and Clinical Applications of Laryngeal High-Speed Videoendoscopy

The pioneering works using LHSV date back to 1958 with the seminal works of Timcke, Lenden and Moore [8-10]. They measured the vibratory amplitude of each vocal fold separately, and plotted its variations as a function of time. They measured glottal parameters such as open quotient (OQ) and speed quotient (SQ) during opening and closing phases. They reported observations of normal and abnormal vocal-folds vibrations during phonation. They found that changes in air pressure exert a considerable influence on individual components of the vibratory cycle. They also found that the anatomical configuration of the vocal folds plays a major role in the vibratory patterns, and highlighted the needs for descriptive terms to objectively measure the vocal-folds vibratory patterns. In these times, each frame within the glottal cycle was processed manually, which was very precise but time consuming. In [11] a machine-aided procedure to extract glottal waveform and other glottal measurements was reported.

Twenty years afterwards, videokymography has emerged [12]. Clinicians had access to the vocal-folds vibratory pattern of a single line with commonly-used low-speed cameras. This smart approach provided an in-depth and real-time understanding of vocal function. It was complemented with the availability of high-speed cameras. Subsequent works [13-17] pointed out the advantages of high-speed monitoring of vocal-folds vibration for detecting asymmetries, transients, breaks, opening phase events, closing phase events, and irregularities. Correlation between vocal-folds vibratory patterns and voice quality could be studied on speakers in combining image-based analyses and acoustic-signal processing [18, 19]. LHSV was combined with biomechanical models to quantify the spatio-temporal vibrations of the vocal folds. In [20], the two-mass model proposed in [21] was used with LHSV to estimate physiological parameters such as vocal-folds mass and stiffness. A genetic algorithm was used to match the behaviour of a two-mass model parameters to that of a real vocal fold [5]. Different parameters were extracted, including masses, spring and damper constants. In [22], an automatic optimization procedure was developed for fitting a multi-mass model [23] to the observed vocal-folds oscillations, and estimating stiffness and mass distribution along the entire vocal folds. One of the latest works [24] used an optimization method, which combines genetic algorithms and a quasi-Newton method to extract some physiological parameters of the vocal folds and reproduce some complex behaviors, such as the ones that occur in different types of pathologies.

LHSV also has been used to highlight the importance of visualizing mucosal-wave propagation for an accurate diagnosis and evaluation of voice disorders. In [25], the presence of atypical magnitude and symmetry of the vocal-folds mucosal wave was shown for normal speakers. In [26], the mucosal-wave propagation was detected and quantified by combining image-processing techniques with physiological knowledge of its lateral component, with the aim to replace subjective assessment of mucosal wave in the clinical environment. The authors in [27] discussed the benefits, the disadvantages, and the clinical applicability of different mucosal-wave measurement techniques. They found the necessity of additional research to broaden the use of LHSV for an accurate and objective diagnosis of vocal disorders.

Different singing styles have also been analyzed using LHSV. The bass-type Kargyraa mechanism in Mongolian throat singing was studied [28], demonstrating vibratory movements of both vocal and vestibular folds during singing contributing to subharmonics in the voice acoustic signal. Similar vibratory behaviour was evidenced in Sardinian Bassu singing [29]. In [30] the characteristics of hard-rock singing, also known as distorted singing, were investigated. Modulations of vocal-folds vibrations by means of periodic or aperiodic motion in the supraglottic mucosa were found to add special expressivity to loud and high tones in rock singing.

The contribution of vestibular folds has been evidenced in shouted speech and several types of singing by means of LHSV [31]. In this study, physiological data derived from LHSV image processing has been used as an input signal to a physical model of phonation.

LHSV has been used for clinical voice-research purposes. For instance, the applicability of LHSV to diagnose functional voice disorders was demonstrated in [4], where non-stationary activities of vocal folds during onset were investigated and described with two variables for pathological and normal voices. The first variable describes the growth of vocal-folds vibratory amplitude during phonation onset, and the second one draws conclusions on voice efficiency with respect to the necessary subglottal pressure and myoelastic forces. A computer-aided method was presented in [32] for automatically and objectively classifying individuals with normal and abnormal vocal-folds vibratory patterns. First, a set of image-processing techniques was employed to visualize vocal-folds dynamics. Later, numerical features were derived, capturing the dynamic behavior and the symmetry in oscillation of the vocal folds. Lastly, a support vector machine was applied to classify between normal and pathological vibrations. The results indicate that an objective analysis of abnormal vocal-folds vibration can be achieved with considerably high accuracy. In [3] a set of parameters were proposed to differentiate between healthy and organic voice disorders in men speakers. The parameters were chosen based on spatio-temporal information of the vocal-folds vibratory patterns. The spatial parameters provided information about opening and closing phase. Meanwhile, the temporal parameters reflected the influence of organic pathologies on the periodicity of glottal vibrations. The results obtained suggest that the differences between male healthy voices and male organic voice disorders may be more pronounced within temporal characteristics that cannot be visually detected without LHSV. The authors in [33] reported a procedure to discriminate between malignant and precancerous lesions by measuring the characteristics of vocal- folds dynamics on a computerized analysis of LHSV sequences.

13.1.3. From a Big Amount of Data to a Synthesized View

LHSV makes it possible to characterize laryngeal-tissue dynamics and vocal-folds vibratory patterns. However, this can not be done easily by simply observing the successive recorded frames. With the appropriate image-processing techniques, the time-varying data can be synthesized in a few static images, or in a unidimensional temporal sequence. In this way, clinicians or researchers can follow the dynamics of anatomical features of interest without substituting the rich visual content with scalar

numbers. The literature reports several proposals to represent the LHSV information in a more simple way. They are able to objectively identify the presence of organic voice disorders [3], classify functional voice disorders [32], vibratory patterns [13], discriminate early stages of malignant and precancerous vocal folds lesions [33], and other applications [15, 16]. These representations improve the quantification accuracy, facilitate the visual perception, and increase the reliability of visual ratings while preserving the most relevant characteristics of glottal vibratory patterns. Such representations are named as facilitative playbacks [34]. They can be grouped in local- or global-dynamics playbacks depending on the way they assess the glottal dynamics. Local-dynamics playbacks analyze vocal-folds behavior along a single line, while global-dynamics playbacks present glottal vibrations along the whole vocal-folds length. Table 13.1 presents the main studies carried out to synthesize vocal-folds vibratory patterns.

Table 13.1. Playbacks proposed in the literature to synthetize vocal-folds vibratory patterns.

Author	Year	Playback	Dynamics
Timcke et al. [9]	1958	Glottal area waveform	Global
Westphal et al. [35]	1983	Discrete Fourier Transform Analysis	Global
Švec and Schutte [12]	1996	Videokymography	Local
Palm et al. [36]	2001	Vibration Profiles	Global
Neubauer et al. [37]	2001	Empirical Orthogonal Eigen-functions Analysis	Global
Li et al. [38]	2002	Eigenfolds Analysis	Global
Zhang et al. [39]	2007	Non linear Dynamics Analysis	Global
Lohscheller et al. [40]	2007	Vocal Folds Trajectories	Local
Yan et al. [19]	2007	Hilbert Transform Analysis	Global
Deliyski et al. [34]	2008	Mucosal Wave Kymography	Local
Lohscheller et al. [16]	2008	Phonovibrogram	Global
Sakakibara et al. [41]	2010	Laryngotopography	Global
Karakozoglou et al. [42]	2012	Glottovibrogram	Global
Unger et al. [43]	2013	Phonovibrogram Wavegram	Global
Ikuma et al. [44]	2013	Waveform Decomposition Analysis	Global
Rahman et al. [45]	2014	Dynamic time Warping Analysis	Global
Chen et al. [46]	2014	Glottaltopogram	Global
Hersbt et al. [47]	2016	Phasegram Analysis	Global

Despite of great current advances, the generalized use of LHSV into routine clinical practice still requires additional developments. Therefore, the researchers need new methods for data visualization to overcome the drawbacks of existing ones, providing simultaneously features that would integrate time dynamics, such as: velocity, acceleration, instants of maximum and minimum velocity and vocal-folds displacements during phonation. In this chapter, the use of Optical Flow (OF) to characterize glottal dynamics of LHSV sequences in a compact manner is presented and discussed. In Section 13.2, the general principles of computation are explained, and its applicability to the LHSV problem is discussed. Section 13.3 describes the databases of LHSV sequences used in the study, and presents several OF-based playbacks focusing on local/global

dynamics, and glottal velocity. Section 13.4 evaluates the proposed OF playbacks with regard to commonly-used ones. Conclusions and perspectives are given in Section 13.5.

13.2. Optical-Flow Computation

13.2.1. Theoretical Background

The video motion analysis is one of the main tasks to precisely and faithfully models the dynamical behavior of observed objects, such as motion is typically represented using motion vectors, also known as vector displacements or OF.

OF estimation has been used for the last 35 years since the seminal works of Horn Schunck and Lucas-Kanade [48, 49]. Many innovative methods have been proposed to solve its computation [50]. The OF has been used in a variety of situations, including time-to-collision calculations, segmentation, structure of objects, motion analysis, image registration [51] or moving-objects detection. It has been applied in biomedical context to analyze dynamic properties of tissues or cellular objects, deformation of organs [52], estimation of blood flow [53], detection of cell deformation in microscopy [54], motion estimation of cardiac ultrasound images [55, 56], analysis of displacements detection of breast-tumor cancer [57], tracking colonoscopy videos [58], among others.

