HHS Public Access

Author manuscript

Conf Proc IEEE Eng Med Biol Soc. Author manuscript; available in PMC 2019 October 23.

Published in final edited form as:

Conf Proc IEEE Eng Med Biol Soc. 2015 August ; 2015: 735–738. doi:10.1109/EMBC.2015.7318467.

Voice pathology classification based on High-Speed Videoendoscopy

D. Panek,
AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Krakow, Department of Measurement and Electronics, Poland

A. Skalski [Member, IEEE],
AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Krakow, Department of Measurement and Electronics, Poland

T. Zielinski [Member, IEEE],
AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Krakow, Department of Telecommunications, Poland

D.D. Deliyski
Communication Sciences Research Center, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Abstract

This work presents a method for automatical and objective classification of patients with healthy and pathological vocal fold vibration impairments using High-Speed Videoendoscopy of the larynx. We used an image segmentation and extraction of a novel set of numerical parameters describing the spatio-temporal dynamics of vocal folds to classification according to the normal and pathological cases and achieved 73,3% cross-validation classification accuracy. This approach is promising to develop an automatic diagnosis tool of voice disorders.

I. Introduction

In the last few years, the analysis of high frequency endoscopic video sequences called High-Speed Videoendoscopy (HSV), enabling recording the movement of vocal cords, has became one of the fastest growing diagnosis methods of voice disorders. Examination of vocal cords is an essential part of clinical evaluation of voice. However, in clinical practice an acoustic analysis still remains the most instrumental measure used for assessing glottal aerodynamics and providing valuable information on a speakers vocal function. Acoustic parameters are free from bias and may provide a quantitative measure for perceptual voice characteristics [1, 2]. To gain more accurate information about the glottal source and supra-glottal glottis, it is needed to add visual information of the entire vocal fold. High-Speed Videoendoscopy offer significant advantages over other techniques. Extraction and analysis

dpanek@agh.edu.pl.

of the vocal waveform can provide quality information of the vocal fold movement as well as the changes in glottal airflow depending on time.

HSV is particularly useful for visualization and quantitative evaluation of pathology, which affects the dynamics of the vocal folds [3]. A common quality criterion for the discrimination between healthy and pathological vocal fold vibration patterns is the symmetry and oscillating constancy of the vocal folds [4,12]. Endoscope records the vibrations of the vocal cords performed during the phonation. Fundamental frequency for adult ranges from 80–300 Hz. Due to this fact the temporal resolution of visual system should be able to capture in detailed oscillating vocal folds. HSV enables recording every cycle of the oscillation irrespective of perturbation, enables recording of highly disturbed and aperiodic signals, registers the intra-cycle vibratory behavior through a full image of the vocal folds [5, 6]. Beside the advantages mentioned above, the main disadvantage of the High-Speed Videoendoscopy is high acquisition costs and time consuming data preprocessing such as segmentation.

Image segmentation is an essential tool for further HSV-based vocal fold analysis. Many authors present segmentation algorithms based on region growing or thresholding [7, 8, 22]. In literature there are different approaches focused on edge detection methods of the vocal folds and active contour models for the segmentation process [9, 10, 22].

Many promising approaches have been revealed to access the subjective and objective analysis methods to appraise HSV recordings [11–14]. The most frequent parameters are occurrence of mucosal wave, glottal closure, vibratory amplitude [15, 16]. Additionally, instabilities of the fundamental frequencies, amplitude, symmetry and regularity parameters can be calculated giving quantitative and qualitative information about irregular fold vibrations [17–20, 28]. The work of [21] evaluate parameters quantifying the spatio-temporal correlation along the anterior-posterior dimension of the varying number of pixels between left and right vocal fold contours. No conclusion can be deduced about the lateral symmetry and synchronicity due to the fact that no distinction is made between left and right vocal fold vibrations. There is still no common standard to automatically analyze the entire oscillation pattern of both vocal folds.

In this work, we present a novel method to segment the glottal area and vocal fold edges (fig. 1) being a modification of our previous approach [22]. Based on the segmentation results we describe the spatio-temporal dynamics of vocal folds and shape by means of 8 parameters. In the paper the HSV recordings of 30 healthy and pathological subjects have been analyzed and their classification to normal and pathological cases have been done. Obtained features describing the behavior of vocal folds during the phonation were classified using K-means algorithm.

