

Visual and Automatic Evaluation of Vocal Fold Mucosal Waves Through Sharpness of Lateral Peaks in High-Speed Videokymographic Images

*S. Pravin Kumar, *Ketaki Vasant Phadke, †Jitka Vydrová, ‡Adam Novozámský, ‡Aleš Zita, ‡Barbara Zitová, and *Jan G. Švec, *Olomouc, and †,‡Prague, Czech Republic

Abstract: Introduction. The sharpness of lateral peaks is a visually helpful clinical feature in high-speed videokymographic (VKG) images indicating vertical phase differences and mucosal waves on the vibrating vocal folds and giving insights into the health and pliability of vocal fold mucosa. This study aims at investigating parameters that can be helpful in objectively quantifying the lateral peak sharpness from the VKG images.

Method. Forty-five clinical VKG images with different degrees of sharpness of lateral peaks were independently evaluated visually by three raters. The ratings were compared to parameters obtained by automatic image analysis of the vocal fold contours: *Open Time Percentage Quotients* (OTQ) and *Plateau Quotients* (PQ). The OTQ parameters were derived as fractions of the period during which the vocal fold displacement exceeds a predetermined percentage of the vibratory amplitude. The PQ parameters were derived similarly but as a fraction of the open phase instead of a period.

Results. The best correspondence between the visual ratings and the automatically derived quotients were found for the OTQ and PQ parameters derived at 95% and 80% of the amplitude, named OTQ₉₅, PQ₉₅, OTQ₈₀ and PQ₈₀. Their Spearman's rank correlation coefficients were in the range of 0.73 to 0.77 ($P < 0.001$) indicating strong relationships with the visual ratings. The strengths of these correlations were similar to those found from inter-rater comparisons of visual evaluations of peak sharpness.

Conclusion. The Open time percentage and Plateau quotients at 95% and 80% of the amplitude stood out as the possible candidates for capturing the sharpness of the lateral peaks with their reliability comparable to that of visual ratings.

Keywords: Mucosal waves—Lateral peak sharpness—Kymography—Vocal fold vibration—Image analysis—Quantification.

INTRODUCTION

The occurrence of mucosal waves on the vibrating vocal folds has been generally recognized as a crucial indicator for healthy voice. Mucosal waves originate at the inferior surface of the vocal fold mucosa, propagate vertically along the medial surface, and then horizontally along the superior surface, creating a wave-like motion on the vocal folds.¹⁻⁸ A soft and pliable superficial layer of the lamina propria is necessary for their occurrence.^{1,9} In other words, health and pliability of vocal fold mucosa may be indicated by the presence of mucosal waves.¹⁰ Reduced mucosal wave amplitude

is clinically observed in cases of increased mucosal stiffness due to, eg, lesions or scarring.^{9,11}

Observations on excised hemilarynges^{5,7,12-15} and lately also ultrasonic laryngeal observations *in vivo*¹⁶ have shown that mucosal waves are associated with the phase-delayed movements of the upper vocal fold margin (lip or edge) trailing the lower margin. This delay is termed “vertical phase difference”, and it facilitates the delivery of airflow energy to vocal fold tissue.^{2,3,10,17-19} Titze et al (1993)⁵ stroboscopically tracked the flesh-points in excised larynges to quantify the phase delay and demonstrated its relationship with mucosal wave propagation velocity.

In vivo laryngoscopic imaging techniques such as videostroboscopy and high-speed videoendoscopy (HSV) have enabled easier visualization and quantitative evaluation of the presence, absence, or reduction of mucosal waves in clinical practice.^{1,20-28} An alternative view for clinical evaluation of the mucosal waves has been offered by kymographic (ie, single-line) imaging techniques such as videokymography (VKG), digital kymography (DKG) or strobokymography (SVKG).²⁹

Kymography assesses mucosal waves based on (1) vertical phase differences and (2) laterally traveling mucosal waves.^{10,30,31} Vertical phase differences show up as sharp lateral peaks in kymograms, and laterally running mucosal

Accepted for publication August 30, 2018.

A part of the work was presented at the Voice Foundation's 47th Annual Symposium – Care of the Professional Voice, May 30-June 3, 2018, Philadelphia, USA. At this symposium S.P.K. received the Hamdan International Presenter Award 2018 and Sataloff New Investigator Award 2018 for this work. The research of the authors was supported by Technological Agency of the Czech Republic project no. TA04010877 (2014-2017) and the Czech Science Foundation (GA CR) project no. GA16-01246S (2016-2018). S. P. K. thanks Sri Sivasubramaniya Nadar (SSN) College of Engineering, Kalavakkam, India, for granting the leave of absence during his sabbatical stay (2016-2018) at Palacký University in Olomouc, Czech Republic.

From the *Voice Research Lab, Department of Biophysics, Faculty of Science, Palacký University, Olomouc, Czech Republic; †Voice and Hearing Centre, Medical Healthcom Ltd., Prague, Czech Republic; and the ‡Department of Image Processing, Institute of Information Theory and Automation of the Czech Academy of Sciences, Prague, Czech Republic.

Address correspondence and reprint requests to S. Pravin Kumar or Jan G. Švec, Voice Research Lab, Department of Biophysics, Faculty of Science, Palacký University, 17. listopadu 12, Olomouc 77416, Czech Republic. E-mail: pravin.subbaraj@upol.cz; jan.svec@upol.cz

Journal of Voice, Vol. ■■■, No. ■■■, pp. 1–9
0892-1997

© 2018 The Voice Foundation. Published by Elsevier Inc. All rights reserved.
<https://doi.org/10.1016/j.jvoice.2018.08.022>

waves appear on the kymogram as lines running obliquely sideways along the upper margin during the medial excursion of the vocal fold^{10,30-32} (Figure 1).

The sharpness and roundedness as observed from the shape of the lateral peaks are resulting from the vertical phase differences between the lower and upper margins of the vibrating vocal folds^{30,33} (Figure 2). Looking from above the vocal folds, the boundary between the glottis and the vocal fold is created by the most medial part of the vocal fold. Due to the vertical phase differences, during the opening phase, this boundary is formed by the position of the upper margin of the vocal fold, whereas during the closing phase the boundary is normally formed by the position of the lower margin of the vocal fold. At the point of transition from opening to the closing phase, the glottal edge shifts from the upper to the lower margin (Figure 2A). When the vertical phase differences are large, the shift from upper to lower vocal fold margin is abrupt. In the kymogram, this sudden transition results in a sharp lateral peak within the oscillating vocal fold contour. In smaller vertical phase differences this transition happens rather gradually, causing the lateral peak to be rounded (Figure 2B).

