

Quantitative Analysis of Healthy and Pathological Vocal Fold Vibrations using an Optical Flow based Waveform

Heyfa Ammar

King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia
University of Tunis El Manar, Tunisia

Abstract—The objective assessment of the vocal fold vibrations is important in diagnosing several vocal diseases. Given the high speed of the vibrations, the high speed videoendoscopy is commonly used to capture the vocal fold movements into video recordings. Commonly, two steps are carried out in order to automatically quantify laryngeal parameters and assess the vibrations. The first step aims to map the spatial-temporal information contained in the video recordings into a representation that facilitates the analysis of the vibrations. Numerous techniques are reported in the literature but the majority of them require the segmentation of all the images of the video, which is a complex task. The second step aims to quantify laryngeal parameters in order to assess the vibrations. To this aim, most of the existing approaches require an additional processing to the representation in order to deduce those parameters. Furthermore, for some reported representations, the assessment of the symmetry and the periodicity of the vocal fold dynamics needs setting up parameters that are specific to the representation under consideration; which makes difficult the comparison between the existing techniques. To alleviate these problems, the present study investigates the use of a recently proposed representation named optical flow based waveform, in order to objectively quantify the laryngeal parameters. This waveform is retained in this study as it does not require the segmentation of all the images of the video. Furthermore, it will be shown in the present work that the automatic quantification of the vibrations using this waveform can be carried out without applying any additional processing. Moreover, common laryngeal parameters are exploited; hence, no specific parameters are needed to be defined for the automatic assessment of the vibrations. Experiments conducted on healthy and pathological phonation show the accuracy of the waveform. Besides, it is more sensitive to pathological phonation than the state-of-the-art techniques.

Keywords—Quantification; vocal fold vibrations; optical flow based waveform; pathology

I. INTRODUCTION

Diseases that affect our ability to speak can have disastrous impacts on our lives as our voice remains the most important way to communicate. It is then important to diagnose and treat vocal disorders as soon as signs like hoarseness, difficulty of speaking and many others, take place. To make a sound, the right and left vocal folds (VF) consisting of two membranes located in the larynx, open and close periodically, and their vibrations are symmetric. For the aim of diagnosing their functional behavior, the examination of these high-frequency vibrations is made possible thanks to the analysis of high speed videoendoscopy (HSV) recordings. HSV allows to capture the VF movements at a frequency that could exceed 4000

images/second. However, this high frequency makes it hard to visually assess the vibrations. For this reason, in a first step, the spatial-temporal information contained in the HSV recordings are commonly mapped into representations appropriate for the human visualization and analysis. Then, in the second step, the VF vibrations are assessed either visually based on the shape of the considered representation, or automatically through the quantification of some vibratory parameters. Many efforts are invested by researchers to propose objective techniques for the assessment of the VF vibrations. However, they suffer from one or more of the following limitations:

- 1) In order to be generated, the representation used to reflect the VF vibrations requires the glottal segmentation in all the images of a HSV recording. This is a hard and a complex task as the segmentation of the VF necessitates the user intervention and the number of images in a video is large given the high sampling rate. Furthermore, the segmentation may fail when the VF are completely closed. Examples of such representations are the glottal area waveform (GAW) [1], the digital kymography (DKG) [2], the phonovibrography (PVG) [3], the glottovibrography (GVG) [4] and more [5], [6].
- 2) The objective assessment of the vibrations through a representation, mostly requires a further processing of it in order to be able to derive the laryngeal parameters. For instance, the DKG should be segmented in order to derive parameters like vibration amplitudes, periodicity indices and so on [7]. The same holds for the PVG [3], [8] and the GVG [4].
- 3) The automatic assessment of the vibrations is sometimes based on parameters that are specific to the representation under consideration. Researchers agree on the features to be assessed namely the periodicity and the symmetry of the vibrations; but, different parameters aiming to quantify these features can be found in the literature. Consequently, it becomes difficult to compare between the representations in terms of efficiency and accuracy of the vibration assessment. For instance, parameters specific to PVG are proposed in [3], [8]. A comparative study conducted in [9] investigates the efficiency of parameters derived from different methods like GAW, DKG and the laryngotopography [10].