II. Data

The vocal fold recordings of more than 30 patients have been captured with HSV recording equipment. The diagnoses were made by physicians, what stands for a gold standard for evaluation of automatic classification results. This way a population of 15 patients with

diagnosed different voice impairments was found. Moreover, as a reference we used a population of normal, healthy voices, 15 healthy candidates were recorded. During the examination the patients were asked to keep the phonation of the sustained vowel /i/. This is due to the fact that the subject cannot close the vocal tract with an inserted endoscope. The temporal resolution of visual system captures 2000 frames per second.

III. Methods

A. Image segmentation

In order to extract and evaluate the change of vocal folds parameters over time, the glottal area was first segmented in the recorded HSV recordings. Applied segmentation methodology was based on our previous work [22] in which the level set method described by Osher and Sethian [23] had been used. Level sets is useful for capturing changing topologies, because merging and breaking are made automatically in it. This is very important feature in HSV segmentation context (fig. 1, second image from left) because it frequently happens that vocal folds are temporarily, partially glued during vibration.

The contour is defined as zero level set of time dependent function $\varphi(t, x, y)$. Its evolution equation is described by [22].

$$\frac{\partial \phi}{\partial t} + F|\nabla \phi| = 0 \tag{1}$$

where function F means a speed function and is related to the function φ and image data. In this work, in contrary to [22], we use variational formulation of the function F [24]. The final evolution equation is based on [23]:

$$\frac{\partial \phi}{\partial t} = \omega\left[\Delta\phi - div\left(\frac{\nabla\phi}{|\nabla\phi|}\right)\right] + \lambda\delta(\phi)\, div\left(g\frac{\nabla\phi}{|\nabla\phi|}\right) + \nu g\delta(\phi) \tag{2}$$

where g signify an edge indicator function, $\delta(\varphi)$ means a regularized Dirac function and ν, λ, ω are constant values. More details can be found in [22]. Segmentation effectiveness in qualitative and quantitative form are presented in fig. 1 and in section IV, respectively.

To evaluate the performance of the automatic segmentation we used the Dice similarity coefficient (DSC), which is commonly used in evaluating segmentation performance and its values is extends 0, which means no overlap and 1, meaning ideal overlap. DSC is one of measures describing spatial overlap between binary images. DSC was obtained using [25]:

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \quad (3)$$

where A is a binary image from automatic segmentation, B is a binary image from manual segmentation. Manual segmentation was done by 3 different specialists. We have randomly chosen 63 frames from 30 different recordings and calculated DSC. The final Dice similarity coefficient is the mean of 3 comparisons.

B. Feature extraction

The first step towards feature extraction based on HSV recordings consisted in automatic detection of the oscillation cycles. Due to the fact, that the first and the last vocal fold oscillation cycle could be incomplete in the HSV recording, we removed two outermost cycles. We computed following parameters from the glottal signal: fundamental frequency, Jitter and Shimmer coefficients, bigger to smaller side of glottal area ratio, curvature and correlation coefficient of displacement vectors on both sides of the vocal folds, Bhattacharyya distance and a parameter describing the difference in respect to the regular phonation.

The first parameter, the fundamental frequency, was possible to calculate using the glottal area waveform. Fundamental frequency enables to measure the number of glottal oscillation cycles occurring in one second.

Jitter coefficient means a short-term cycle-to-cycle perturbation in the fundamental frequency, whereas Shimmer coefficient means a short-time cycle-to-cycle perturbation in amplitude. In this paper those parameters were used to describe frequency and amplitude variability. The Jitter and Shimmer coefficients were calculated as the percentage of standard deviation value to the mean value of the cycle-to-cycle variation in frequency and amplitude [15].

To calculate the next parameter, the bigger to smaller side of glottal area ratio, it was necessary to designate the major axis of the glottal area. To determine this we have proposed a method for finding the ellipse, which is surrounding the segmented object, assuming minimization of the area (fig. 2B). Using the slope of this line it was possible to divide the contour of the vocal cords into two parts: left and right, respectively Thus, it allowed the measurement of the surface of the glottal area on the right and left side. Since the endoscope may be placed with different distance to the glottis, a ratio of the bigger side of the glottal area to the smaller one was being calculated.