The shape of the lateral peak has been found to be a clinically useful parameter revealing the vocal fold vibration characteristics that are not easily observable in non-kymographic imaging methods.³⁴ It has gained attention due to its diagnostic importance in assessing various voice disorders such as mucosal inflammations, scarring or tumors related to increased mucosal stiffness.^{10,30,31,34-37} Increased vertical thickness of the vocal folds and increased pliability of the mucosa likely lead to larger vertical phase differences producing sharper lateral peaks in kymography. In contrast, increased stiffness of the mucosa is expected to reduce vertical phase differences, thus producing more rounded lateral peak

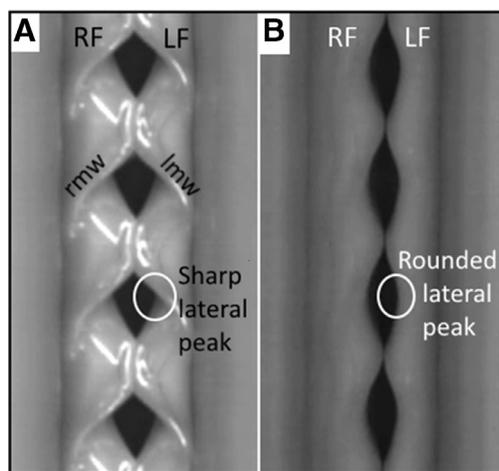


FIGURE 1. Videokymographic images (four vibratory cycles each) showing (A) sharp lateral peaks (encircled) and laterally running mucosal waves (rmw, lmw) on the right and left vocal fold, respectively; (B) rounded lateral peaks (encircled) with no mucosal waves. RF, LF – right and left vocal fold. Total time displayed in the kymograms: 17.6 ms (time direction from top to bottom).

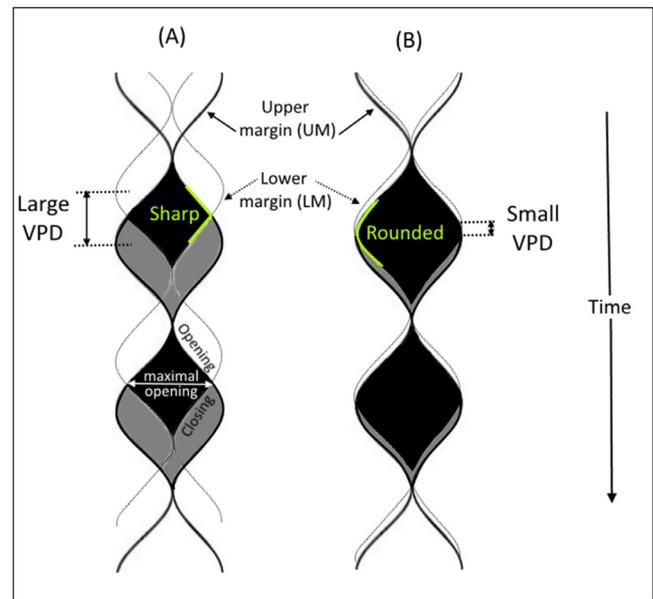


FIGURE 2. Formation of sharp (A) and rounded (B) lateral peaks in the kymogram. Movements of the lower and upper margins of the vocal folds are indicated by thin-dotted and thick-solid curves, respectively. The vibratory displacement of the lower margin precedes that of the upper margin, thus creating a vertical phase difference between their respective motions. During the opening phase, the motion of the lower margin is invisible—this is indicated by the thin-dotted line; it becomes only visible during the closing phase. A sharp lateral peak (A) is seen when the vertical phase difference is large and a rounded lateral peak (B) is seen when the vertical phase difference is small (indicated in green). LM, lower margin; UM, upper margin; VPD, vertical phase difference. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

shapes.^{10,30,32} The magnitude of the vertical phase differences and the related sharpness of lateral peaks can also reveal on vocal fold vibratory behavior in different voice tokens, such as vocal registers.^{33,38}

Efforts have been made to assess vertical phase differences and laterally traveling mucosal waves using image analysis methods. Shaw and Deliyski (2008)³⁹ used mucosal wave playback and qualitatively assessed the variations in mucosal wave magnitude and symmetry. Voigt et al (2010)²³ managed to detect the laterally traveling mucosal waves in high-speed endoscopic videos using automated image analysis techniques. Lately, Andrade-Miranda et al (2017)⁴⁰ used the optical flow method to detect mucosal wave propagation from high-speed endoscopic videos.

Chen, Woo, and Murry⁴¹⁻⁴³ applied spectral analysis to vocal fold waveforms obtained from digital kymograms and reported its usefulness in quantifying the waveforms and their changes due to different vocal tokens, pathologies, and surgical interventions. In principle, the spectral features can be expected to reflect the sharpness of the lateral peaks through

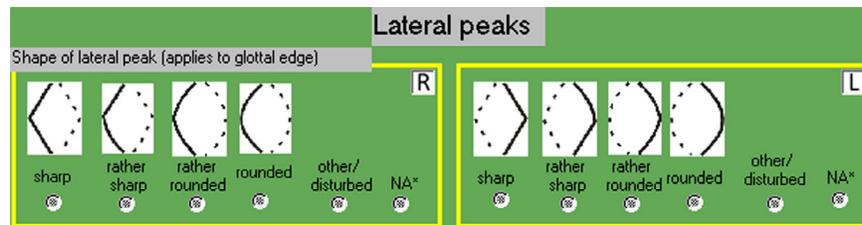


FIGURE 3. The form for visual evaluation of the VKG images with the descriptive pictograms representing varied degrees of lateral peak sharpness in the right (R) and left (L) vocal folds.

increased energy in upper harmonics, but such spectral changes can occur also due to, eg, the occurrence of closed phase; thus the spectral analysis of kymographic waveform makes it difficult to clearly distinguish the sharpness of the lateral peaks from other factors.