However, to date, there is no unique method to characterize all the possible motion scenarios at minimal computational cost, including those with disturbing phenomena such as lighting changes, reflection effects, modifications of objects properties, motion discontinuities, or large displacements. The definition of OF takes its roots from a physiological description of images formed on human retina as a change of structured light caused by a relative motion between the eyeball and the scene. In the field of computer vision, Horn-Schunck defined OF in [48] as "the apparent motion of brightness patterns observed when a camera is moving relative to the objects being imaged". Given an image sequence $I(\mathbf{x}, t)$, the basic OF assumption is that at any pixel $\mathbf{x}_{ij}$, at time t_k, the intensity $I(\mathbf{x}_{ij}, t_k)$ would remain constant during a short interval of time Δt_k, the so-called Brightness Constancy Constraint (BBC) or data term:

$$I(\mathbf{x}_{ij}, t_k) = I(\mathbf{x}_{ij} + \vec{\boldsymbol{w}}(\mathbf{x}_{ij}, t_k), t_k + \Delta t_k), \forall \mathbf{x}_{ij}\,, \tag{13.1}$$

where $\vec{\boldsymbol{w}}(\mathbf{x}_{ij}, t_k)$ is the vector displacement of $\mathbf{x}_{ij}$ in a time interval Δt_k. The vector displacement $\vec{\boldsymbol{w}}(\mathbf{x}_{ij}, t_k)$ has two components: one in the x-axis direction $u(\mathbf{x}_{ij}, t_k)$ and another in the y-axis direction $v(\mathbf{x}_{ij}, t_k)$. Consequently, the total motion field at time t_k is defined as:

$$\mathcal{W}(\mathbf{x}, t_k) = (U(\mathbf{x}, t_k), V(\mathbf{x}, t_k)\,)\ \forall \vec{\mathrm{w}}(\mathbf{x}_{ij}, t_k), \tag{13.2}$$

where $U(\mathbf{x}, t_k)$ and $V(\mathbf{x}, t_k)$ are the components in the x- and y-axis of the total motion field, respectevely. The BCC provides only one equation to recover the two unknown components of $\mathcal{W}(\mathbf{x}, \mathrm{t_k})$. Therefore, it is necessary to introduce an additional constraint

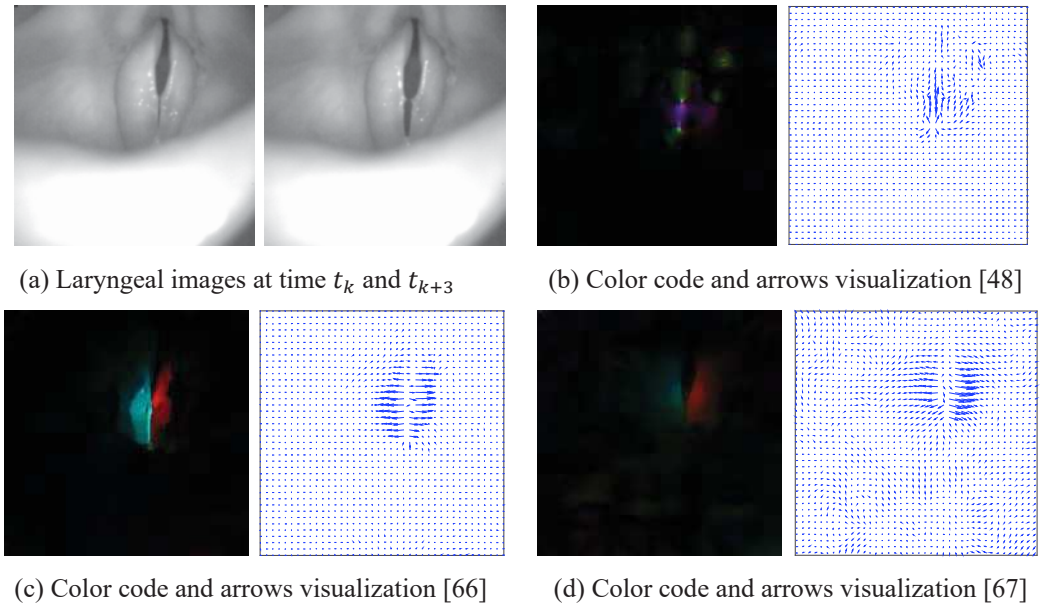
encoding a priori information of $\mathcal{W}(\mathbf{x}, t_k)$. Such information comes from the spatial coherency imposed by either local or global (regularization term).

In practice, the BBC assumption is an imperfect photometric expression of the real physical motion in the scene that cannot be applied in case of changes in the illumination sources of the scene, shadows, noise in the acquisition process, specular reflections or large and complex deformations. Therefore, several matching costs (also called penalty functions) have been explored to overcome the drawback of the BBC, in particular its sensitivity to noise and illumination changes. Over the last few years, the number of OF algorithms with increasingly good performance has grown dramatically and it becomes difficult to summarize all contributions and their categories. For instance, studies back to the nineties [50, 59] classify the OF into six groups: intensity-based differential methods, frequency methods, correlation-based method, multiple motion methods and temporal refinement methods. On the other hand, some of the last studies focus their attention on variational approaches [60-62] since they are quite versatile, allowing one to model different forms of flow fields by combining different data and regularization terms. But more important is that they have shown the most accurate results to the OF problem in the literature [63]. Other OF algorithms with outstanding performance are based on discrete optimization [64]. Its main advantage over the continuous approaches is that they do not require differentiation of the energy and can thus handle a wider variety of data and regularization terms. On the counterpart, a trade-off has to be found between the accuracy of the motion labelling and the size of the search space. For a most recently survey about the OF techniques, formulation, regularization and optimization refer to [65].

13.2.2. Optical Flow Evaluation

There are two main visualization techniques to assess OF: via arrow visualization or via color coding. The first one represents the displacement vectors and offers a good intuitive perception of physical motion. However, it requires to under-sample the motion field to prevent the overlapping between displacement vectors. Meanwhile, the color coding visualization associates a color hue to a direction and a saturation to the magnitude of the displacement vectors. It allows a dense visualization of the flow field and a better visual perception of subtle differences between neighbor motion vectors. Fig. 13.2 depicts the arrow and color code visualization using three different OF formulations: Horn and Schunck formulation [48]; Zach and Pock formulation, which is based on the minimization of a functional containing a data term using the L1 norm and a regularization term using the total variation of the flow [66]; and Drulea and Nedevschi formulation, which is based on correlation transform [67].

Additionally, there are two quantitative evaluation methods based on error metrics that are used to compare the performance of the OF methods when a ground truth is available, namely the Angular Error (AE) and the Endpoint Error (EPE) [68]. AE measures the 3D angle error between the estimated and reference vectors; meanwhile EPE measures their Euclidean distance.



(a) Laryngeal images at time t_k and t_{k+3} (b) Color code and arrows visualization [48]

(c) Color code and arrows visualization [66] (d) Color code and arrows visualization [67]

Fig. 13.2. Color code and arrows visualization. (a) Two laryngeal images taken during the opening phase of the vocal folds at time t_k and t_{k+3}; (b) OF based on Horn and Schunk [48]; (c) OF based on Total variation L1 [66]; (d) OF based on Correlation transform [67].

13.2.3. Optical Flow in LHSV

The purpose of LHSV analysis is to characterize vocal-folds motion by identifying their movements from one frame to the followings. However, this task requires to isolate the glottis and track it along time. Advantageously, OF computation allows the possibility to track unidentified objects solely based on its motion, with no need of additional segmentation techniques.

The LHSV sequences present challenging scenarios such as complex reflectance phenomena that appear due to intrinsic mucosal surface properties, motion discontinuities due to mucosal-wave dynamics and occlusion in the glottal-area region. On the other hand, the OF accuracy is improved by the high frame rate of LHSV, since it reduces the temporal aliasing not only for areas with large displacements but also for areas with small displacements and high spatial frequencies. Additionally, the BBC assumption becomes even more valid with high frame rates. In two consecutive frames the OF should describe precisely the vocal-folds motion pattern. The motion-field direction is expected to be inwards during glottal closing phase and outwards during glottal opening. In order to illustrate this idea, Fig. 13.3 presents a synthetic representation of vocal-folds motion along the glottal main axis for two consecutive frames during the opening phase.

Additionally, the motion-field fluctuations over time have to reflect the glottal dynamics solely. In order to prove this fact, the OF changes in magnitude with respect to x-axis are analyzed in a line located in the middle of glottal main axis for a complete glottal cycle (see Fig. 13.4). As expected, the flow is concentrated in the glottal region where strongest

movements occur. Another remarkable feature is the valley formed between two peaks. The valley can be understood as the region inside the glottis, in which the motion field is zero. The two peaks can be interpreted as the pixels along the selected line with maximal positive and negative displacements. Despite its suitability to the problem under study, the use of OF for assessing the vocal folds dynamics has only recently been introduced in [69, 70]. Nevertheless, the authors in [71] already used motion-estimation techniques to describe vocal-folds deformation, but only around glottal edges.

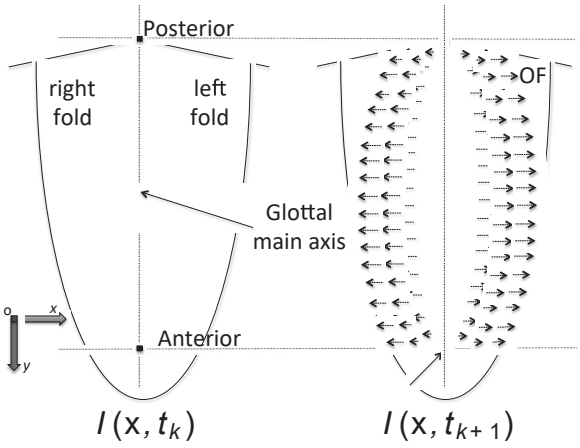


Fig. 13.3. Illustration of a synthetic motion field along the glottal main axis, between anterior and posterior parts of the vocal folds, in two consecutive instants of time, t_k and t_{k+1}.

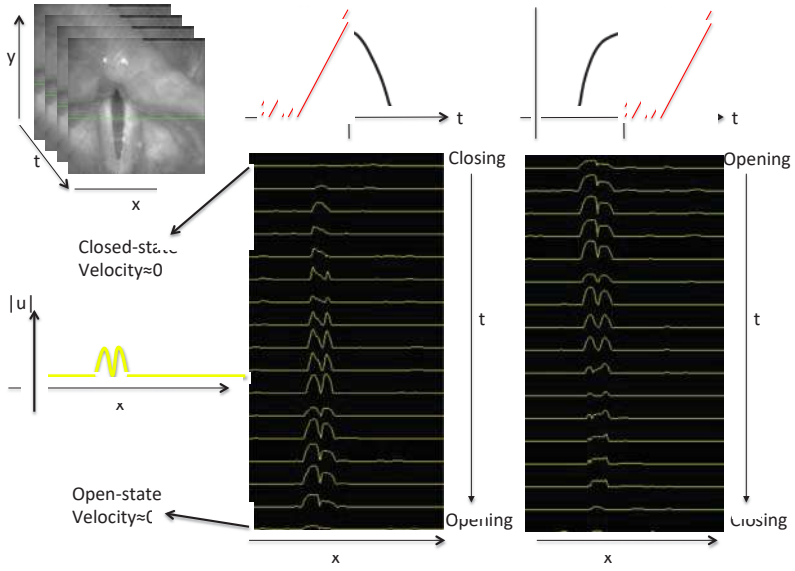


Fig. 13.4. Fluctuations of u along one line located in the middle of the glottal main axis for a complete glottal cycle.