The next parameter was a curvature. The curvature κ is the change in tangent direction as it is moved along the curve, and hence is approximated by the ratio between the change in the tangent direction and Δs [24,26]:

$$\kappa = \frac{\alpha}{\Delta s} \tag{4}$$

where α is an angle describing how the tangent direction changes from point to point along the sampled curve, what is called *turning angle*, and $\Delta s = L/n$, where L is the length of analyzed curve sampled at n uniformly-spaced points. We took this parameter into account as it is a quantitative measure describing the degree to which the analyzed object deviates in its properties from a straight line. The curvature was calculated separately on right and left side of GA with the usage of the algorithm described in [26]. To measure the similarity between left and right side of the curvature we have used the correlation coefficient.

To analyze the symmetry of vocal folds we have calculated the displacement vectors related to the left and right fold (fig. 2C–D). The vectors on both sides were compared in analogy to the curvature using the correlation coefficient. For healthy patients, the correlation coefficient should be equal to 1, as both vocal folds should move symmetrically.

The 7th calculated parameter was a Bhattacharyya distance, which enables to measure the similarity between two histograms containing information about the area on right and left side of GA. The Bhattacharyya distance is calculated as follows:

$$B = \frac{1}{8}(\mu_2 - \mu_1)^T \left[1/2 \cdot (\Sigma_1 + \Sigma_2)\right]^{-1}(\mu_2 - \mu_1) + \\ + \frac{1}{2}\ln\frac{\left|1/2 \cdot (\Sigma_1 + \Sigma_2)\right|}{\left|\Sigma_1\right|^{1/2}\left|\Sigma_2\right|^{1/2}} \tag{5}$$

where μ_i and $\sum_i$ are the mean vector and covariance matrix of class i.

The last calculated parameter aimed at indicating the difference between measured signal and the regular, cyclic one. Firstly, we have measured the duration and amplitude of each cycle of the proper move of vocal folds. It represented correct, regular cycle of movement of the vocal folds during a phonation. In the next step, we compared this signal with the one, which was obtained during the patient examination. Integrating the difference between two signals we received the last required measure.

To organize feature characteristics according to their discriminative ability and, as a consequence, to obtain stable and consistent results we have used Principal Component Analysis (PCA). PCA also enabled reduction of the number of parameters up to 5, leaving

99% variance in the data. The PCA was used as a pre-treatment step prior to further data analysis, simplifying subsequent calculations.

IV. Results

The Dice similarity coefficient representing spatial overlap between automatic and manual segmentation was equal 0,84±0.02.

In order to classify the data into two groups, healthy or pathological voice, we used K-means algorithm, which ensures that the total distance is minimized between the group's members and its corresponding centroid. To validate created feature vector we calculated 10-fold cross validation. To estimate the quality assessment the parameters such as: accuracy, precision for healthy and pathological cases and specificity for healthy and pathological cases were calculated [27].

V. Conclusion

The observation of the vocal folds movement during the phonation is an essential part of clinical examination. In this paper a novel approach of a quantitative analysis of vocal folds during the phonation is presented. The purpose of this study was to design an automatic segmentation of the vocal folds and provide measurements and automatic classification between healthy patients and those with voice impairments. We achieved 73,3% accuracy for all analyzed material. This result is very encouraging and it is expected that the presented tools will be useful to preventative health care in laryngology.

Acknowledgment

This work was funded by the Ministry of Science and Higher Education in Poland under the Diamond Grant program, decision number 0136/DIA/2013/42 (AGH 68.68.120.364). Funding for part of this study was provided by the National Institutes of Health-NIDCD: Grant R01-DC007640 "Efficacy of Lanryngeal High-Speed Videoendoscopy".

References

[1]. Yuling Y, Damrose E, Bless D. "Functional analysis of voice using simultaneous high-speed imaging and acoustic recordings", Journal of Voice, vol. 21, no. 5, 2007, pp. 604–616. [PubMed: 16968665]

[2]. Panek D, Skalski A, Gajda J, Tadeusiewicz R, "Acoustic analysis assessment in speech pathology detection (accepted for publication)", International Journal of Applied Mathematics and Computer Science, to be published.

[3]. Ikuma T, Kunduk M, McWhorter AJ, "Objective Quantification of Pre-and Postphonosurgery Vocal Fold Vibratory Characteristics Using High-Speed Videoendoscopy and a Harmonic Waveform Model", Journal of Speech, Language, and Hearing Research, vol. 57, no. 3, 2014, pp. 743–757.

[4]. Hoppe U, "Mechanisms of hoarseness - visualization and interpretation by means of nonlinear dynamics", Aachen, Germany: Shaker, 2001.