According to our knowledge, two methods have tried to quantify the shape of the lateral peak from kymograms so far.^{44,45} Jiang et al estimated the shape of the peak indirectly by quantifying the vertical phase difference from kymographic images using a sinusoidal model approximation.^{44,46-51} While this method is mathematically elegant, it becomes troublesome and difficult to interpret when the vocal fold motion becomes rather complex. The second method by Yamauchi et al (2015) quantified the peak sharpness from digital kymograms by the “lateral peak index”, defined as an angle formed by two lines between the start of open phase and lateral peak, and between the lateral peak and the end of open phase.⁴⁵ However, this index disregards the changes of curvature of the vocal fold waveform that influence the peak sharpness. Its value is additionally influenced also by the closed quotient and the vibratory amplitude, thus making it also sensitive to other factors than vertical phase differences.

Therefore, there is a need to search for other parameters that could help to improve the reliability of visual evaluation of clinical kymographic images of the vocal folds. Due to limited inter- and intrarater reliability, the approach of subjective rating limits the comparability of quantitative parameters on a large set of data. It further prevents the acquisition of reliable standard reference values for clinicians whose treatment decisions are dependent on assessment of such parameters. In contrast, objectification helps to find the accuracy of visual ratings.⁵² The purpose of this study was therefore to investigate parameters which could quantify the lateral peak sharpness seen in the kymographic images and could easily be measured automatically from the detected contours of the vibrating vocal folds.

The work was done in the following steps: (1) A set of clinically obtained videokymographic images was evaluated visually to obtain ratings of the lateral peak sharpness. (2) The same images were subjected to automatic image analysis, in order to detect and compute the contours of the vibrating vocal folds as waveforms. (3) The resulting waveforms were quantified in order to obtain

numerous parameters expected to reflect the lateral peak sharpness. (4) The obtained values of the parameters were compared to the visual ratings from step (1), in order to determine the parameters that show the best correlation with the visual ratings.

METHODS

Dataset

The dataset used in this work consisted of 45 videokymographic (VKG) images retrospectively selected from clinical records of patients examined for voice complaints at the Voice and Hearing Centre, Medical Healthcom, Ltd, Prague. The VKG recordings were obtained with the second generation VKG camera (Kymocam, CYMO, b.v. Groningen, the Netherlands, image rate 7200 lines/s), which was connected to a laryngoscope (Xion Medical, Germany, 10 mm diameter, 90° angle) using a C-mount objective adapter (R. Wolf, Germany, type 85261.272, 27 mm focal length). The larynx was illuminated by a 300 W endoscopic xenon light source (type FX 300 A, Fentex Medical, Germany). The VKG recordings were stored digitally by means of an EndoSTROB video capturing unit (Xion Medical, Germany). The images were extracted from the video records using the recently developed VKG Analyzer software.⁵³ The images were selected so that they demonstrated varied degrees of sharpness of lateral peaks.

Visual rating

Three raters independently evaluated the sharpness of the lateral peaks from the VKG images using a visual form (Figure 3).⁵⁴ The raters used the pictogram descriptions of the sharpness features as a reference for evaluation. The rating was done on a four points rating scale (1-sharp; 2-rather sharp; 3-rather rounded; 4-rounded) for left and right vocal folds separately, thus making a total of 90 ratings per rater from 45 images.

In order to assess the intra-rater reliability, each rater performed the evaluation twice, with a pause of 7-10 days in between. During the second evaluation, the order of the images was changed to minimize the memory effect. The ratings from the two evaluations, for the three raters, were consolidated, and an average (visual average – VA) was obtained. A common consensus (visual consensus – VC)

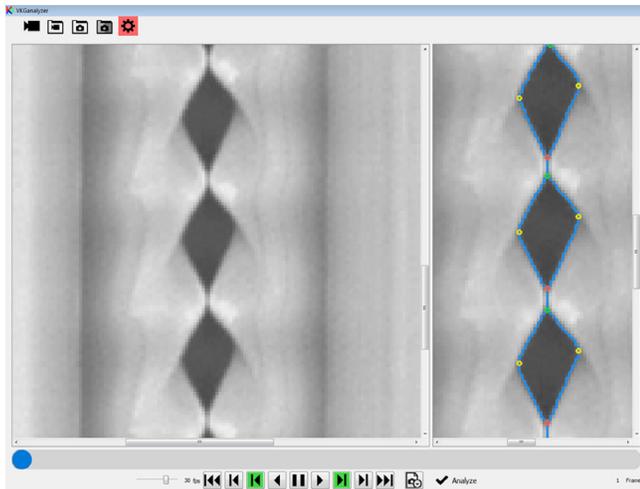


FIGURE 4. The screenshot of the VKG analyzer software showing the VKG image on the left and the detected glottal edge contours on the right.

was also arrived through the discussion among the three raters afterwards.

Image analysis

The recently developed VKG analyzer software⁵³ was used to detect and extract the contours defining the glottal edge boundary of both the left and right vocal folds (Figure 4). The image brightness and contrast were manually adjusted to improve the accuracy of the edge detection whenever required. The contours extracted from each of the VKG images were saved in a text file as a set of data defining the glottal edges of the left and right vocal folds, along with their respective time instances. A custom MATLAB script was then used to process the vocal fold contours and to obtain parameters capturing lateral peak sharpness, which could be included in the VKG analyzer software in future versions.

Quantification of lateral peak sharpness

Two kinds of parameters were defined for their simplicity in quantifying the vocal fold waveforms and their expected capability of reflecting the sharpness of the lateral peaks: the *Open Time Percentage Quotients* (OTQ) and *Plateau Quotients* (PQ).