13.3. Application of OF Computation to LHSV Image Analysis

13.3.1. Database

The recording took place at the University Medical Center Hamburg-Eppendorf (UKE) in Germany [42] and two male subjects (one speaker, one singer) participated in the experiment. Highspeed laryngeal images were acquired by means of a Wolf high-speed cinematographic system. The system consists of synchronized high-speed, audio and EGG recordings. The LHSV sequences were filmed by a rigid endoscope (Wolf 90 E 60491) equipped with a continuous source of light (Wolf 5131) driven by optic fiber. The database is composed for 60 high-speed sequences that are sampled at either 2000 or 4000 fps with a spatial resolution of 256×256 pixels.

The LHSV sequences include different phonatory tasks: sustained sounds with specific voice qualities (creaky, normal, breathy, pressed), pitch glides, sung vowels at different pitches and loudness. Additionally, they cover a huge variety of vocal-folds vibratory movements, including symmetrical and asymmetrical left-right movements, transients, aperiodicities, and antero-posterior modes of vibration. The processed sequences are composed of 501 frames, which correspond to roughly 125 msec. of phonation.

13.3.2. Image Processing Implementation

In order to obtain a more accurate information, reduce computational burden and mitigate the effect produced by noisy regions, the OF has been computed only inside a region of interest (ROI). Such region can be detected automatically or manually to include only vocal-folds area.

The OF techniques used for the implementation of the playbacks are Total Variation L1 Optical-Flow (TVL1-OF) [66], Motion Tensor Optical-Flow (MT-OF) [72] and Lukas Kanade Optical-Flow (LK-OF) [49]. The principal reason for this selection is to explore the performance of different OF implementations since these methods use different strategies to deal with the complex reflectance phenomena and motion discontinuities. Other algorithms were also explored [48, 67] but due to computational burden needed to process a whole video and similarities in the computation of the flow field with the aforementioned, they were not included in the OF-based playback evaluation.

Although TVL1-OF and LK-OF are based on the BBC assumption, they differ in the approach followed to compute OF. TVL1-OF uses a spatial coherence constraint for the global form of the motion field, which is imposed by an explicit regularization term. Contrariwise, LK-OF uses a parametric approach where the flow is computed in small squared or circular patches, the center of which is taken as velocity vector. Meanwhile, MT-OF does not have a direct connection with BBC-based methods since the flow field is computed by orientation tensors.

The implementation provided in the C++ OpenCV library was adopted for TVL1-OF and MT-OF flow computation. Since LK-OF is one of the fastest algorithms to compute OF,

it was programmed in Matlab. The implementation procedure is represented graphically in Fig. 13.5 and the computation of each playback is explained below.

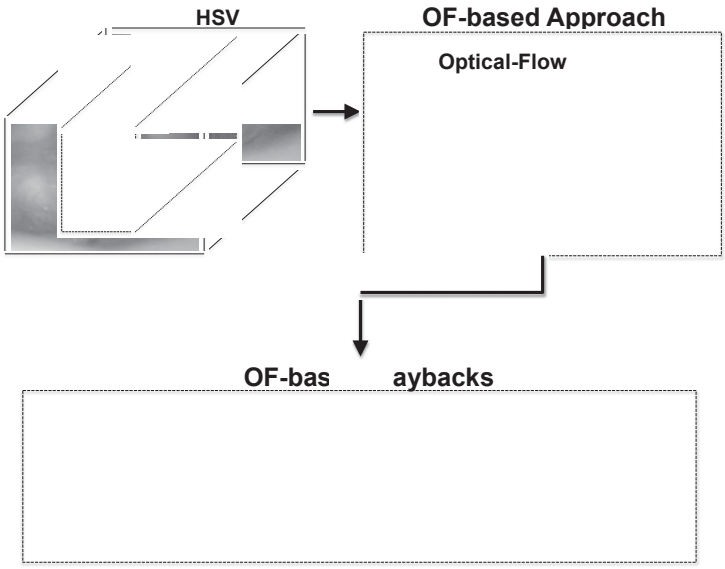


Fig. 13.5. Graphical representation of the procedure followed to compute OF-based playbacks.

13.3.3. Optical Flow Facilitative Playbacks

Three facilitative playbacks can be extracted: Optical Flow Kymogram (OFKG), which depicts local dynamics along one line; Optical Flow Glottovibrogram (OFGVG) that represents global dynamics along the whole vocal-folds length; and Glottal Optical Flow Waveform (GOFW), which plots glottal velocity.

13.3.3.1. Optical Flow Kymogram (OFKG)

The OFKG playback shows the direction and magnitude of vocal-folds motion in a single line. It follows the same idea as Digital Kymogram (DKG) [13] to compact LHSV information. However, the information used to synthesize the data comes from displacements produced in the x-axis ($U(\mathbf{x}, t_k)$) at each time t_k . For rightwise displacements, the direction angle ranges from $[-\pi/2, \pi/2]$, and it is coded with red intensities. Conversely, the angle for leftwise displacements ranges from $[\pi/2, 3\pi/2]$ and is coded with blue tonalities. The OFKG playback is depicted in Fig. 13.6 for a sequence of six glottal cycles.

13.3.3.2. Optical Flow Glottovibrogram (OFGVG)

The OFGVG playback represents the vocal-folds global dynamics by plotting glottal-velocity movement per cycle as a function of time. The OFGVG playback has the

goal to complement the spatiotemporal information provided by common techniques (GVG, PVG), adding velocity information of vocal-folds cycles. It is obtained by averaging each row of OF x component ($U(\mathbf{x}, t_k)$) and representing it as a column vector. This procedure is repeated along time for each new frame. Fig. 13.7 depicts the graphic representation of OFGVG playback.

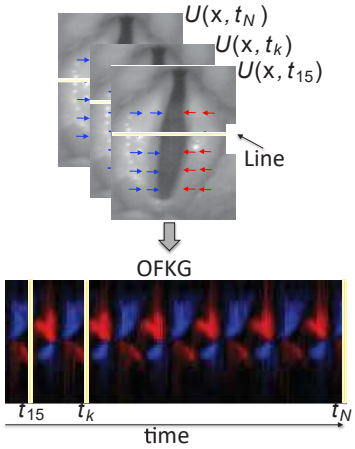


Fig. 13.6. Schematic view of an OFKG playback for the line represented in yellow, which is located in the median part of the vocal folds; the new local playback distinguishes the direction of motion (rightwise: red; leftwise: blue).

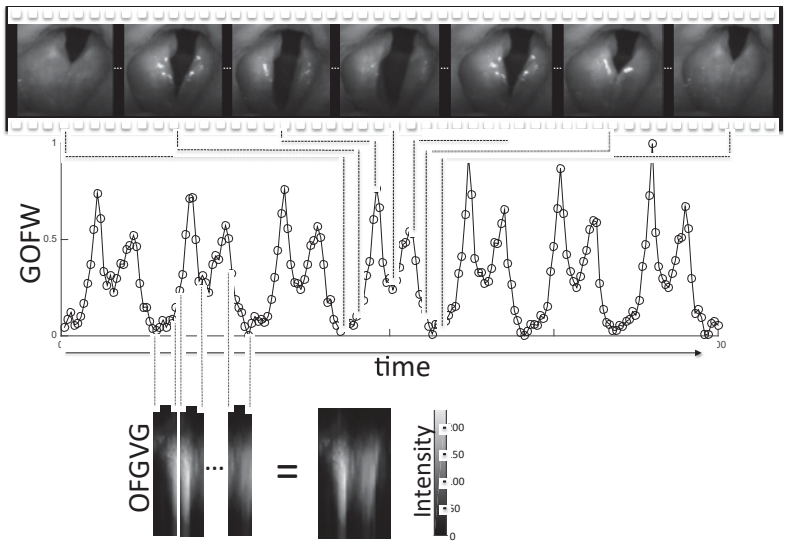


Fig. 13.7. First row: frames representation of one glottal cycle. Second row: schematic view of GOFW. Each point in the playback (dark circles) is obtained by averaging the absolute magnitude of $U(\mathbf{x}, t_k)$. Third row: schematic view of one OFGVG cycle. Dark regions indicate no velocity ($U(\mathbf{x}, t_k) = 0$).

13.3.3.3. Glottal Optical Flow Waveform (GOFW)

The GOFW playback is an 1D representation of glottal velocity. It is computed following the same criteria of the Glottal Area Waveform (GAW) but averaging the absolute magnitude of $U(\mathbf{x}, t_k)$. Additionally, overlapping this information with GAW highlights the velocity variation in each instant of the glottal cycle. The second row of Fig. 13.7 explains schematically how GOFW is computed, showing different velocity instants (black circles).