To alleviate the limitations mentioned above, the present work investigates the use of a recently proposed optical flow based

waveform (OFW) [11] in order to quantify the vibratory characteristics. The exploitation of this waveform is retained in this work for mainly three reasons. First, to generate the OFW, the segmentation of only one image per vibratory cycle is necessary. This allows a remarkably decrease in the amount of images to be segmented. Furthermore, the segmented image corresponds to the one of maximal opening of the glottis in each vibratory cycle; and hence, results of high accuracy could be obtained. Second, the OFW can be directly exploited for the automatic assessment of the vibrations without any further processing. Third, the quantification of the vibratory characteristics can be easily and accurately computed through the OFW based on the most commonly used parameters. The most important and common parameters in the literature are [7], [12], [13]: the fundamental frequency, the left-right phase symmetry, the left-right amplitude symmetry, the time periodicity and the amplitude periodicity. The accuracy of the objective analysis of the OFW is evaluated in comparison to the measures obtained by the analysis of the electroglottographic (EGG) signals [14] and DKG. Experiments show that by using the same parameters, the OFW is more sensitive to a pathological phonation than DKG.

The remainder of this paper is organized as follows. Section II describes the quantification technique and the data used for the evaluation of the obtained measures. The results are depicted in Section III and discussed in Section IV. Finally, some conclusions are drawn in Section V.

II. MATERIALS AND METHODS

A. Materials

Video recordings corresponding to healthy and pathological vibrations of the VF are analyzed in order to evaluate the accuracy of the quantification measures. HSV recordings of healthy phonation are provided by E. Bianco and G. Degottex-IRCAM [15], [16] and contains about 48 videos along with the corresponding EGG and audio signals. Videos corresponding to disordered VF of different pathologies are publicly available online¹.

B. Brief Overview of the OFW Technique

The rOFW (resp. the lOFW) carried out at a level \mathcal{L} from the posterior to the anterior commissure is the trajectory of a point located on the right VF (resp. left VF) at the glottal level \mathcal{L} , during phonation. Typical values of \mathcal{L} are 25%, 50% and 75%. Fig. 1 is an illustration of the OFWs that correspond to healthy vocal folds at the three glottal levels. The amplitudes of the lOFW are displayed in the negative part for more clarity.

To generate the waveforms, the OFW technique proceeds according to the following steps. After sampling the HSV video recording into K images $\{I^{(k)}\}_{k=1,\dots,K}$, the region that includes the glottis and the VF is firstly detected according to the technique described in [4]. All the processing is carried out on this region of interest in order to alleviate the computations. The second step aims to partition the set of K images into \mathcal{C} sets of images $\{\{I_c^{(k')}\}_{k'=1,\dots,K}\}_{c=1,\dots,\mathcal{C}}$ to allow

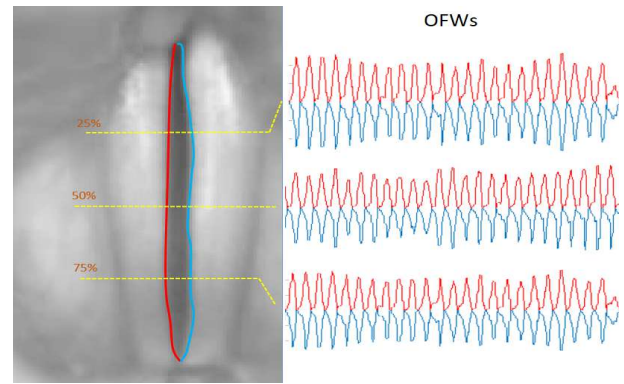


Fig. 1. Optical flow based waveforms at three levels of the glottal area. rOFW in red. -lOFW (in blue) is displayed instead of lOFW for more clarity.

a cycle-to-cycle interpretation of the VF vibratory behavior. Each set $\{I_c^{(k')}\}_{k'}$ consists of K_c images composing the vibratory cycle $c = 1, \dots, \mathcal{C}$ among the \mathcal{C} cycles of the phonation under consideration. K_c is the number of images of the cycle c and k' is a number that defines the temporal location of the image in the cycle. Hence, in a healthy phonation, the set $\{I_c^{(k')}\}_{k'}$ starts by an image of a closed status of the VF which gradually open until reaching the image I_c^{ref} of maximal opening. Then, the VF gradually close during a closing stage until the end of the cycle.