The *Open Time Percentage Quotients* (OTQ_R) were inspired by the OT50 parameter published by Woo (1996),⁵⁵ who investigated the time for which the glottal area waveform exceeded 50% of the amplitude. Here, we defined the OTQ_R parameter as the proportion of time during which the vocal fold displacement exceeds a chosen percentage (R) of the vibration amplitude within a period (Figure 5):

$$OTQ_R = \frac{D_R}{T}$$

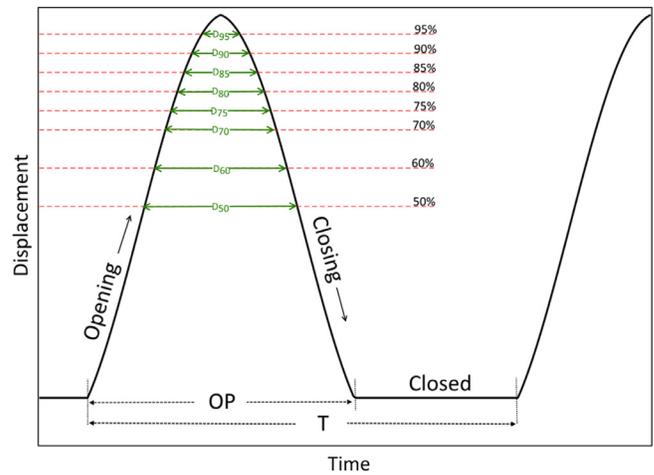


FIGURE 5. Parameterization of the vocal fold waveform for obtaining the Open Time Percentage Quotients (OTQ_R) and the Plateau Quotients (PQ_R) as indicators for peak sharpness. OP is the open phase, T is the period, and D_R is the duration of the phase during which the waveform exceeds a specified R percentage of the amplitude. The R percentages are indicated by the dashed red lines.

where D_R is the duration of the phase where the lateral displacement is greater than $R\%$ of the vibration amplitude and T is the period of the vocal fold vibratory cycle. The vibration amplitude was determined as the difference between the most lateral and most medial position of the vocal fold during the open phase.

The *Plateau Quotients* (PQ_R) used here were inspired by the work of Mehta et al,⁵⁶ who investigated the proportion of open phase for which the glottal area was larger than 95% of its maximum. Here, we defined PQ_R as the proportion of time during which the vocal fold displacement

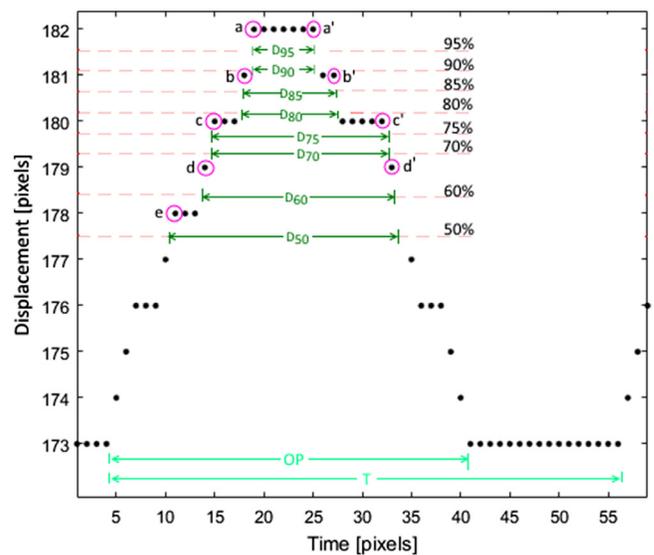


FIGURE 6. Implementation of the parameterization of the waveform illustrating the procedure followed to calculate the D_R durations from the discrete samples.

exceeds R% of vibration amplitude within the open phase (Figure 5):

$$PQ_R = \frac{D_R}{OP}$$

where OP is the duration of the open phase.

When implementing the automatic analysis procedure, it was necessary to deal with the fact that the waveforms were not continuous, but consisted of samples of limited temporal and spatial resolution. While the contour samples are defined by integer pixel coordinates, the R% levels usually correspond to noninteger subpixel coordinates. An example of the procedure adapted to calculate the OTQ and PQ parameters from the discrete samples is shown in Figure 6. When digitized, the discrete contour data points were located at specific pixels with coordinates defined by integer numbers. Therefore, sometimes the same pixel coordinates pertained to multiple consecutive time points (see Figure 6). In order to measure the time intervals at which the vocal fold displacement exceeds the criterion level R%, the first and last samples with the values above the R% criterion level in the opening and closing phases, respectively, were selected (marked by circles and indicated as a, b, c, d, e in the opening phase, and a', b', c', d' in the closing phase in Figure 6). Thus, for the R% levels at 95%, 90%, 85%, 80%, 75%, 70%, 60% and 50%, the intervals between $a-a', a-a', b-b', b-b', c-c', c-c', d-d'$ and $e-d'$, respectively, were considered to calculate the D_R durations in the example shown in Figure 6.

Statistical analysis

Statistical analysis was performed using the SPSS (version 24) software. Spearman's rank correlation coefficient was computed to determine the inter- and intrarater reliability of the visual ratings. The intrarater reliability was also tested with Cronbach's Alpha value. To estimate the correlation between the objective measures and the visual ratings, Spearman's rank correlation coefficient was again used.

RESULTS

Visual rating

Results from the repeated visual evaluations of the lateral peak sharpness in VKG images by the three raters were compared to find the intrarater and inter-rater reliability. The intrarater comparisons between the two repeated evaluations resulted in the Cronbach's Alpha values around 0.92 for all three raters, indicating excellent reliability of the raters. The intrarater Spearman's rank correlation coefficients for the individual raters varied between 0.84 and 0.85 ($P < 0.001, N = 90$) indicating very strong and significant correlations between the repeated evaluations.

The inter-rater comparisons showed Spearman's rank correlation coefficients in the range of 0.67 to 0.82, with a mean value of 0.73. These coefficients indicated strong and significant correlations ($P < 0.001, N = 90$) between the evaluations of the different raters, but also hinted at some discrepancies among the raters. Therefore, a consensus among the raters was established by mutual discussions.