13.3.3.4. Vocal Folds Displacements Trajectories

The Vocal Folds Displacement Trajectories (VFDT) follow the same framework introduced in [40] with the difference that the displacement accuracy is measured rather than the distance between vocal-folds edges and glottal axis. Firstly, a trajectory line $\mathbf{L}(t_k)$ at time t_k, which intersects perpendicularly with glottal main axis $\mathbf{G}(t_k)$ in a predefined point $\mathbf{g}_{pc}(t_k)$, is defined and updated for each image:

$$\mathbf{g}_{pc}(t_k) = \mathbf{p}(t_k) + \big(\mathbf{p}(t_k) - \mathbf{a}(t_k)\big)\left(\frac{pc(\%)}{100\ \%}\right) \in \mathbf{G}(t_k), \tag{13.3}$$

where $\mathbf{p}(t_k)$ is the posterior commissure, $\mathbf{a}(t_k)$ the anterior commissure and the subscript $pc(\%)$ indicates a percentage of the total length of the glottal main axis. Following, the intersection between vocal-folds edges, $\mathbf{C}^{l,r}(t_k)$, and trajectory line, $\mathbf{L}(t_k)$, is computed (Fig. 13.8). Then, the displacements trajectories at time t_k and position $pc(\%)$ are defined as:

$$\hat{\delta}_{\mathcal{W}}^{l,r}(pc, t_k) = \mathcal{W}\big(\mathbf{c}_{pc}^{l,r}(t_k)\big), \tag{13.4}$$

$\mathbf{c}_{pc}^{l,r}(t_k)$ represents the intersection between vocal-folds edges $\mathbf{C}^{l,r}(t_k)$ and trajectory line $\mathbf{L}(t_k)$. From displacements trajectories, two additional trajectories can be derived: $\hat{\delta}_{\mathcal{W}_\mathrm{u}}^{l,r}(pc, t_k) = U\big(\mathbf{c}_{pc}^{l,r}(t_k)\big)$ and $\hat{\delta}_{\mathcal{W}_\mathrm{v}}^{l,r}(pc, t_k) = V\big(\mathbf{c}_{pc}^{l,r}(t_k)\big)$. However, as glottal edges have a motion pattern mainly perpendicular to the glottal main axis, $\hat{\delta}_{\mathcal{W}_\mathrm{v}}^{l,r}(pc, t_k)$ is negligible. Hence $\hat{\delta}_{\mathcal{W}}^{l,r}(pc, t_k)$ reflects primarly the fluctuations along t_k produced by $\hat{\delta}_{\mathcal{W}_\mathrm{u}}^{l,r}(pc, t_k)$. From now, both termns are used indistinctly and denoted for simplicity only as $\hat{\delta}_{\mathcal{W}}^{l,r}(pc, t_k)$. The graphical procedure followed to plot $\hat{\delta}_{\mathcal{W}}^{l,r}(pc, t)$ is described in Fig. 13.8, the VFDT is positive when glottal edges are moving from right to left, and contrariwise, negative when edges are moving from left to right.

13.3.4. Reliability Assessment of Optical Flow Playbacks

Due to the high amount of data in LHSV and the complexity of vocal-folds motion, it is difficult to create a ground-truth to evaluate OF performance [65]. Therefore, it is necessary to find alternative options to assess the reliability of the proposed new playbacks. An intuitive way to evaluate the accuracy of OF playbacks is to compare them

with those obtained using glottal-segmentation algorithms, since both results should be related. This premise comes from the fact that these two techniques represent the motion originated in the vocal folds, with the difference that the motion is reflected only on vocal-folds edges in glottal segmentation, while OF-based techniques analyze the entire vocal-folds region.

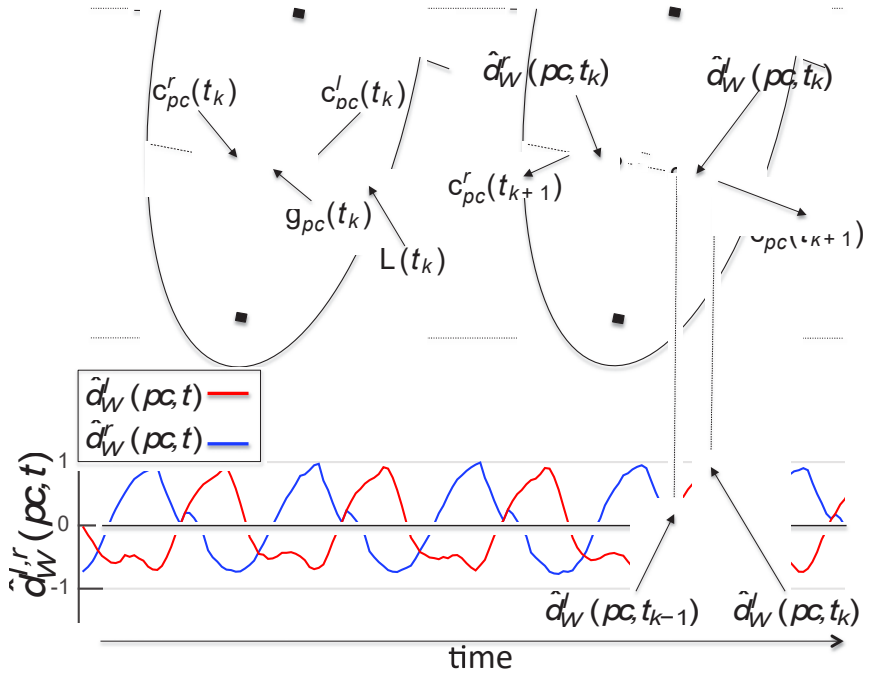


Fig. 13.8. Schematic procedure to compute VFDT during glottal opening phase.

For this purpose, the database was segmented automatically, resulting in both well-segmented videos and videos with minor errors in the segmentation. In this way, the benefits of OF playbacks are explored when the segmentation is not fully reliable. Three assessments are carried out. Firstly, the VFDT obtained by OF are correlated with the one obtained via glottal segmentation, which are defined as:

$$\hat{\delta}^{l,r}_{\text{seg}}(pc,t_k) = \mathbf{c}^{l,r}_{pc}(t_{k+1}) - \mathbf{c}^{l,r}_{pc}(t_k). \tag{13.5}$$

Since three different OF methods are used, $\hat{\delta}^{l,r}_{\text{seg}}(pc,t)$ is compared with each of them. The OF displacement trajectories are renamed as: $\hat{\delta}^{l,r}_{\text{TVL1}}(pc,t)$, $\hat{\delta}^{l,r}_{\text{MT}}(pc,t)$ and $\hat{\delta}^{l,r}_{\text{LK}}(pc,t)$ for TVL1-OF, MT-OF and LK-OF respectively. All displacement trajectories are computed in the medial glottal-axis position ($pc = 50$ %). The second assessment tries to find out the similarities of traditional playbacks with respect to OF playbacks by visually analyzing their common features.

13.4. Assessment of OF-Based Playbacks

13.4.1. Displacements Trajectories Comparison

The correlation between the segmentation trajectory and OF-based trajectories is depicted in Fig. 13.9. Each point of the graphic corresponds to the correlation of one LHSV sequence. Best correlations are obtained when TVL1-OF is compared with the segmentation. The average correlation achieved for that case is 0.74 while the average correlations with LK-OF and MT-OF only reached values of 0.51 and 0.63, respectively. The greatest correlation is 0.98, which is obtained with TVL1. Meanwhile, the values reached with LK-OF and MT-OF do not exceed 0.93. Additionally, 62 % of the trajectories computed via TVL1-OF presented a correlation greater or equal than 0.8 while only 23 % and 8 % of the trajectories reached this value using LK-OF and MT-OF respectively. On the other hand, only 8 VFDT computed with TVL1-OF have values below to 0.5, representing 13 % of the videos in the database.

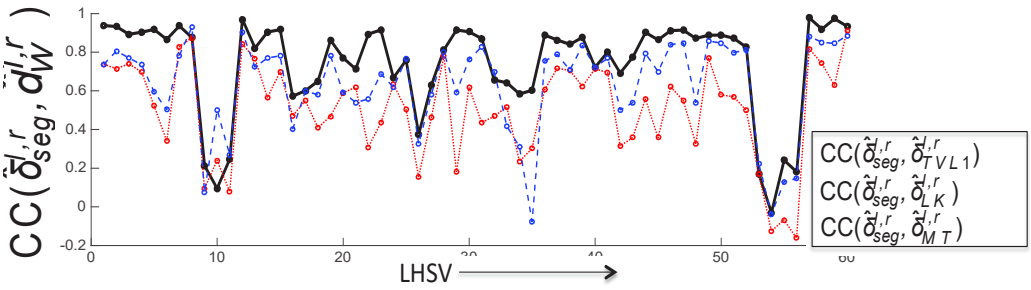


Fig. 13.9. Correlation between OF trajectories and segmentation trajectory for each sequence in a line located at *pc* = 50 %.

Fig. 13.10 illustrates four displacement trajectories for a normal phonatory task. The coherence in the vocal-folds dynamic can be noticed when segmentation and OF are compared. However, the amplitude and shape of the trajectories differ more when the segmentation is contrasted with LK-OF and MT-OF flows. Contrariwise TVL1-OF looks more robust and its shape seems like the trajectory of the segmentation.

Lastly, the fluctuations of the VFDT using TVL1-OF and segmentation methods are studied in detail for two different phonatory tasks (breathy and creaky) in *pc*=50 % (see Fig. 13.11). The trajectories computed via TVL1-OF are smoother than the ones obtained via segmentation but the shape and the amplitude of both are comparable. Additionally, during a short period of time TVL1-OF (regions enclosed by dashed lines in black at Fig. 13.11) presents some fluctuations originated from a vibration of the vocal folds, while the segmentation-based method does not show any motion. This fact means that the segmentation does not delineate correctly the glottal area, causing an erroneous estimation of the trajectory displacements.

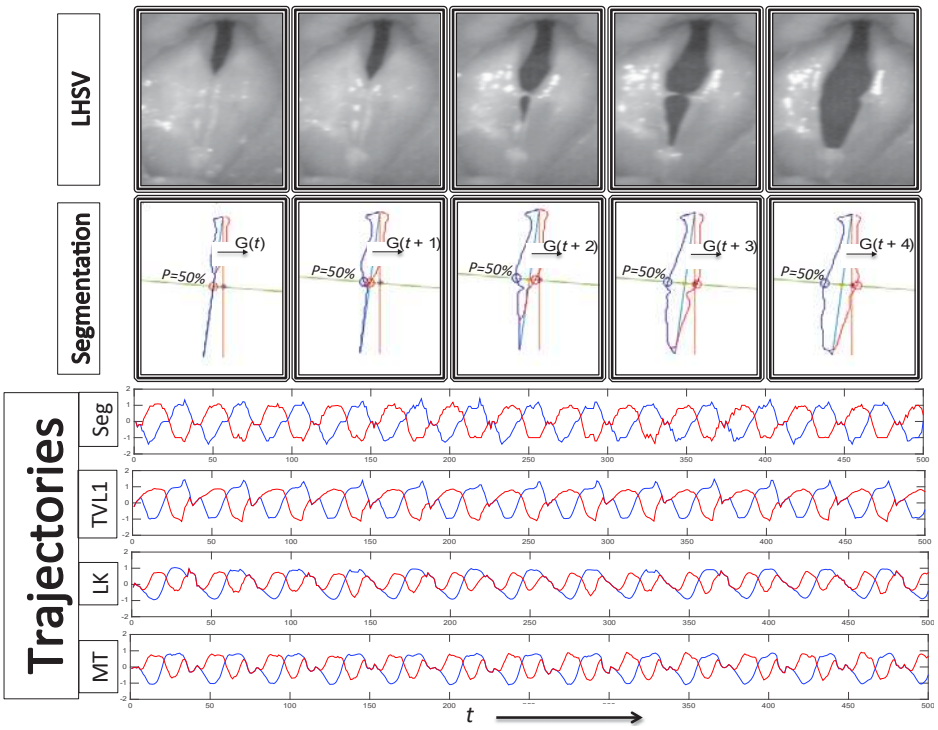


Fig. 13.10. Top: the LHSV sequence cropped after applying a manual ROI selection. Middle: Glottal segmentation (the blue and red contours represent the right and left vocal folds edges respectively; The line to be tracked is colored in green and it is located at *pc* = 50 %). Bottom: trajectories of the vocal folds movement; from up to down: segmentation, TVL1-OF, LK-OF and MT-OF.