The image I_c^{ref} considered as the *reference image* of the cycle c , is of a particular importance. It allows an accurate detection of the glottal edges and consequently, an accurate localization of the points to be tracked, located on the right and the left vocal folds at the chosen level \mathcal{L} of the glottis. The optical flow (OF) estimation is carried out in the backward and the forward directions from the reference image within the same cycle, in order to better reflect the vibratory behavior of the VF. The displacements of the points of interest along the cycle are cumulated with respect to their positions in I_c^{ref} and constitute the OFW. Further details are given in [11] in which the visual interpretation of the OFW either related to a healthy or a pathological phonation showed its accuracy in assessing the important features of the vibrations such as the periodicity and the symmetry between the two vocal fold dynamics. Despite the efficiency of the visual analysis of the waveforms, an objective and automatic assessment is necessary [17]. In the present study, we investigate the objective and automatic analysis of the OFW by quantifying vibratory parameters as described in the following paragraph.

C. Objective Assessment of the VF Vibrations based on OFW

The diversity of the reported representations aiming to facilitate the visual analysis of the VF vibrations (like the glottal area waveform [1], the phonovibrography [3], the glottovibrography [4], the DKG [2]) implied the appearance of numerous quantification measures. Although most of these measures are closely related to their respective representations, their common objective is to assess important features such as the fundamental frequency, the amplitude/phase symmetry and the periodicity of the vibrations. Additionally to these features,

¹at www.entusa.com

we propose to evaluate the amplitude similarities between the two VF vibrations at each instant.

a) Amplitude Similarity: It is important to know how similar are the amplitudes of the two waveforms over time. For this reason, the one-way ANOVA test is carried out in order to evaluate the similarity between the right and the left vibration amplitudes at all instants.

b) Fundamental frequency: The fundamental frequency F_{0R} (resp. F_{0L}) of the trajectory $rOFW$ (resp. $lOFW$) of the right (resp. the left) vocal fold is estimated using the non linear curve fit model defined by [18]:

$$rOFW(k) = s_0 + a_1 \sin\left(\frac{2\pi}{T_R}k\right) + b_1 \cos\left(\frac{2\pi}{T_R}k\right) + a_2 \sin\left(\frac{4\pi}{T_R}k\right) + b_2 \cos\left(\frac{4\pi}{T_R}k\right). \quad (1)$$

where s_0 is the direct component related to the average value of $rOFW$ over time, a_1, b_1, a_2 and b_2 are coefficients. Hence, the vibration period T_R is estimated and the fundamental frequency F_{0R} related to the trajectory of the right vocal fold is deduced as $F_{0R} = \frac{1}{T_R}$. Analogously, the vibration period T_L related to the trajectory of the left vocal fold is estimated and its fundamental frequency $F_{0L} = \frac{1}{T_L}$ is computed.

c) Symmetry: A large left-right asymmetry can cause voice problems especially when frequency differences between the right and the left vocal folds are significant. This behavior appears when the patient suffers from an unilateral laryngeal paralysis. The symmetrical aspect of the vibrations can be viewed in several ways: amplitude, phase and frequency differences [19]. They can be assessed using the following quantification measures [7].

- **Left-right amplitude symmetry index (ASI):** This indicates the degree of similarity between the amplitudes of the two vocal folds when they are at their maximum value (which corresponds to the maximum closing in the waveform) within a given cycle. It is defined by the difference between the maximum right displacement a_R and the maximum left displacement a_L divided by the sum of them as shown in Fig. 2:

$$ASI = \frac{a_R - a_L}{a_R + a_L} \quad (2)$$

A value of ASI that approaches 0 indicates a perfect symmetry in amplitude between the two vocal folds at the selected level of the glottis.

- **Left-right phase symmetry index (PSI):** When the two vocal folds reach their maximal opening at the same time, their vibrations can be qualified by phase-symmetric. The PSI is defined as the difference between the instants t_R and t_L when respectively the right and the left vocal folds reach their maximal opening, divided by the mean vibration period [7]:

$$PSI = \frac{t_R - t_L}{T_R}. \quad (3)$$

A value of PSI that approaches 0 indicates a perfect symmetry in phase between the vocal folds.