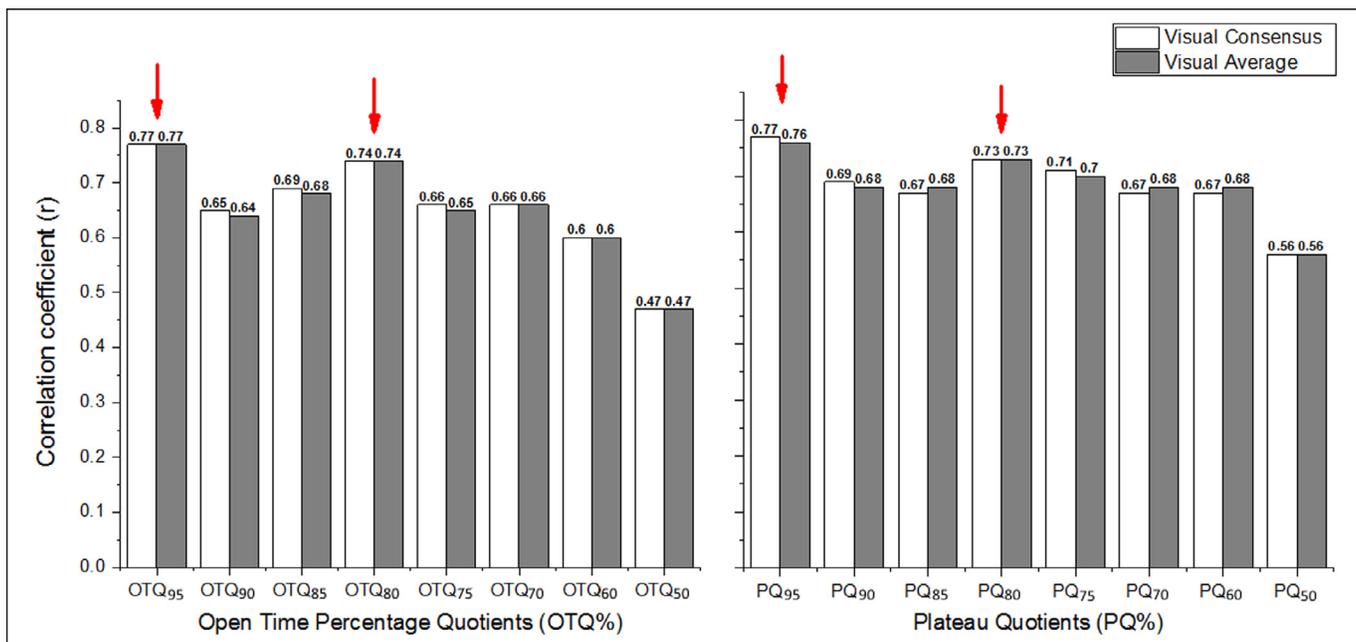


FIGURE 7. Spearman's rank correlation coefficients indicating the agreement between the visual ratings and the measured parameters OTQ and PQ. The highest correlation coefficients are indicated by red arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

The visual consensus versus visual average comparison revealed very strong Spearman's rank correlation ($r = 0.99$, $P < 0.001$). Furthermore, both the visual consensus and visual average values very strongly correlated with the values of all three raters ($r = 0.81-0.91$, $P < 0.001$) in both evaluations. Therefore, the visual consensus and visual average values were deemed appropriate for further analysis of the correlations between the visual and automatic image analysis.

Correlation between visual ratings and the analyzed parameters

The correlations between the different OTQ and PQ parameters with the visual consensus and visual average ratings are shown in Figure 7. All correlations had a significance level of $P < 0.001$, indicating that all parameters were well related to the visual ratings. Highest

correlations were found for the parameters measured at 95% amplitude (OTQ₉₅, PQ₉₅) and at 80% amplitude (OTQ₈₀, PQ₈₀). In Figure 7, these are indicated by arrows. There were minimal differences between the OTQ and PQ parameters measured at the same percentage. Also, there were minimal differences between the visual average and visual consensus. Lowest correlations were found for the parameters measured at 50% amplitude (OTQ₅₀, PQ₅₀).

The relationships between the values of the four best correlating parameters OTQ₉₅, OTQ₈₀, PQ₉₅, and PQ₈₀, and the visual ratings are revealed in Figure 8. As expected, all these quotients clearly increase their values when the peak shape changes from sharp to rounded. There is, however, some spread of the measured data around the best fit line, which indicates that some discrepancies exist between the visual and automatic evaluations. The Spearman's rank correlation values between