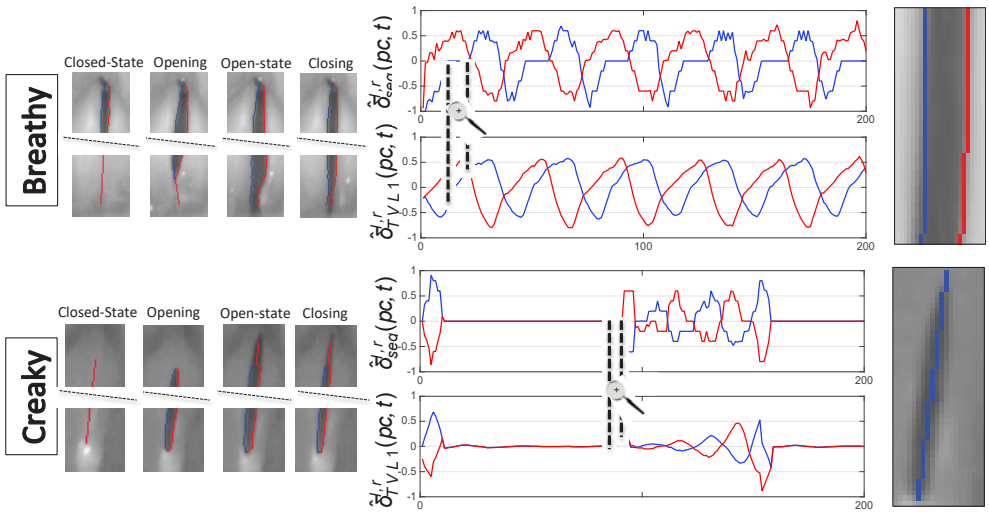


Fig. 13.11. VFDT of two phonatory tasks: breathy and creaky. The left panel shows four frames of each LHSV with their respective segmentation and trajectories. The right panel shows a close up of two frames with segmentation errors corresponding to the interval in dashed lines.

13.4.2. Comparison between OFGVG, GVG and |dxGVG|

Five playbacks are depicted in Fig. 13.12 for three phonation cases: GVG, its derivative |dxGVG|, and three OFGVG. Similarities between |dxGVG| and OFGVG playbacks can be noticed, especially in shape appearance. In pressed phonation there is a long closed-state that can be observed along the five playbacks, taking place at the same time for all of them.

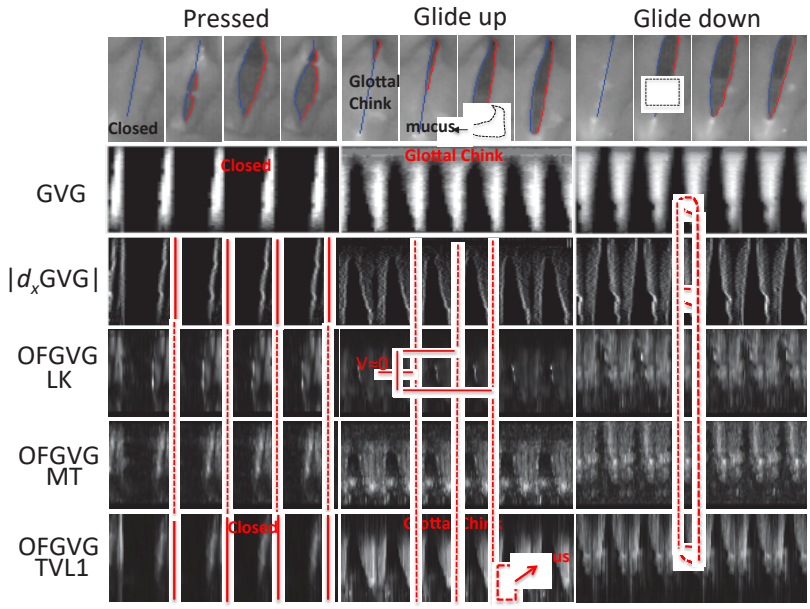


Fig. 13.12. Illustration of GVG, |dxGVG|, OFGVG-LK, OFGVG-MT and OFGVG-TVL1 playbacks for three different phonatory tasks (pressed, glide up and glide down).

Glide up phonation has a posterior glottal chink that produces a constant tonality of gray at the top part of the GVG playback. In contrast, this is perceived as a no-motion region in the |dxGVG| and in the OFGVGs, so it is depicted in black for these playbacks. In the glide-down sequence, the anterior and posterior part of the vocal folds open separately until they close in a short period of time. This effect can be easily observed in the GVG (dashed circle in red) and in its derivative. However, due to the blurring effect induced by the presence of mucus, it is not obviously readable in the OFGVG. Additionally, two peculiarities are observed in the OFGVGs representation of Fig. 13.12. Firstly, the playbacks do not show gray tonalities in the middle part of the glottal cycle (open-state), which means that there is no motion of the vocal folds (velocity close to 0). Secondly, the presence of mucus is depicted as gray regions that produce a blurring effect (bottom panels in Fig. 13.12). Lastly, for all the phonatory tasks a certain degree of noise is found when the OF is computed via LK-OF and MT-OF. Contrariwise, OFGVG based on TVL1-OF is more readable and its shape pattern resembles |dxGVG|.

Figs. 13.13 and 13.14 show two examples where the vibratory patterns are more distinctly represented in the OFGVG-TVL1 than in the GVG. Fig. 13.13 presents an example with a glottal chink in the posterior part, so the motion only appears at the anterior part of the vocal folds. Nevertheless, |dxGVG| indicates a vibratory pattern in the posterior part of the vocal-folds edges due to an imprecise contour detection. Contrariwise, OFGVG synthesizes the motion of the anterior part and includes the vibration of the mucosal wave as blurring gray tonalities during the closed-phase. Fig. 13.14 shows a LHSV sequence with a glottal chink in the posterior part. The glottal-edge contour detected by segmentation does not completely reach the anterior part of the vocal folds, affecting the legibility of the GVG. For instance, a close look to the frame t_{13} and t_{32} shows that there is no left glottal edge defined for the anterior part (red edge). So the distance between the glottal edges is different to zero in spite of the glottis being closed, producing vertical gray lines in the |dxGVG| playback. In contrast, the vibratory pattern of OFGVG is more readable and remains similar for all the glottal cycles. Lastly, its tolerance to highly asymmetrical vocal folds vibration is illustrated in Fig. 13.15 during a glissando with a transition between two laryngeal mechanisms. In this example, OFGVG and |dxGVG| playbacks have features in common such as cycle shape and time of occurrence of mechanism transition.

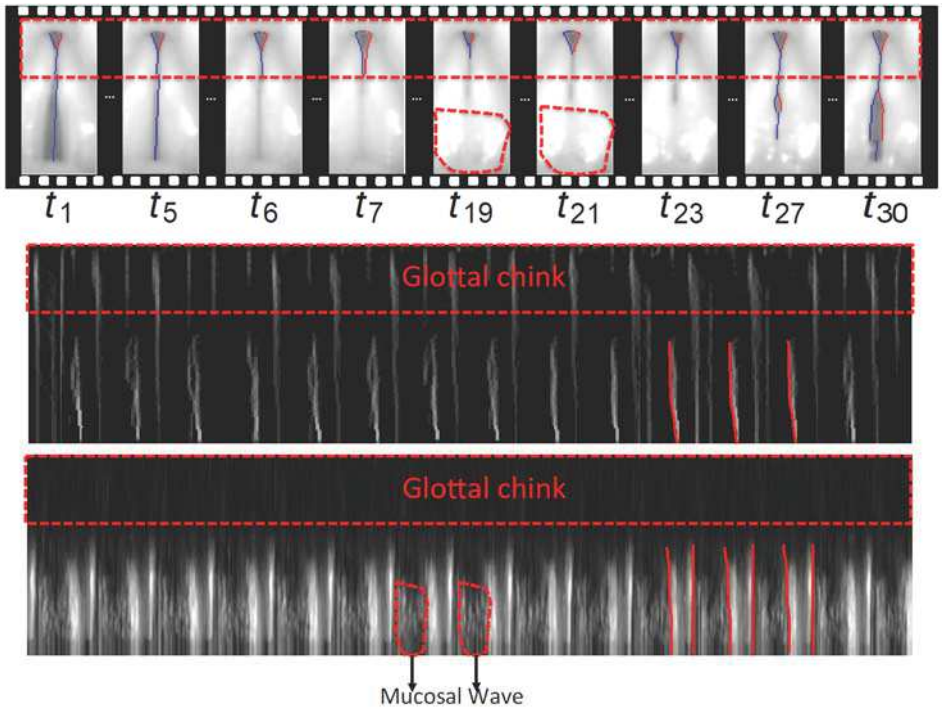


Fig. 13.13. Upper panel: nine segmented frames, the rectangle dotted with red corresponds to the space between the margin of the ROI and to the area with a glottal chink; middle panel: |dxGVG| playback; lower panel: OFGVG with a vertical length that depends on the ROI size. The effect caused by the mucosal wave motion and the vibratory shape patterns for three consecutive cycles are marked with dotted and continuous red lines respectively.