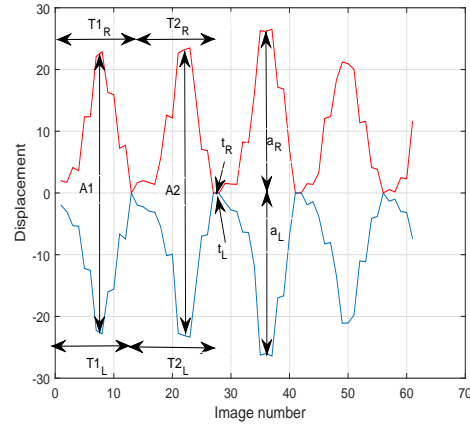


Fig. 2. The waveform parameters. $rOFW$ in red, $(-lOFW)$ in blue. a_R and a_L : maximum displacements within one cycle of the right and left VF. t_R and t_L : instants when the right and the left VF reach their maximal opening. T_{1R}, T_{2R}, T_{1L} and T_{2L} : duration periods of two cycles related to the right and the left vibrations.

d) Periodicity: The periodicity can be defined as the repetition of the same spatial-temporal vibratory behavior along many cycles. It is evaluated by assessing the time and amplitude periodicity.

- **Time periodicity index (TPI):** This is the ratio between the shorter duration of a cycle and the larger duration in two successive cycles [7]. We analyze the time periodicity in the right and the left vocal fold waveforms and determine the respective time periodicity indices TPI_R and TPI_L as follows:

$$TPI_R = \frac{\min(T_{1R}, T_{2R})}{\max(T_{1R}, T_{2R})}, \quad (4)$$

$$TPI_L = \frac{\min(T_{1L}, T_{2L})}{\max(T_{1L}, T_{2L})}.$$

where T_{1R} and T_{2R} (resp. T_{1L} and T_{2L}) the durations of two successive cycles in the right (resp. left) vocal fold waveform. The values range between 0 and 1. A vibration is perfectly periodic when the corresponding time periodicity index approaches 1.

- **Amplitude periodicity index (API):** This is the ratio between the smaller amplitude and the larger one in two consecutive cycles. It is defined by:

$$API = \frac{\min(A_1, A_2)}{\max(A_1, A_2)} \quad (5)$$

where A_1 and A_2 are the sum of the right and the left amplitudes when the vocal folds reach their maximal closing respectively calculated in two successive cycles as illustrated in Fig. 2.

III. RESULTS

In order to evaluate the accuracy of the objective assessment based on the OFW, healthy and pathological phonation are explored.

TABLE I. QUANTIFICATION OF THE GLOTTAL PARAMETERS. RELAXED VOICE (103 CYCLES, 4485 IMAGES). BY USING EGG, $F_0 = 155$

| | $\mathcal{L} = 25\%$ | | $\mathcal{L} = 50\%$ | | $\mathcal{L} = 75\%$ | |
|-------|----------------------|-----------|----------------------|--------------------|----------------------|----------|
| | DKG | OFW | DKG | OFW | DKG | OFW |
| F_0 | 153 | L/R:152 | 153 | L/R:152 | 153 | L/R:152 |
| ASI | -0.04 | 0.02 | 0.06 | -0.03 | 0.02 | -0.05 |
| PSI | -0.02 | L/R:0.001 | 0.01 | L/R:0.0016 | 0 | L/R:0 |
| TPI | 0.96 | L/R:0.87 | 0.96 | R: 0.9, L: 0.93 | 0.96 | L/R:0.92 |
| API | 0.92 | 0.93 | 0.92 | 0.92 | 0.92 | 0.82 |

A. Healthy vocal folds

In healthy phonation, ideally, the waveforms of the left and right vocal folds should be as superimposed as possible ensuring phase and amplitude symmetries. In addition, the same pattern should be observed among many cycles. Quantitatively, the P-values resulting from the ANOVA test should be greater than 0.05, the estimates of the fundamental frequencies of the left and right vibrations should be close to each other, the symmetry indices (such as ASI and PSI) should approach 0 and the periodicity indices (as TPI and API) should approach 1. The quantitative measures obtained by the OFW are compared to those obtained by the analysis of the EGG signals and the DKG. The analysis of the EGG signal is performed according to the technique described in [14] and using the MOQ software².