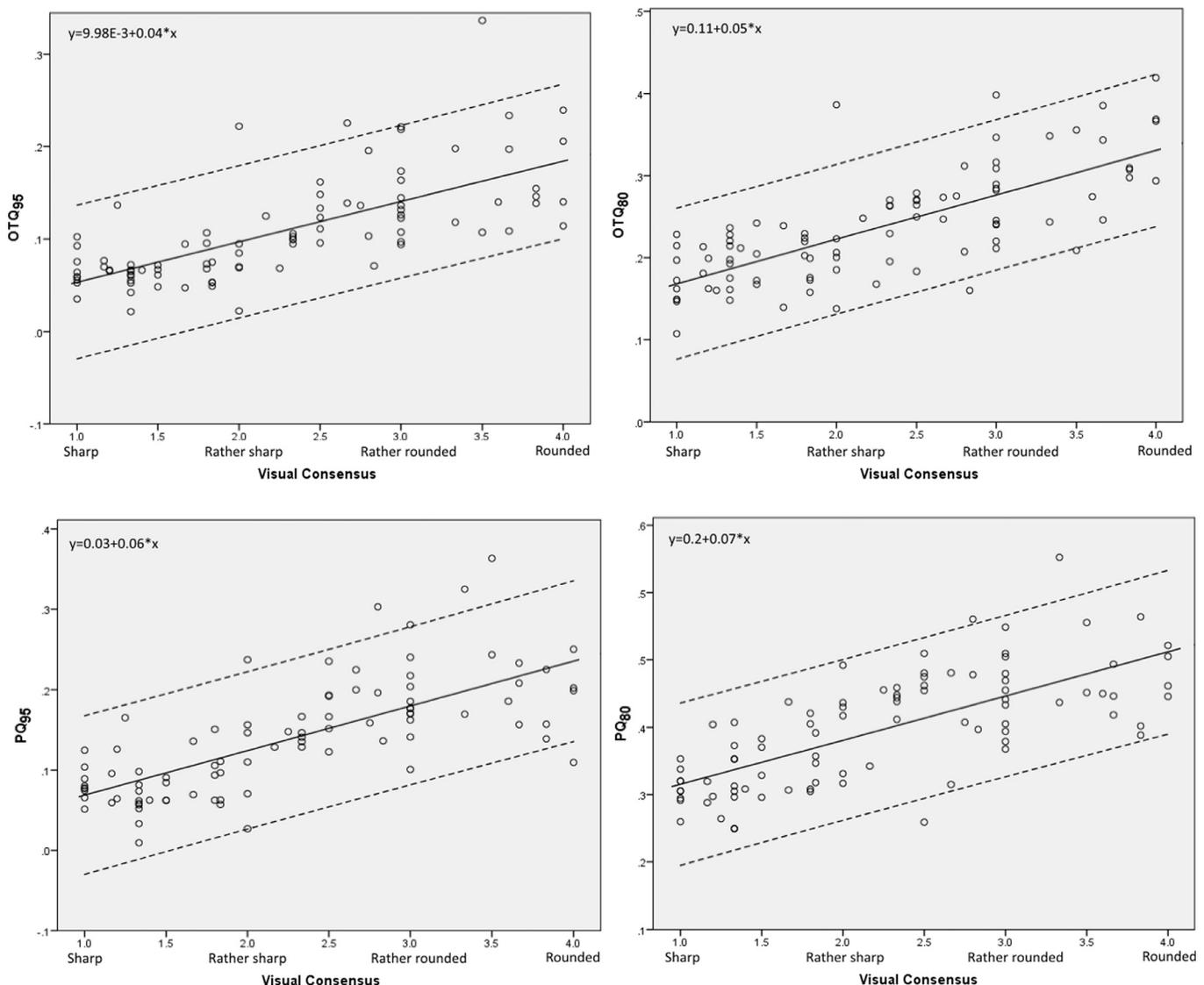


FIGURE 8. The relationship between the measured values and the visual ratings for the four parameters with the highest correlations – OTQ₉₅, OTQ₈₀, PQ₉₅, and PQ₈₀. The lines indicate the best fit linear relationship (solid) and 95% confidence intervals (dashed).

these analyzed quotients and the visual ratings (0.73-0.77, as shown in Figure 7) were comparable to those found between different raters (0.67-0.82) indicating that the discrepancies in the automatic-to-visual comparisons are similar to those found in inter-rater comparisons.

DISCUSSION

Sharpness of lateral peaks has been recognized previously as a useful visual feature that can indicate pliability and health of the vocal fold mucosa.^{30,45} In a recent study, the lateral peak sharpness has been identified as one of the most helpful visual features for clinical evaluation of voice disorders using videokymography.³⁴ The peak sharpness is directly related to vertical phase differences between the motions of the upper and lower margin of the vocal folds and results from projection of the vocal fold motion into the laryngoscopic view from above of the vocal folds.^{10,30,57} Biomechanically, stiffening of the mucosa leads to increased mucosal wave speed⁶ and decreased vertical phase differences, causing the peak to become more rounded.^{30,32,35} Apart from physiological factors related, eg, to pitch increase and voice registration, stiffening of the mucosa is considered to be a direct result of pathological processes on the vocal folds. Therefore, evaluation of peak sharpness can help clinicians to better diagnose the health of the vocal fold mucosa, particularly in phonations produced at comfortable pitch in modal/chest register where the mucosa is expected to be pliable.

Visual evaluation, however, is subjective and differences among evaluations of different raters can be expected. This can be spotted also in our results: while the intraindividual Spearman's rank correlations were very strong ($r=0.84-0.85$), the inter-rater Spearman's rank correlations were lower ($r=-0.67-0.82$) indicating more disagreements between the visual evaluations of different raters than between repeated evaluations of the same rater.

This study searched for objective parameters that are related to the visual ratings of peak sharpness in kymograms and can be used as "peak sharpness indicators". For this purpose, the OTQ and PQ were defined by relating the durations of different phases of the vibratory cycle to each other, applying the same concept as used for the well-established traditional parameters such as the Closed Quotient (CQ), Open Quotient (OQ) or Speed Quotient (SQ).⁵⁸⁻⁶⁰ As such, these parameters are relatively simple to measure. As far as their interpretation is concerned, smaller OTQ and PQ values correspond to sharper lateral peaks of the vocal fold waveform detected in the kymogram (recall Figure 8).

The OTQ and PQ parameters measured from the time intervals at different percentages of vibratory amplitude were compared to the visual ratings of the peak sharpness in order to evaluate the congruence of these two approaches. The OTQ and PQ parameters showed very similar correlations to the visual ratings which indicate that they quantified the visual impressions similarly. The best correlations with

the visual evaluations were found for the OTQ and PQ parameters measured at 95% and 80% of the amplitude, the worst correlations appeared at 50% of the amplitude. Since the peak corresponds to 100% of the amplitude, it appears logical that the best correlations for peak sharpness should be obtained for the measurements made as closely to the peak as possible – this explains the finding of the worst correlations at 50% and best correlations at 95% of the amplitude (recall Figure 7). However, the correlations at 90% and 85% of the amplitude were worse than those at 80%. This seemingly contradictory finding could be attributed to the contour artifacts due to the limited pixel and temporal resolution (compare the ideal waveform in Figure 5 with the real detected waveform in Figure 6). The clinical videokymographic images analyzed here showed the average vocal fold vibratory amplitudes around 8 pixels (range 5-15 pixels). A change of 1 pixel, in this case, corresponds to the spatial resolution of 12.5% of the amplitude (range 7-20%). This means that it is hardly possible to reliably distinguish levels that are close together, such as those at 85%, 90%, and 95% of the amplitude.

Preliminary investigations using synthetic kymograms generated by a kinematic model of the vocal folds⁶¹ with known vertical phase differences (not included here for brevity reasons) showed that the limited spatial and temporal resolution of the kymographic images can influence the accuracy of the results, particularly of those quotients measured at the proximity of the peak, and these artifacts need to be taken into account. Thus the measurements at 80% amplitude could potentially be used as a compromise to reduce the influence of the possible waveform artifacts, but still reflect the peak sharpness and vertical phase differences reasonably well. The waveform artifacts present a general limitation which is inherent in the laryngeal kymographic techniques. Increased spatial resolution of the kymographic images is desirable for improving the quantification accuracy of the vocal fold vibratory patterns in future.

In principle, the OTQ and PQ parameters can be implemented also for analyzing the glottal area waveforms (GAWs) obtained from full high-speed endoscopic videos, as done by Mehta et al (2011).⁵⁶ GAWs offer better pixel resolution than kymography due to the fact that the glottal area is distributed over multiple image lines and thus over considerably more pixels. In this respect, GAWs may possibly offer better accuracy than kymographic waveforms in measuring the OTQ and PQ parameters as defined here. However, a more detailed study is needed to elucidate these factors and to better understand the influence of limited spatial and temporal resolution on the accuracy of these parameters.