[5]. Deliyski DD, Petrushev PP, Bonilha HS, Gerlach TT, Martin-Harris B, Hillman RE, "Clinical implementation of laryngeal high-speed videoendoscopy: challanges and evolution", Folia Phoniatrica Logopaedica, no. 60, 2008, pp.33–44.

[6]. Patel R, Dailey S, Bless D, “Comparison of high-speed digital imaging with stroboscopy for laryngeal imaging of glottal disorders”, Annals of Otology, Rhinology & Laryngology, no.117, 2008, pp. 413–424.

[7]. Chen X, Bless D, Yan Y, “A segmentation scheme based on Rayleigh distribution model for extracting glottal waveform from high-speed laryngeal images”, in Engineering in Medicine and Biology Society, Jan. 2006, pp. 6269–6272.

[8]. Lohscheller J, Toy H, Rosanowski F, Eysholdt U, Döllinger M, “Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos”, Medical Image Analysis, vol. 11, no. 4, 2007, pp. 400–413. [PubMed: 17544839]

[9]. Scholl I, Sovakar A, Lehmann T, Spitzer K, “Motion analysis of vocal folds using adaptive snakes”, Advances in Quantitative Laryngoscopy, Verlang Abteilung Phoniatrie, 1997, pp. 29–38.

[10]. Dollinger M, Hoppe U, Hettlich F, Lohscheller J, Schuberth S, Eysholdt U, “Vibration parameter extraction from endoscopic image series of the vocal folds”, Biomedical Engineering, IEEE Transactions on, vol. 49, no. 8, 2002, pp. 773–781.

[11]. Inwald EC, Döllinger M, Schuster M, Eysholdt U, Bohr C, “Multiparametric analysis of vocal fold vibrations in healthy and disordered voices in high-speed imaging”, Journal of Voice, vol. 25, no. 5, 2011, pp. 576–590. [PubMed: 20728308]

[12]. Voigt D, Döllinger M, Braunschweig T, Yang A, Eysholdt U, Lohscheller J, “Classification of functional voice disorders based on phonovibrograms”, Artificial Intelligence in Medicine, vol. 49, no. 1, 2011, pp. 51–59.

[13]. Döllinger M, Kunduk M, Kaltenbacher M, Vondenhoff S, Ziethe A, Eysholdt U Bohr C, “Analysis of vocal fold function from acoustic data simultaneously recorded with high-speed endoscopy”, Journal of Voice, vol. 26, no. 6, 2012, 726–733. [PubMed: 22632795]

[14]. Mehta DD, Deliyski DD, Zeitels SM, Quatieri TF, Hillman RE, “Voice production mechanisms following phonosurgical treatment of early glottic cancer”, Annals of Otology, Rhinology & Laryngology, vol. 119, no. 1, 2010, pp. 1–9.

[15]. Yan Y, Ahmad K, Kunduk M, Bless D, “Analysis of vocal-fold vibrations from high-speed laryngeal images using a Hilbert transform-based methodology”, Journal of Voice, vol. 19, no. 2, 2005, pp. 161–175. [PubMed: 15907431]

[16]. Lohscheller J, Eysholdt U, Toy H, Dollinger M, “Phonovibrography: mapping high-speed movies of vocal fold vibrations into 2-D diagrams for visualizing and analyzing the underlying laryngeal dynamics”, Medical Imaging, IEEE Transactions on, vol. 27, no.3, 2008, pp. 300–309.

[17]. Mergell P, Herzel H, Titze IR, “Irregular vocal-fold vibration—high-speed observation and modeling”, The Journal of the Acoustical Society of America, vol. 108, no. 6, 2000, pp. 2996–3002. [PubMed: 11144591]

[18]. Eysholdt U, Rosanowski F, Hoppe U, “Measurement and interpretation of irregular vocal cord fold vibrations”, HNO, vol. 51, no.9, 2003, pp. 710–716. [PubMed: 12955248]

[19]. Tokuda I, Herzel H, „Detecting synchronizations in an asymmetric vocal fold model from time series data”, Chaos An Interdisciplinary Journal of Nonlinear Science, vol. 15, no. 1, 2005, pp. 013702–1–11.