Tests are conducted on four different types of phonation using several laryngeal mechanisms on healthy vocal folds. The aim is to evaluate the accuracy of the objective measures computed through the analysis of the OFW. The P-values computed through the ANOVA test confirm the similarity in vibrations' amplitudes of the right and the left vocal folds for all the tested mechanisms for 97% of the cycles at the glottal level $\mathcal{L} = 25\%$, 94% of the cycles at the level $\mathcal{L} = 50\%$ and 93% of the cycles at the level $\mathcal{L} = 75\%$ level. As depicted in Tables I, II, III and IV, the fundamental frequency estimated according to the proposed approach is the same for both vocal folds, and within the three glottal levels, in all the sequences related to healthy phonation. Moreover, the estimates are close to the ones estimated by DKG and EGG for all the sequences.

To evaluate the periodicity of the vibrations, the TPI and API are used and their values are found to be above 0.8 for all the sequences. Note that the final decision about how well the vocal folds are healthy is delegated to the clinician based on the provided quantitative measures and his experience. The vibratory parameters measured using DKG in one hand, and using the proposed approach on the other hand, have the same values in 70% of the tested situations including the different glottal levels. In the remaining 30%, the difference between the values ranges between 0.02 and 0.14.

Concerning the symmetry indices, they are 0 for the majority of cases and multiples of 0.001 and 0.01 in some tests.

B. Pathological Vocal Folds

HSVs related to different vocal fold disorders are used to evaluate the efficiency and the reliability of the OFW in assisting clinicians in the diagnosis. The first video is about a right true vocal cord paralysis where the cause of

TABLE II. QUANTIFICATION OF THE GLOTTAL PARAMETERS. (SAMPLE USC2008.02.050.OUV.RENF.50.FRAMES OF THE DATABASE, 130 CYCLES, 7200 IMAGES). BY USING EGG, $F_0 = 115$

| | $\mathcal{L} = 25\%$ | | $\mathcal{L} = 50\%$ | | $\mathcal{L} = 75\%$ | |
|-------|----------------------|---------------------|----------------------|---------------------|----------------------|---------|
| | DKG | OFW | DKG | OFW | DKG | OFW |
| F_0 | 124 | L/R: 121 | 124 | L/R:121 | 124 | L/R:121 |
| ASI | -0.04 | 0.005 | -0.03 | 0.01 | -0.05 | -0.04 |
| PSI | -0.03 | L/R:0 | -0.03 | L/R: 0 | -0.02 | L/R:0 |
| TPI | 0.8 | R: 0.86, L: 0.87 | 0.8 | R: 0.83, L: 0.81 | 0.8 | L/R:0.8 |
| API | 0.95 | 0.9 | 0.98 | 0.9 | 0.96 | 0.8 |

TABLE III. QUANTIFICATION OF THE GLOTTAL PARAMETERS (SAMPLE USC2008.02.065.GLISSENDO.65.FRAMES81 CYCLES, 1158 IMAGES). BY USING EGG, $F_0 = 510$

| | $\mathcal{L} = 25\%$ | | $\mathcal{L} = 50\%$ | | $\mathcal{L} = 75\%$ | |
|-------|----------------------|---------------------|----------------------|---------------|----------------------|---------------------|
| | DKG | OFW | DKG | OFW | DKG | OFW |
| F_0 | 467 | L/R:471 | 467 | L/R:471 | 467 | L/R: 472 |
| ASI | 0.01 | -0.007 | -0.09 | 0.02 | -0.14 | 0.01 |
| PSI | -0.07 | R: 0, L: 0 | -0.02 | R: 0, L: 0 | -0.05 | L/R: 0 |
| TPI | 0.93 | R: 0.91, L: 0.92 | 0.93 | L/R: 0.93 | 0.93 | R: 0.92, L: 0.93 |
| API | 0.9 | 0.8 | 0.97 | 0.9 | 0.98 | 0.9 |