The detailed comparisons between the visual ratings and the OTQ and PQ parameters shown in Figure 8 reveal that the relationship is not perfect and some discrepancies exist here. Besides of the influence of the limited spatial and temporal resolution of the images (7200 kymographic lines per second with 720 pixels per line used here), these discrepancies could possibly be also due to contour detection artifacts

resulting from the image analysis procedure. Furthermore, it is known that the visual perception process is rather complex and visual judgments of the peak shape may also be influenced by, eg, the grayscale shadings which are not captured in the contours. All these factors may contribute to the differences between the automatic analysis and the visual ratings. Nevertheless, the Spearman's rank correlations between the visual ratings and the OTQ and PQ parameters measured at 80% and 95% amplitude ($r = 0.73-0.77$, recall Figure 7) are similar to those found between different raters. Therefore the reliability of the parameters, although not perfect, is considered acceptable here.

While the shape of the lateral peak appears as a useful clinical feature, ultimately it should be related to the vertical phase differences. These differences cannot be exactly measured laryngoscopically *in vivo*. Therefore, we were not able to establish their direct relationship with the defined parameters, which poses another potential limitation of this study. However, this relationship may be derived and investigated using synthetic kymograms obtained from a mathematical model of the vocal folds with known vertical phase differences⁶¹, which is planned to be addressed in a future study.

CONCLUSION

The PQ₉₅, PQ₈₀, OTQ₉₅ and OTQ₈₀ parameters stood out as the possible candidates for capturing the sharpness of the lateral peaks. The reliability of these parameters appears comparable to the inter-individual reliability of visual ratings. The results provide basic insights into developing the computer algorithms to automatically quantify the sharpness of lateral peaks from the VKG images.

REFERENCES

- Hirano M. *Clinical Examination of Voice*. Wien, Austria: Springer-Verlag; 1981.
- Titze IR. The physics of small-amplitude oscillation of the vocal folds. *J Acoust Soc Am*. 1988;83:1536-1552.
- McGowan R. An analogy between the mucosal waves of the vocal folds and wind waves on water. *Haskins Lab Status Rep Speech Res*. 1990;101:243-249.
- Yumoto E, Kurokawa H, Okamura H. Vocal fold vibration of the canine larynx: observation from an infraglottic view. *J Voice*. 1991;5:299-303.
- Titze IR, Jiang JJ, Hsiao T-Y. Measurement of mucosal wave propagation and vertical phase difference in vocal fold vibration. *Ann Otol Rhinol Laryngol*. 1993;102:58-63.
- Berke GS, Gerratt BR. Laryngeal biomechanics: an overview of mucosal wave mechanics. *J Voice*. 1993;7:123-128.
- Boessenecker A, Berry DA, Lohscheller J, et al. Mucosal wave properties of a human vocal fold. *Acta Acust united Ac*. 2007;93:815-823.
- Krausert CR, Olszewski AE, Taylor LN, et al. Mucosal wave measurement and visualization techniques. *J Voice*. 2011;25:395-405.
- Hirano M, Bless DM. *Videostroboscopic Examination of the Larynx*. San Diego, California: Singular Publishing Group; 1993.
- Švec JG, Šram F, Schutte HK. Videokymography. In: Fried M, Ferlito A, eds. 3 ed *The Larynx*. Vol 1, San Diego, CA: Plural Publishing; 2009:253-271.
- Bless DM, Hirano M, Feder RJ. Videostroboscopic evaluation of the larynx. *Ear Nose Throat J*. 1987;66:289-296.
- Hiroto I. Vibration of vocal cords: an ultra high-speed cinematographic study (film). Kurume, Japan: Department of otolaryngology, Kurume University; 1968.
- Berry DA, Montequin DW, Tayama N. High-speed digital imaging of the medial surface of the vocal folds. *J Acoust Soc Am*. 2001;110:2539-2547.
- Döllinger M, Berry DA, Kniesburges S. Dynamic vocal fold parameters with changing adduction in ex-vivo hemilarynx experiments. *J Acoust Soc Am*. 2016;139:2372-2385.
- Herbst CT, Hampala V, Garcia M, et al. Hemi-laryngeal setup for studying vocal fold vibration in three dimensions. *J Vis Exp*. 2017; 129:e55303. <http://dx.doi.org/10.3791/55303>.
- Jing B, Ge Z, Wu L, et al. Visualizing the mechanical wave of vocal fold tissue during phonation using electroglottogram-triggered ultrasonography. *J Acoust Soc Am*. 2018;143:EL425-EL429.
- Ishizaka K, Flanagan J. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Syst Tech J*. 1972;51:1233-1268.
- Titze IR. Comments on the myoelastic-aerodynamic theory of phonation. *J Speech Hear Res*. 1980;23:495-510.
- Titze IR. *Principles of Voice Production (Second Printing)*. Iowa City, IA: National Center for Voice and Speech; 2000.
- Dejonckere PH, Bradley P, Clemente P, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *Eur Arch Otorhinolaryngol*. 2001;258:77-82.
- Poburka BJ. A new stroboscopy rating form. *J Voice*. 1999;13:403-413.
- Deliyiski DD, Petrushev PP, Bonilha HS, et al. Clinical implementation of laryngeal high-speed videoendoscopy: challenges and evolution. *Folia Phoniatr Logop*. 2008;60:33-44.
- Voigt D, Döllinger M, Eysholdt U, et al. Objective detection and quantification of mucosal wave propagation. *J Acoust Soc Am*. 2010;128:EL347-EL353.
- Kaneko M, Shiromoto O, Fujiu-Kurachi M, et al. Optimal duration for voice rest after vocal fold surgery: randomized controlled clinical study. *J Voice*. 2017;31:97-103.
- Poburka BJ, Patel RR, Bless DM. Voice-vibratory assessment with laryngeal imaging (VALI) form: reliability of rating stroboscopy and high-speed videoendoscopy. *J Voice*. 2017;31:513e1-513.e14.
- El-Demerdash A, Fawaz SA, Sabri SM, et al. Sensitivity and specificity of stroboscopy in preoperative differentiation of dysplasia from early invasive glottic carcinoma. *Eur Arch Otorhinolaryngol*. 2015;272:1189-1193.
- Zacharias SRC, Deliyiski DD, Gerlach TT. Utility of laryngeal high-speed videoendoscopy in clinical voice assessment. *J Voice*. 2018;32:216-220.
- Patel RR, Awan SN, Barkmeier-Kraemer J, et al. Recommended minimum protocols for instrumental assessment of voice: American Speech-Language Hearing Association Committee on Instrumental Voice assessment protocols. *Am J Speech Lang Pathol*. 2018;27:887-905.
- Švec JG, Schutte HK. Kymographic imaging of laryngeal vibrations. *Curr Opin Otolaryngol Head Neck Surg*. 2012;20:458-465.
- Švec JG, Šram F, Schutte HK. Videokymography in voice disorders: what to look for. *Ann Otol Rhinol Laryngol*. 2007;116:172-180.
- Švec JG, Frič M, Šram F, Schutte HK. Mucosal waves on the vocal folds: conceptualization based on videokymography. *Fifth International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*. Firenze, Italy: Firenze University Press; 2007:171-172.
- Švec JG, Šram F. Videokymographic examination of voice. In: Ma EPM, Yiu EML, eds. *Handbook of Voice Assessments*. San Diego, CA: Plural Publishing; 2011:129-146.
- Sundberg J, Högset C. Voice source differences between falsetto and modal registers in counter tenors, tenors and baritones. *Logoped Phoniatr Vocol*. 2001;26:26-36.