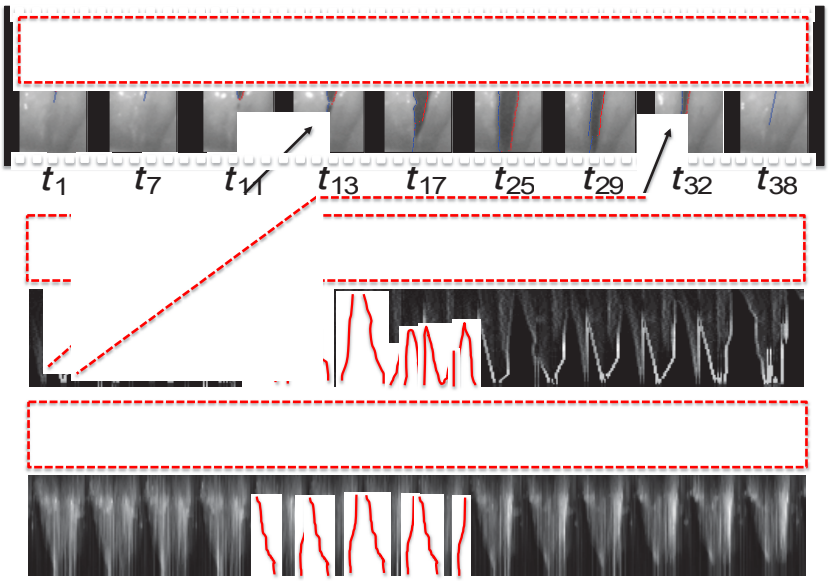


Fig. 13.14. Upper panel: nine segmented frames (the area delimited with red dotted line corresponds to posterior glottal chink); middle panel: |dxGVG| playback; lower panel: OFGVG with a vertical length that depends on the ROI size. The misleading calculation of the distance between edges is observed as gray vertical lines in |dxGVG| plot. The vibratory shape pattern for three consecutive cycles is marked with a continuous red line.

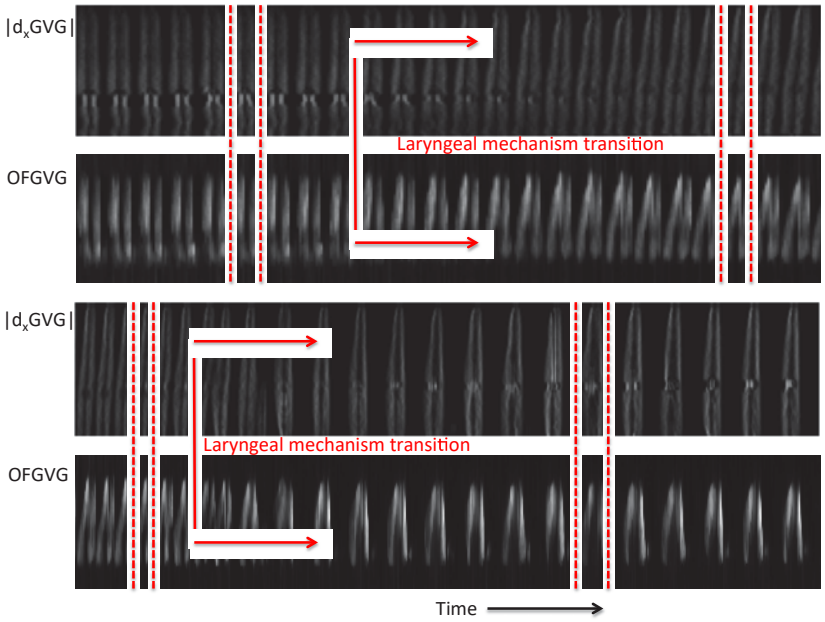


Fig. 13.15. |dxGVG| and OFGVG visualization of peculiar vocal-folds vibratory movements during a glissando with a laryngeal-mechanism transition. Upper panel: 24 glottal cycles. Lower panel: 23 glottal cycles. The laryngeal mechanism transition is pointed out with red arrows and the dashed lines in red indicate different glottal cycles.

13.4.3. Glottal Velocity: Derivative of Glottal Area Waveform and Glottal Optical FlowWaveform

Since GOFW computes an absolute velocity, it is possible to obtain a similar representation by differentiating GAW and computing its absolute value (|dGAW|). The GOFW provides valuable information about the total velocity of the vocal-folds motion for each instant of time. Additionally, if |dGAW| is overlapped with GAW (as shown in Fig. 13.16), it is feasible to analyze the velocity variation with respect to the glottal cycles. Fig. 13.16 shows that in the open-state the velocity decreases, creating a valley in the |dGAW| and in the GOFW playbacks.

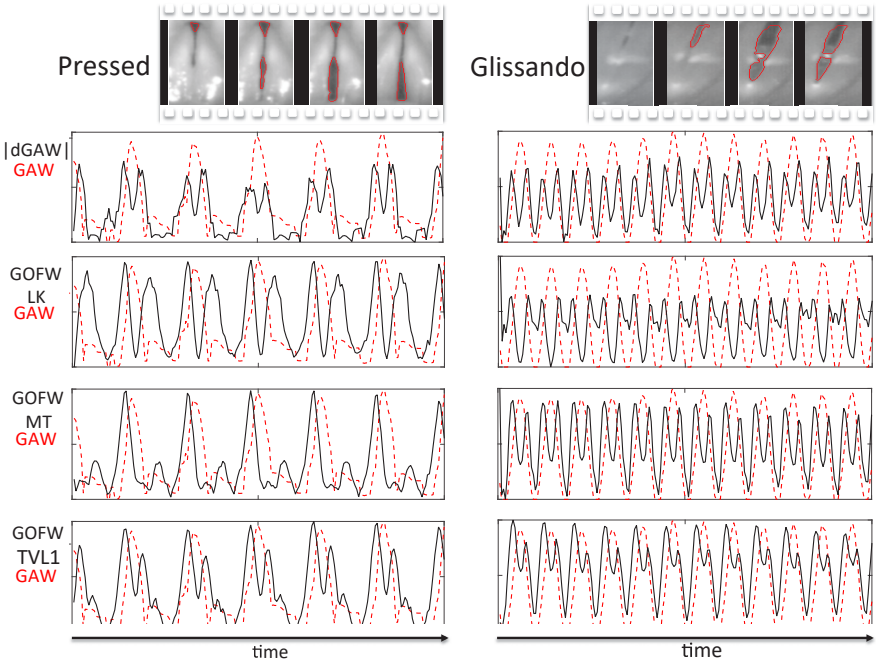


Fig. 13.16. GAW vs GOFW representation for a pressed and glissando task. First row: GAW and |dGAW|; second row: GAW and GOFW-LK; third row: GAW and GOFW-MT; fourth row: GAW and GOFW-TVL1.

Additionally, it shows that the maximum velocities take place in the same instants of time but with different amplitude values depending on the OF techniques. A velocity variation can be seen in all |dGAW| playbacks since in some glottal cycles the maximum occurs during the opening, in others during the closing phase, and sometimes both amplitudes are similar. This fact can be clearly observed in Fig. 13.16 for the pressed voice quality where the amplitude of the peaks oscillates around different values. Contrariwise, GOFW always has its maximum velocity during the opening phase, but the amplitude values are different depending on the OF method used. In the glissando task, |dGAW| and GOFW have a discrepancy with respect to the maximal velocity instant. In |dGAW|, it occurs during the closing-state, while in GOFW, during the opening. Among GOFW playbacks,

the main dissimilarity relies on the peak amplitude. For instance, in the glissando phonation, GOFW-LK and GOFW-MT maximum fluctuates between opening and closing states. In contrast, GOFW-TVL1 always has maximum velocity during the opening phase.

13.4.4. Optical Flow Using Local Dynamics along One Line: Digital Kymogram and Optical Flow Kymogram

OFKG is computed using TVL1-OF, LK-OF and MT-OF for three different glottal locations, each of them corresponding to a percentage of the glottal axis (pc_1 = 10 %, pc_2 = 50 % and pc_3 = 90 %) as shown in Fig. 13.17. The results show that OFKG has a shape similar to DKG, yet blurred over the vocal folds. Such blurring effect is caused by mucosal-wave propagation. One outstanding characteristic appears during the change between opening and closing phases due to the presence of a discontinuity in the OFKG. This can be understood as an instant for which velocity decreases considerably. In pc_1, there is a quasi-static vibratory behavior due to a glottal chink. The DKG represents the absence of motion when the shape of glottal gap (dark region) does not change over time. Meanwhile, OFKG is displayed with low intensity tonalities ($u(pc_1, \mathrm{t})$). The lines located at pc_2 and pc_3 present a visible triangular pattern in OFKG, which is a characteristic of DKG for a normal voice production. LKOF and MT-OF computation approximate the shape expected for OFKG. Yet the images are blurred, and this effect is propagated to the closed-state and to the inner part of the glottis. Contrariwise, OFKG-TVL1 motion pattern is more readable and distinguishable.

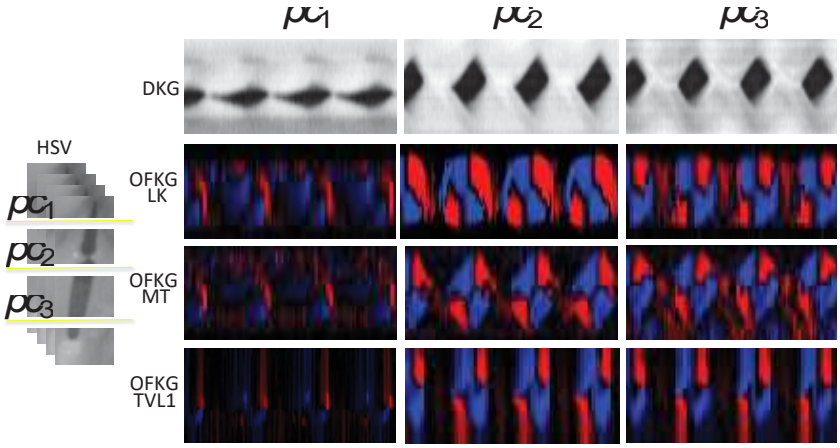


Fig. 13.17. Illustration of DKG and OFKG at three different positions of the LHSV sequence. First row: VKG playback; second row: OFKG using LK-OF; third row: OFKG using MT-OF; fourth row: OFKG using TVL1-OF.