the paralysis is unknown or idiopathic. The video is sampled into 940 images corresponding to 9 cycles according to the cycle detection approach proposed in [11]. The waveforms of the right and the left VF are generated and the ANOVA test is carried out. The majority of the P-values are less than the confidence level confirming the vibration asymmetries between the two VF. In addition, the estimated fundamental frequencies of both waveforms related to the left and the right VF are different and have suspicious values especially at 25% of the glottis, as shown in Table V. These values could suspect the presence of an anomaly, precisely in the right cord where the frequency is 0.02 Hz at 25% and 9 Hz at the two other glottal levels, compared to a frequency of 21 Hz at the left vocal cord for the three levels. Even though the fundamental frequency estimated by DKG has a suspicious value compared to the frequency range related to healthy VF, the OFW is found to be more sensitive to a pathologic phonation. Another suspicious value is given by the TPI parameter which is 0.5, expected to approach 1 in a healthy phonation.

The patient of the second video suffers from a cancer in the larynx involving the left vocal fold. The majority of the P-values is less than the confidence level. The values of the TPI and the API parameters in Table VI let suspect an aperiodicity of the VF vibrations. Also, the estimates of the

TABLE IV. QUANTIFICATION OF THE GLOTTAL PARAMETERS (SAMPLE USC2008.02.004.M1.4.FRAMES, 51 CYCLES, 655 IMAGES). BY USING EGG, $F_0 = 590$

| | $\mathcal{L} = 25\%$ | | $\mathcal{L} = 50\%$ | | $\mathcal{L} = 75\%$ | |
|-------|----------------------|----------|----------------------|--------------------|----------------------|----------|
| | DKG | OFW | DKG | OFW | DKG | OFW |
| F_0 | 522 | L/R:518 | 522 | L/R:517 | 522 | L/R:518 |
| ASI | -0.09 | 0.09 | 0.09 | -0.04 | -0.02 | 0.04 |
| PSI | -0.08 | L/R: 0 | -0.05 | L/R: 0 | 0.005 | L/R: 0 |
| TPI | 0.9 | L/R: 0.9 | 0.92 | R: 0.9, L: 0.92 | 0.92 | L/R:0.88 |
| API | 0.9 | 0.8 | 0.94 | 0.8 | 0.94 | 0.88 |

²available at <http://voiceresearch.free.fr/egg/>

TABLE V. QUANTIFICATION OF THE GLOTTAL PARAMETERS OF DISORDERED VFS. RIGHT TRUE VOCAL CORD PARALYSIS (9 CYCLES)

| | $\mathcal{L} = 25\%$ | | $\mathcal{L} = 50\%$ | | $\mathcal{L} = 75\%$ | |
|-------|----------------------|------------------------------------|----------------------|------------------------------------|----------------------|-------------------------------|
| | DKG | OFW | DKG | OFW | DKG | OFW |
| F_0 | 79 | R: 0.02 , L: 21 | 79 | R: 9 , L: 21 | 79 | R: 9 , L: 20 |
| ASI | 0.003 | -0.04 | -0.01 | -0.01 | 0.09 | 0.16 |
| PSI | 0.02 | L/R:0 | 0.01 | L/R 0 | 0.12 | L/R:0 |
| TPI | 0.5 | R: 0.58 , L: 0.43 | 0.5 | R: 0.58 , L: 0.47 | 0.5 | L/R: 0.56 |
| API | 0.87 | 0.74 | 0.91 | 0.75 | 0.86 | 0.76 |

TABLE VI. QUANTIFICATION OF THE GLOTTAL PARAMETERS OF DISORDERED VFS. CANCER OF THE LEFT TRUE VOCAL CORD (5 CYCLES,552 IMAGES)

| | $\mathcal{L} = 25\%$ | | $\mathcal{L} = 50\%$ | | $\mathcal{L} = 75\%$ | |
|-------|----------------------|--------------------------|----------------------|--------------------|----------------------|--------------------|
| | DKG | OFW | DKG | OFW | DKG | OFW |
| F_0 | 61 | R: 30, L: 0.04 | 61 | R: 20, L: 60 | 61 | R: 20, L: 41 |
| ASI | 0.17 | -0.15 | 0.01 | -0.08 | 0.04 | -0.24 |
| PSI | 0.05 | L/R:0 | 0.04 | L/R 0 | 0.008 | L/R:0 |
| TPI | 0.82 | R: 0.43, L: 0.72 | 0.82 | R: 0.46, L: 0.6 | 0.8 | R: 0.58, L: 0.8 |
| API | 0.51 | 0.63 | 0.67 | 0.7 | 0.58 | 0.48 |

fundamental frequencies of both waveforms have suspicious values, especially the left vocal fold at 25% of the glottis, leading to conclude an anomaly in that glottal part which is in concordance with the ground truth.