[20]. Unger J, Hecker D, Kunduk M, Schuster M, Schick B, Lohscheller J, “Quantifying spatiotemporal properties of vocal fold dynamics based on a multiscale analysis of phonovibrograms”, IEEE Engineering in Medicine and Biology Society, vol. 61, no. 9, 2014, pp. 2422–2433.

[21]. Krausert CR, Liang Y, Zhang Y, Rieves AL, Geurink KR, Jiang JJ, “Spatiotemporal analysis of normal and pathological human vocal fold vibrations”, American Journal of Otolaryngology, vol. 33, no.6, 2012, 641–649. [PubMed: 22841342]

[22]. Skalski A, Zielinski T, Deliyski D, “Analysis of vocal folds movement in high speed videoendoscopy based on level set segmentation and image registration”, Int. Conf. on Signals and Electronic Systems ICSES, Krakow 2008, pp. 223–226.

[23]. Osher S, Sethian JA, “Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations”, Journal of Computational Physics, vol. 114, 1988, pp.12–49.

[24]. Li C, Xu CG, Fox MD, “Level Set Evolution without Reinitialization: A New Variational Formulation”, IEEE CVPR, 2005, pp. 430–436.

[25]. Dice LR, “Measures of the amount of ecologic association between species” Ecology, vol.26, no. 3, 1945, pp. 297–302.

[26]. Feldman J, Singh M, “Information along contours and object boundaries”, Psyhological Review, vol. 112, no.1, 2005, pp. 243–252.

[27]. Panek D, Skalski A, Gajda J, “Quantification of Linear and Nonlinear Acoustic Analysis Applied to Voice Pathology Detection”, in Information Technologies in Biomedicine, vol. 4, 2014, pp. 355–364.

[28]. Doellinger M, Lohscheller J, McWhorter A, Kunduk M, “Variability of normal vocal fold dynamics for different vocal loading in one healthy subject investigated by phonovibrograms”, Journal of Voice, vol. 23, no. 2, 2009, pp. 175–181. [PubMed: 18313896]

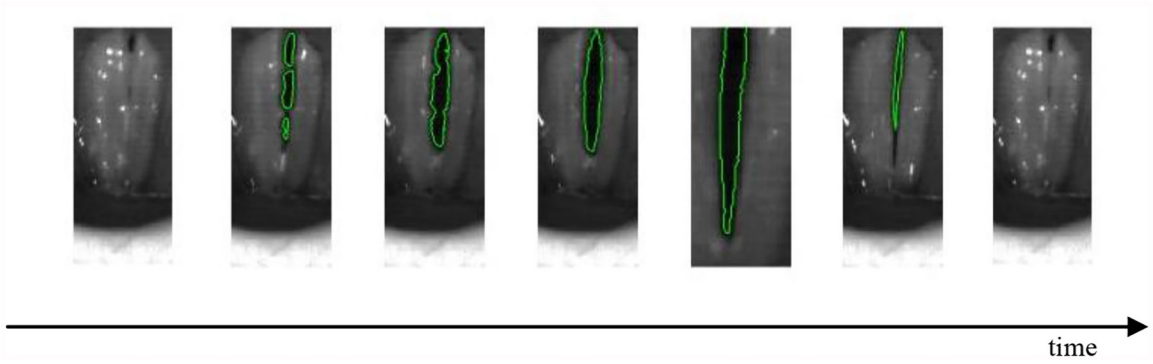

Figure 1.
Examplary results: single HSV video oscilation presenting the vocal fold movement and countour being the result of the segmentation.

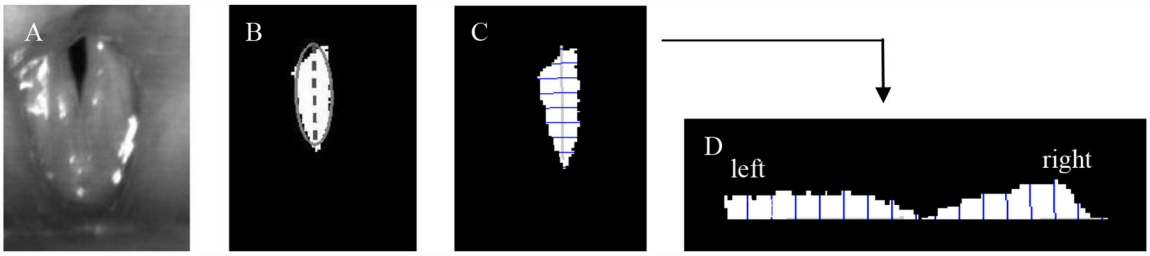

Figure 2.