13.5. Conclusions

In this review, the use of OF techniques is explored to synthesize the vocal-folds dynamics. The OF techniques allow to track unidentified objects solely based on its

motion, with no need of additional segmentation techniques. Therefore, not only the points belonging to glottal edges are included but also those regions that originated such movements. Three new playbacks are presented: two of them, called Optical Flow Glottovibrogram (OFGVG) and Glottal Optical Flow Waveform (GOFW), analyze the global dynamics; and the remaining one, called Optical Flow Kymogram (OFKG), analyzes the local dynamics. These new ways for data visualization have the goal to overcome the drawbacks of existing playbacks, providing simultaneously features that integrate the time dynamics, such as velocity, acceleration, instants of maximum and minimum velocity, vocal-folds displacements during phonation and motion analysis. The proposed OF-based playbacks demonstrate a great correlation in shape with traditional playbacks, allowing to identify important instants of glottal cycle, such as closed-state and maximal opening. In addition, the playbacks based on OF computation provide complementary information to the common spatiotemporal representations when segmentation is not available, or when it is not reliable enough due to failures in glottal-edge detection. In comparison to traditional playbacks, OF-based ones are slightly blurred due to the effect introduced by the mucosal-wave movements.

OF-based LHSV analysis is a promising approach that would require further development, such as to find out additional ways to reduce the dimensionality of vector motion field and include the information of y-axis, to consider the deflections of left and right folds separately, to compare the accuracy of the OF algorithm with respect to reality, and not only by comparison with conventional segmentation techniques. Functional voice disorders could be classified using OF playbacks. Individual contribution of mucosal wave during phonation could be studied. It might be helpful to combine OF-based techniques with other glottal-activity signals such as electroglottography (EGG) in order to provide a more complete assessment of vocal-folds vibration.

Acknowledgements

We are very grateful to the ENT team of University Medical Center Hamburg-Eppendorf: Frank Müller, Dr. Götz Schade and Pr. Markus Hess for the database collection, to Cédric Gendrot and Robert Expert who volunteered as subjects, and to Sevasti-Zoi Karakozoglou for manual glottal segmentation of the selected sequences.

References

[1]. J. Kreiman, D. Sidtis, Introduction, in Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception, *Wiley*, 2011, pp. 1-24.

[2]. S. R. Schwartz, S. M. Cohen, S. H. Dailey, R. M. Rosenfeld, E. S. Deutsch, M. B. Gillespie, et al., Clinical practice guideline: Hoarness (dysphonia), *Otolaryngology-Head and Neck-Surgery*, Vol. 141, 2009, pp. S1-S31.

[3]. C. Bohr, A. Kräck, A. Dubrovskiy, U. Eysholdt, J. Švec, G. Psychogios, A. Ziethe, M. Döllinger, Spatiotemporal analysis of high-speed videolaryngoscopic imaging of organic pathologies in males, *Journal of Speech, Language, and Hearing Research*, Vol. 57, Issue 4, 2014, pp. 1148-1161.

[4]. T. Braunschweig, J. Flaschka, P. Schelhorn-Neise, M. Döllinger, High-speed video analysis of the phonation onset, with an application to the diagnosis of functional dysphonias, *Medical Engineering and Physics*, Vol. 30, Issue 1, 2008, pp. 59-66.
[5]. C. Tao, Y. Zhang, J. J. Jiang, Extracting physiologically relevant parameters of vocal folds from high-speed video image series, *IEEE Transactions on Biomedical Engineering*, Vol. 54, Issue 5, 2007, pp. 794-801.
[6]. D. Deliyski, E. H. Robert, State of the art in laryngeal imaging: research and clinical implications, *Current Opinion in Otolaryngology & Head and Neck Surgery*, Vol. 18, Issue 3, 2010, pp. 147-152.
[7]. D. D. Mehta, R. E. Hillman, The evolution of methods for imaging vocal fold phonatory function, *SIG 5 Perspectives on Speech Science and Orofacial Disorders*, Vol. 22, Issue 1, 2012, pp. 5-13.
[8]. P. Moore, H. Von Leden, Dynamic variations of the vibratory pattern in the normal larynx, *Folia Phoniatrica et Logopaedica*, Vol. 10, Issue 4, 1958, pp. 205-238.
[9]. R. Timcke, H. Von Leden, P. Moore, Laryngeal vibrations: Measurements of the glottic wave. I. The normal vibratory cycle, *Archives of Otolaryngology*, Vol. 68, Issue 1, 1958, pp. 1-19.
[10]. H. Von Leden, P. Moore, R. Timcke, Laryngeal vibrations: Measurements of the glottic wave, Part III. The Pathologic Larynx, *Archives of Otolaryngology*, Vol. 71, Issue 4, 1960, pp. 16-35.
[11]. D. G. Childers, A. Paige, P. Moore, Laryngeal vibration patterns machine-aided measurements from high-speed film, *Archives of Otolaryngology*, Vol. 102, Issue 7, 1976, pp. 407-410.
[12]. J. G. Švec, H. K. Schutte, Videokymography: High-speed line scanning of vocal fold vibration, *Journal of Voice*, Vol. 10, Issue 2, 1996, pp. 201-205.
[13]. T. Wurzbacher, R. Schwarz, M. Döllinger, U. Hoppe, U. Eysholdt, J. Lohscheller, Model-based classification of nonstationary vocal fold vibrations, *The Journal of the Acoustical Society of America*, Vol. 120, Issue 2, 2006, pp. 1012-1027.
[14]. J. Lohscheller, U. Eysholdt, Phonovibrogram visualization of entire vocal fold dynamics, *The Laryngoscope*, Vol. 118, Issue 4, 2008, pp. 753-758.
[15]. D. D. Mehta, D. D. Deliyski, T. F. Quatieri, R. E. Hillman, Automated measurement of vocal fold vibratory asymmetry from high-speed videoendoscopy recordings, *Journal of Speech, Language, and Hearing Research*, Vol. 54, Issue 1, 2013, pp. 47-54.
[16]. J. Lohscheller, J. G. Švec, M. Döllinger, Vocal fold vibration amplitude, open quotient, speed quotient and their variability along glottal length: kymographic data from normal subjects, *Logopedics Phoniatrics Vocology*, Vol. 38, Issue 4, 2013, pp. 182-192.
[17]. C. T. Herbst, J. Lohscheller, J. G. Švec, N. Henrich, G. Weissengruber, W. T. Fitch, Glottal opening and closing events investigated by electroglottography and super-high-speed video recordings, *The Journal of Experimental Biology*, Vol. 217, Issue 6, 2014, pp. 955-963.
[18]. Y. Yan, E. Damrose, D. Bless, Functional analysis of voice using simultaneous high-speed imaging and acoustic recordings, *Journal of Voice*, Vol. 21, Issue 5, 2007, pp. 604-616.
[19]. K. Ahmad, Y. Yan, D. Bless, Vocal fold vibratory characteristics in normal female speakers from high-speed digital imaging, *Journal of Voice*, Vol. 26, Issue 2, 2012, pp. 239-253.
[20]. M. Döllinger, U. Hoppe, F. Hettlich, J. Lohscheller, S. Schuberth, U. Eysholdt, Vibration parameter extraction from endoscopic image series of the vocal folds, *IEEE Transactions on Biomedical Engineering*, Vol. 49, Issue 8, 2002, pp. 773-781.
[21]. K. Ishizaka, J. L. Flanagan, Synthesis of voiced sounds from a two mass model of the vocal cords, *Bell Labs Technical Journal*, Vol. 51, Issue 6, 1972, pp. 1233-1268.
[22]. R. Schwarz, M. Döllinger, T. Wurzbacher, U. Eysholdt, J. Lohscheller, Spatio-temporal quantification of vocal fold vibrations using high-speed videoendoscopy and a biomechanical model, *The Journal of the Acoustical Society of America*, Vol. 123, Issue 5, 2008, pp. 2717-2732.

[23]. D. Wong, M. Ito, N. Cox, I. R. Titze, Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases, *The Journal of the Acoustical Society of America*, Vol. 89, Issue 1, 1991, pp. 383-394.
[24]. A. P. Pinheiro, D. E. Stewart, C. D. Maciel, J. C. Pereira, S. Oliveira, Analysis of nonlinear dynamics of vocal folds using high-speed video observation and biomechanical modeling, *Digital Signal Processing*, Vol. 22, Issue 2, 2012, pp. 304-313.
[25]. H. S. Shaw, D. D. Deliyski, Mucosal wave: A normophonic study across visualization techniques, *Journal of Voice*, Vol. 22, Issue 1, 2008, pp. 23-33.
[26]. D. Voigt, M. Döllinger, U. Eysholdt, A. Yang, E. Gürlen, J. Lohscheller, Objective detection and quantification of mucosal wave propagation, *The Journal of the Acoustical Society of America*, Vol. 128, Issue 5, 2010, pp. EL347-EL353.
[27]. C. R. Krausert, A. E. Olszewski, L. N. Taylor, J. S. Mcmurray, S. H. Dailey, J. J. Jiang, Mucosal wave measurement and visualization techniques, *Journal of Voice*, Vol. 25, Issue 4, 2011, pp. 395-405.
[28]. P. A. Lindestad, M. Södersten, B. Merker, S. Granqvist, Voice source characteristics in Mongolian "Throat Singing" studied with high-speed imaging technique, acoustic spectra, and inverse filtering, *Journal of Voice*, Vol. 15, Issue 1, 2001, pp. 78-85.
[29]. L. Bailly, N. Henrich, X. Pelorson, Vocal fold and ventricular fold vibration in period-doubling phonation: Physiological description and aerodynamic modeling, *The Journal of the Acoustical Society of America*, Vol. 127, Issue 5, 2010, pp. 3212-3222.
[30]. D. Z. Borch, J. Sundberg, P. A. Lindestad, M. Thalén, Vocal fold vibration and voice source aperiodicity in 'dist' tones: a study of a timbral ornament in rock singing, *Logopedics Phoniatrics Vocology*, Vol. 29, Issue 4, 2004, pp. 147-153.
[31]. L. Bailly, N. Henrich Bernardoni, F. Müller, A.-K. Rohlfs, M. Hess, The ventricular-fold dynamics in human phonation, *Journal of Speech, Language, and Hearing Research*, Vol. 57, 2014, pp. 1219-1242.
[32]. D. Voigt, M. Döllinger, T. Braunschweig, A. Yang, U. Eysholdt, J. Lohscheller, Classification of functional voice disorders based on phonovibrograms, *Artificial Intelligence in Medicine*, Vol. 49, Issue 1, 2010, pp. 51-59.
[33]. J. Unger, J. Lohscheller, M. Reiter, K. Eder, C. S. Betz, M. Schuster, A noninvasive procedure for early-stage discrimination of malignant and precancerous vocal fold lesions based on laryngeal dynamics analysis, *Cancer Research*, Vol. 75, Issue 1, 2015, pp. 31-39.
[34]. D. Deliyski, P. P. Petrushev, H. Bonilha, T. T. Gerlach, B. Martinharris, R. E. Hillman, Clinical implementation of laryngeal highspeed videoendoscopy: challenges and evolution, *Folia Phoniatrica et Logopaedica*, Vol. 60, Issue 1, 2008, pp. 33-44.
[35]. L. Westphal, D. Childers, Representation of glottal shape data for signal processing, *IEEE Transactions on Acoustics, Speech, Signal Processing*, Vol. 31, Issue 3, 1983, pp. 766-769.
[36]. C. Palm, T. Lehmann, J. Bredno, C. Neuschaefer-Rube, S. Klajman, K. Spitzer, Automated analysis of stroboscopic image sequences by vibration profile diagrams, in *Proceedings of the 5th International Conference Advances in Quantitative Laryngology, Voice and Speech Research*, Groningen, Germany, 2001.
[37]. J. Neubauer, P. Mergell, U. Eysholdt, H. Herzel, Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial modes, *The Journal of the Acoustical Society of America*, Vol. 110, Issue 6, 2001, pp. 3179-3192.
[38]. L. Li, N. P. Galatsanos, D. Bless, Eigenfolds: A new approach for analysis of vibrating vocal folds, in *Proceedings of the 3rd International Symposium on Biomedical Imaging (ISBI'02),* Washington, DC, USA, 2002, pp. 589-592.
[39]. Y. Zhang, J. J. Jiang, C. Tao, E. Bieging, J. K. Maccallum, Quantifying the complexity of excised larynx vibrations from high-speed imaging using spatiotemporal and nonlinear dynamic analyses, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, Vol. 17, Issue 4, 2007, pp. 1-10.