The third video is about a left true vocal cord paralysis from injury to the Vagus Nerve during carotid endarterectomy surgery. The P-values of the ANOVA test are less than the confidence level for more than half of the cycles which confirms the pathological aspect of the vibrations. Besides, the fundamental frequency estimates of the right and the left waveforms are close to each other as shown in Table VII, but their values let suspect an abnormality in the vibrations. Also, the values taken by the periodicity parameters (TPI and API) suspect disordered VF.

IV. DISCUSSION

Numerous techniques are proposed in the literature allowing to map the spatial-temporal information contained in a HSV recording into a representation suitable for the visual and the automatic assessment of the vibrations. Most of the state-of-the-art representations require an additional processing, like applying a segmentation technique, in order to quantify laryngeal parameters. In contrast, it is possible to directly measure these parameters by using the OFW.

TABLE VII. QUANTIFICATION OF THE GLOTTAL PARAMETERS OF DISORDERED VFS. ABDUCTOR PARALYSIS (8CYCLES, 860 IMAGES)

| | $\mathcal{L} = 25\%$ | | $\mathcal{L} = 50\%$ | | $\mathcal{L} = 75\%$ | |
|-------|----------------------|--------------------|----------------------|---------------------|----------------------|----------------|
| | DKG | OFW | DKG | OFW | DKG | OFW |
| F_0 | 63 | R: 33, L: 34 | 63 | R: 31, L: 33 | 63 | R:31, L: 61 |
| ASI | -0.02 | 0.01 | 0.04 | 0.02 | -0.16 | 0.11 |
| PSI | -0.01 | L/R:0 | -0.003 | L/R 0 | -0.04 | L/R:0 |
| TPI | 0.85 | R: 0.62, L: 0.7 | 0.85 | R: 0.68, L: 0.67 | 0.85 | L/R:0.77 |
| API | 0.9 | 0.7 | 0.91 | 0.7 | 0.87 | 0.66 |

Moreover, the present study shows that the exploitation of the OFW allows to quantify the most commonly used laryngeal parameters namely the fundamental frequency of the vibrations, the amplitude/phase symmetry parameters and the time/amplitude periodicity without need of dealing with intermediate metrics specific to OFW.

Healthy and pathological vibrations are explored in this study.

Regarding healthy phonation, different mechanisms used by different persons are considered. The quantification accuracy is evaluated in comparison to the measures obtained through the analysis of the corresponding EGG signals (for the evaluation of the fundamental frequency estimate) and the DKG. For all the tested sequences, the fundamental frequency estimated through the OFW is close to those estimated using the EGG signals and DKG. Also, the measures of the symmetry and the periodicity obtained through analyzing the OFW are close to the measures obtained through DKG, for the three glottal levels.

Regarding the assessment of pathological vibrations, tests conducted on sequences of different disorders show the reliability of the OFW in suspecting a vibratory anomaly. Furthermore, it is shown that the OFW is more sensitive to a pathological phonation than DKG.

V. CONCLUSION

The objective assessment of the vocal fold vibrations is the main focus of the present study. To this aim, the optical flow based waveform is retained. Its generation does not require the segmentation of all the images of the HSV. More importantly, the present paper shows that the quantification of the vocal fold vibrations by using OFW does not necessitate an additional processing, contrary to the majority of the state-of-the-art techniques. Furthermore, it is possible to quantify the commonly used laryngeal parameters. The results show the reliability of the OFW in quantifying the VF vibrations. Moreover, it is more sensitive to a pathological phonation.