A) Sample frame from HSV, B) Segmented glottal area with the ellipse and major axis C) displacement vectors in respect to major axis, D) displacement vectors on the right and left side of the glottal area.

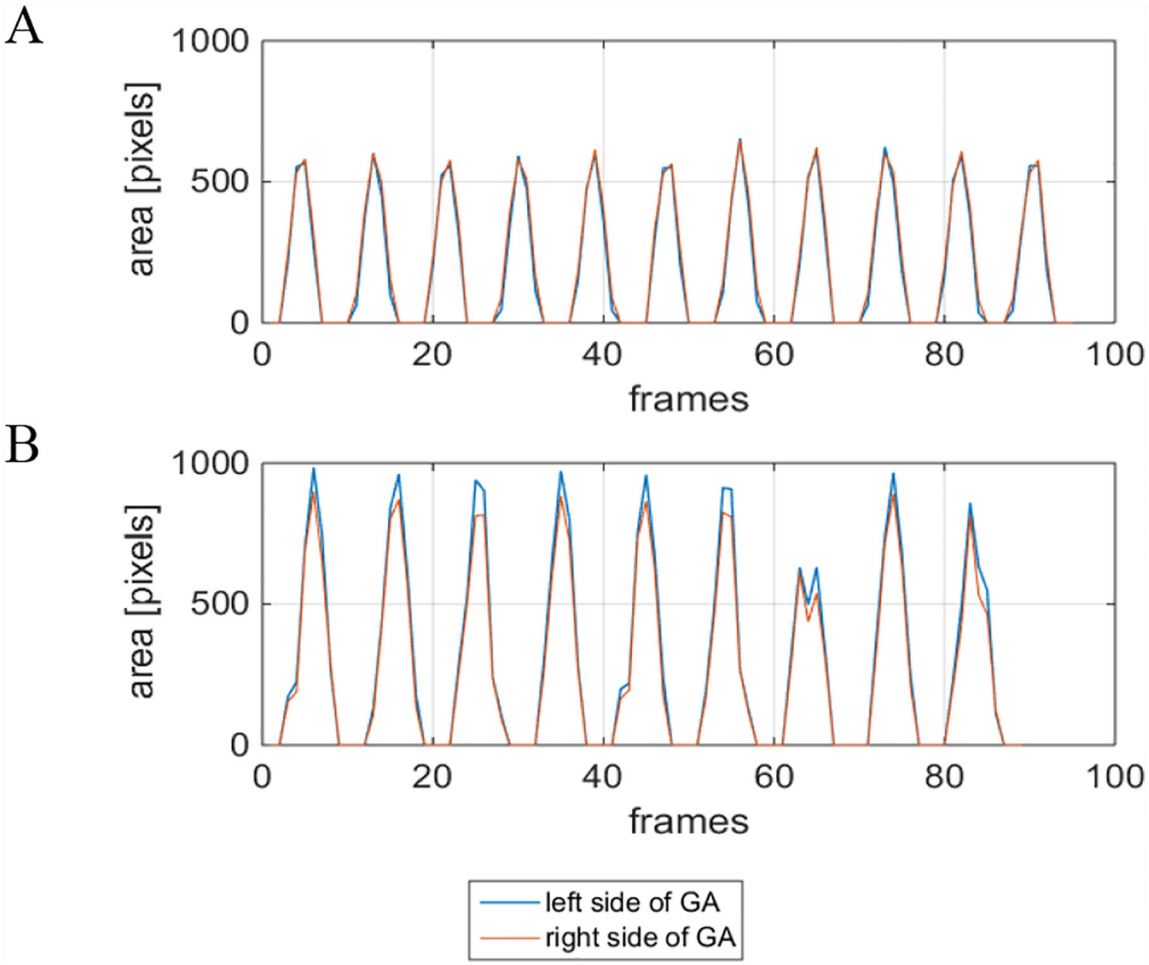

Figure 4.

GA - glottis area, comparison of total area on right and left side of the glottis for: A) healthy patient B) patient with pathological voice.

TABLE I.

Results (%) of the voice pathology classification, ACC-accuracy, HP-precision for healthy cases, PP-precision for pathological cases, HS-specificity for healthy and PS-specificity for patients with pathological voice

ACC	HP	PP	HS	PS
73,33	70,59	76,92	80,00	66,67