[40]. J. Lohscheller, H. Toy, F. Rosanowski, U. Eysholdt, M. Dollinger, Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos, *Medical Image Analysis*, Vol. 11, Issue 4, 2007, pp. 400-413.
[41]. K.-I. Sakakibara, H. Imagawa, M. Kimura, H. Yokonishi, N. Tayama, Modal analysis of vocal fold vibrations using laryngotopography, in *Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH'10)*, Makuhari, Japan, 2010, pp. 917-920.
[42]. S. Z. Karakozoglou, N. Henrich, C. D'alessandro, Y. Stylianou, Automatic glottal segmentation using local-based active contours and application to glottovibrography, *Speech Communication*, Vol. 54, Issue 5, 2012, pp. 641-654.
[43]. J. Unger, T. Meyer, C. T. Herbst, W. T. S. Fitch, M. Döllinger, J. Lohscheller, Phonovibrographic wavegrams: Visualizing vocal fold kinematics, *The Journal of the Acoustical Society of America*, Vol. 133, Issue 2, 2013, pp. 1055-1064.
[44]. T. Ikuma, M. Kunduk, A. J. Mcwhorter, Advanced waveform decomposition for high-speed videoendoscopy analysis, *Journal of Voice*, Vol. 27, Issue 3, 2013, pp. 369-375.
[45]. A. I. A. Rahman, S.-H. Salleh, K. Ahmad, K. Anuar, Analysis of vocal fold vibrations from high-speed digital images based on dynamic time warping, *International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering*, Vol. 8, Issue 6, 2014, pp. 306-309.
[46]. G. Chen, J. Kreiman, A. Alwan, The glottaltopogram: A method of analyzing high-speed images of the vocal folds, *Computer Speech and Language*, Vol. 28, Issue 5, 2014, pp. 1156-1169.
[47]. C. T. Herbst, J. Unger, H. Herzel, J. G. Švec, J. Lohscheller, Phasegram analysis of vocal fold vibration documented with laryngeal highspeed video endoscopy, *Journal of Voice*, Vol. 30, Issue 6, 2016, pp. 771.e1-771.e15.
[48]. B. K. Horn, B. Schunck, Determining optical flow: A retrospective, *Artificial Intelligence*, Vol. 17, 1981, pp. 185-203.
[49]. B. D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI'81)*, Vol. 2, Vancouver, BC, Canada, 1981, pp. 674-679.
[50]. S. S. Beauchemin, J. L. Barron, The computation of optical flow, *ACM Computing Surveys*, Vol. 27, Issue 3, 1995, pp. 433-466.
[51]. B. Glocker, N. Komodakis, N. Paragios, G. Tziritas, N. Navab, Inter and intra-modal deformable registration: Continuous deformations meet efficient optimal linear programming, in *Proceedings of the 20th International Conference on Information Processing in Medical Imaging (IPMI'07)*, 2007, pp. 408-420.
[52]. N. Hata, A. Nabavi, W. M. I. Wells, S. K. Warfield, R. Kikinis, P. M. Black, et al., Three-dimensional optical flow method for measurement of volumetric brain deformation from intraoperative MR images, *Journal of Computer Assisted Tomography*, Vol. 24, Issue 4, 2000, pp. 531-538.
[53]. T. C. Huang, C. K. Chang, C. H. Liao, Y. J. Ho, Quantification of blood flow in internal cerebral artery by optical flow method on digital subtraction angiography in comparison with time-of-flight magnetic resonance angiography, *PLoS ONE*, Vol. 8, Issue 1, 2013, e54678.
[54]. I. H. Kim, Y. C. Chen, D. Spector, R. Eils, K. Rohr, Nonrigid registration of 2-D and 3-D dynamic cell nuclei images for improved classification of subcellular particle motion, *IEEE Transactions on Image Processing*, Vol. 20, Issue 4, 2011, pp. 1011-1022.
[55]. A. S. Hassanein, A. M. Khalifa, Al-W. Atabany, M. T. El-Wakad, Performance of optical flow tracking approaches for cardiac motion analysis, in *Proceedings of the Middle East Conference on Biomedical Engineering (MECBME'14)*, 2014, pp. 4-7.

[56]. Q. Luo, D. C. Liu, Optical flow computation based medical motion estimation of cardiac ultrasound imaging, in *Proceedings of the 5th International Conference on Bioinformatics and Biomedical Engineering (iCBBE'11)*, Vol. 2, 2011, pp. 1-4.
[57]. F. M. M. Shuib, M. Othman, K. Abdulrahim, Z. Zulkifli, Analysis of motion detection of breast tumor based on tissue elasticity from B mode ultrasound images using gradient method optical flow algorithm, in *Proceedings of the 1st International Conference on Artificial Intelligence, Modelling and Simulation (AIMS'13)*, 2013, pp. 278-283.
[58]. J. Liu, K. R. Subramanian, T. S. Yoo, An optical flow approach to tracking colonoscopy video, *Computerized Medical Imaging and Graphics*, Vol. 37, Issue 3, 2013, pp. 207-223.
[59]. J. L. Barron, D. J. Fleet, S. S. Beauchemin, Performance of optical flow techniques, *International Journal of Computer Vision*, Vol. 1, Issue 12, 1994, pp. 43-77.
[60]. A. Mitiche, J. Aggarwal, Computer Vision Analysis of Image Motion by Variational Methods, Springer Topics in Signal Processing, Vol. 10, *Springer*, 2014.
[61]. J. Weickert, A. Bruhn, T. Brox, N. Papenberg, A Survey on variational optic flow methods for small displacements, in Mathematical Models for Registration and Applications to Medical Imaging (O. Scherzer, Ed.), Mathematics in Industry, Vol. 10, *Springer*, Berlin, Heidelberg, 2006, pp. 103-136.
[62]. M. Werlberger, T. Pock, H. Bischof, Motion estimation with non local total variation regularization, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'10)*, San Francisco, USA, 2010, pp. 2464-2471.
[63]. A. Wedel, D. Cremers, Optical flow estimation, in Stereo Scene Flow for 3D Motion Analysis, *Springer*, London, 2011, pp. 5-34.
[64]. M. Menze, C. Heipke, A. Geiger, Discrete optimization for optical flow, in *Proceedings of the German Conference on Pattern Recognition (GCPR'15)*, 2015, pp. 16-28.
[65]. D. Fortun, P. Bouthemy, C. Kervrann, Optical flow modeling and computation: A survey, *Computer Vision and Image Understanding*, Vol. 134, 2015, pp. 1-21.
[66]. C. Zach, T. Pock, H. Bischof, A duality based approach for realtime TV-L1 optical flow, in *Proceedings of the 29th Conference on Pattern Recognition (DAGM'07)*, Heidelberg, Germany, 2007, pp. 214-223.
[67]. M. Drulea, S. Nedevschi, Motion estimation using the correlation transform, *IEEE Transactions on Image Processing*, Vol. 22, Issue 8, 2013, pp. 3260-3270.
[68]. S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, R. Szeliski, A database and evaluation methodology for optical flow, *International Journal of Computer Vision*, Vol. 92, Issue 1, 2011, pp. 1-31.
[69]. G. Andrade-Miranda, N. Henrich Bernardoni, J. I. Godino-Llorente, Synthesizing the motion of the vocal folds using optical flow based techniques, *Biomedical Signal Processing and Control*, Vol. 34, 2017, pp. 25-35.
[70]. A Granados, J. Brunskog, An optical flow-based state-space model of the vocal folds, *The Journal of the Acoustical Society of America*, Vol. 141, Issue 6, 2017, pp. EL543-EL548.
[71]. A. K. Saadah, N. P. Galatsanos, D. Bless, A. Ramos, Deformation Analysis of the Vibrational Patterns of the Vocal Folds, *Bildverarbeitung für die Medizin*, 1996.
[72]. G. Farneback, Fast and accurate motion estimation using orientation tensors and parametric motion models, in *Proceedings of the 15th International Conference on Pattern Recognition (ICPR'2000)*, Vol. 1, Barcelona, Spain, 2000, pp. 135-139.