REFERENCES

- [1] I. R. Titze, "Parameterization of the glottal area, glottal flow, and vocal fold contact area," *The Journal of the Acoustical Society of America*, vol. 75, no. 2, pp. 570–580, 1984.
- [2] T. Wittenberg, M. Tigges, P. Mergell, and U. Eysholdt, "Functional imaging of vocal fold vibration: digital multislice high-speed kymography," *Journal of Voice*, vol. 14, no. 3, pp. 422–442, 2000.
- [3] 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 transactions on medical imaging*, vol. 27, no. 3, pp. 300–309, 2008.
- [4] S.-Z. Karakozoglou, N. Henrich, C. d'Alessandro, and Y. Stylianou, "Automatic glottal segmentation using local-based active contours and application to glottovibrography," *Speech Communication*, vol. 54, no. 5, pp. 641–654, 2012.
- [5] R. Menon, L. Petropoulakis, J. J. Soraghan, H. Lakany, K. MacKenzie, O. Hilmi, and G. Di Caterina, "Automatic quantification of vocal cord paralysis-an application of fibre-optic endoscopy video processing," 2017.

- [6] J. Lin, E. S. Walsted, V. Backer, J. H. Hull, and D. Elson, "Quantification and analysis of laryngeal closure from endoscopic videos," *IEEE Transactions on Biomedical Engineering*, 2018.
- [7] 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*, vol. 55, no. 3, pp. 128–136, 2003.
- [8] J. Unger, D. J. Hecker, M. Kunduk, M. Schuster, B. Schick, and J. Lohscheller, "Quantifying spatiotemporal properties of vocal fold dynamics based on a multiscale analysis of phonovibrograms," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 9, pp. 2422–2433, 2014.
- [9] A. Yamauchi, H. Yokonishi, H. Imagawa, K.-I. Sakakibara, T. Nito, N. Tayama, and T. Yamasoba, "Quantification of vocal fold vibration in various laryngeal disorders using high-speed digital imaging," *Journal of Voice*, vol. 30, no. 2, pp. 205–214, 2016.
- [10] A. Yamauchi, H. Imagawa, K.-I. Sakakibara, H. Yokonishi, T. Nito, T. Yamasoba, and N. Tayama, "Phase difference of vocally healthy subjects in high-speed digital imaging analyzed with laryngotopography," *Journal of Voice*, vol. 27, no. 1, pp. 39–45, 2013.
- [11] H. Ammar, "Optical flow based waveform for the assessment of the vocal fold vibrations," *Australasian Physical & Engineering Sciences in Medicine*, vol. 42, no. 1, pp. 91–109, 2019.
- [12] H. S. Bonilha, D. D. Deliyski, and T. T. Gerlach, "Phase asymmetries in normophonic speakers: visual judgments and objective findings," *American Journal of Speech-Language Pathology*, vol. 17, no. 4, pp. 367–376, 2008.
- [13] C. R. Krausert, D. Ying, Y. Zhang, and J. J. Jiang, "Quantitative study of vibrational symmetry of injured vocal folds via digital kymography in excised canine larynges," *Journal of Speech, Language, and Hearing Research*, vol. 54, no. 4, pp. 1022–1038, 2011.
- [14] D. M. Howard, G. A. Lindsey, and B. Allen, "Toward the quantification of vocal efficiency," *Journal of Voice*, vol. 4, no. 3, pp. 205–212, 1990.
- [15] G. Degottex, E. Bianco, and X. Rodet, "Usual to particular phonatory situations studied with high-speed videoendoscopy," in *International Conference on Voice Physiology and Biomechanics*, 2008, pp. 19–26.
- [16] G. Degottex, "Glottal source and vocal-tract separation: estimation of glottal parameters, voice transformation and synthesis using a glottal model," Ph.D. dissertation, Paris 6, 2010.
- [17] D. Deliyski, K. Kendall, and R. Leonard, "Laryngeal high-speed videoendoscopy," *Laryngeal evaluation: Indirect laryngoscopy to high-speed digital imaging*, pp. 245–270, 2010.
- [18] J. J. Jiang, C.-I. Chang, J. R. Raviv, S. Gupta, F. M. Banzali, and D. G. Hanson, "Quantitative study of mucosal wave via videokymography in canine larynges," *The Laryngoscope*, vol. 110, no. 9, pp. 1567–1573, 2000.
- [19] J. G. Švec, F. Šram, and H. K. Schutte, "Videokymography in voice disorders: What to look for?" *Annals of Otology, Rhinology & Laryngology*, vol. 116, no. 3, pp. 172–180, 2007.