34. Phadke KV, Vydrová J, Domagalská R, et al. Evaluation of clinical value of videokymography for diagnosis and treatment of voice disorders. *Eur Arch Otorhinolaryngol*. 2017;274:3941–3949.
35. Vydrová J, Švec JG, Šram F. Videokymography (VKG) in laryngologic practice. *J Macrotrends Health Med*. 2015;3:87–95.
36. Yamauchi A, Yokonishi H, Imagawa H, et al. Quantification of vocal fold vibration in various laryngeal disorders using high-speed digital imaging. *J Voice*. 2016;30:205–214.
37. Yamauchi A, Yokonishi H, Imagawa H, et al. Visualization and estimation of vibratory disturbance in vocal fold scar using high-speed digital imaging. *J Voice*. 2016;30:493–500.
38. Švec JG, Sundberg J, Hertegard S. Three registers in an untrained female singer analyzed by videokymography, strobolaryngoscopy and sound spectrography. *J Acoust Soc Am*. 2008;123:347–353.
39. Shaw HS, Deliyski DD. Mucosal wave: a normophonic study across visualization techniques. *J Voice*. 2008;22:23–33.
40. Andrade-Miranda G, Bernardoni NH, Godino-Llorente JJ. Synthesizing the motion of the vocal folds using optical flow based techniques. *Biomed Signal Process Control*. 2017;34:25–35.
41. Chen W, Woo P, Murry T. Spectral analysis of digital kymography in normal adult vocal fold vibration. *J Voice*. 2014;28:356–361.
42. Chen W, Woo P, Murry T. Vocal fold vibratory changes following surgical intervention. *J Voice*. 2016;30:224–227.
43. Chen W, Woo P, Murry T. Vocal fold vibration following surgical intervention in three vocal pathologies: a preliminary study. *J Voice*. 2017;31:610–614.
44. Jiang JJ, Chang CIB, Raviv JR, et al. Quantitative study of mucosal wave via videokymography in canine larynges. *Laryngoscope*. 2000;110:1567–1573.
45. Yamauchi A, Yokonishi H, Imagawa H, et al. Quantitative analysis of digital videokymography: a preliminary study on age- and gender-related difference of vocal fold vibration in normal speakers. *J Voice*. 2015;29:109–119.
46. Jiang JJ, Zhang Y, Kelly MP, et al. An automatic method to quantify mucosal waves via videokymography. *Laryngoscope*. 2008;118:1504–1510.
47. Chodara AM, Krausert CR, Jiang JJ. Kymographic characterization of vibration in human vocal folds with nodules and polyps. *Laryngoscope*. 2012;122:58–65.
48. Krausert CR, Ying D, Zhang Y, et al. Quantitative study of vibrational symmetry of injured vocal folds via digital kymography in excised canine larynges. *J Speech Lang Hear Res*. 2011;54:1022–1038.
49. Li L, Zhang Y, Maytag AL, et al. Quantitative study for the surface dehydration of vocal folds based on high-speed imaging. *J Voice*. 2015;29:403–409.
50. Regner MF, Robitaille MJ, Jiang JJ. Interspecies comparison of mucosal wave properties using high-speed digital imaging. *Laryngoscope*. 2010;120:1188–1194.
51. Zhang Y, Huang N, Calawerts W, et al. Quantifying the subharmonic mucosal wave in excised larynges via digital kymography. *J Voice*. 2017;31:123.e7–123.e13.
52. Bonilha HS, Deliyski DD, Gerlach TT. Phase asymmetries in normophonic speakers: visual judgments and objective findings. *Am J Speech Lang Pathol*. 2008;17:367–376.
53. Novozamsky A, Sedlar J, Zita A, et al. Image analysis of videokymographic data. *2015 IEEE International Conference on Image Processing (ICIP)*, 2015,78–82.
54. Švec JG, Švecová H, Herbst C, et al. Evaluation protocol for videokymographic images. (Ms Access software application). Groningen, the Netherlands: Groningen Voice Research Lab, University of Groningen. 2007.
55. Woo P. Quantification of videostrobolaryngoscopic findings—measurements of the normal glottal cycle. *Laryngoscope*. 1996;106:1–27.
56. Mehta DD, Zaňartu M, Quatieri TF, et al. Investigating acoustic correlates of human vocal fold vibratory phase asymmetry through modeling and laryngeal high-speed videoendoscopy. *J Speech Lang Hear Res*. 2011;130:3999–4009.
57. Hiroto I. The mechanism of phonation; its pathophysiological aspects. *Nippon Jibiinkoka Gakkai Kaiho*. 1966;69:2097–2106.
58. Timcke R, von Leden H, Moore P. Laryngeal vibrations - measurements of the glottic wave .I. The normal vibratory cycle. *AMA Arch Otolaryngol*. 1958;68:1–19.
59. Qiu Q, Schutte H, Gu L, et al. An automatic method to quantify the vibration properties of human vocal folds via videokymography. *Folia Phoniatr Logop*. 2003;55:128–136.
60. Lohscheller J, Švec JG, Döllinger M. Vocal fold vibration amplitude, open quotient, speed quotient and their variability along glottal length: kymographic data from normal subjects. *Logoped Phoniatr Vocol*. 2013;38:182–192.
61. Subbaraj PK, Švec JG. Kinematic model for simulating mucosal wave phenomena on vocal folds. In: Manfredi C, ed. *MAVEBA 2017: Models and Analysis of Vocal Emissions for Biomedical Applications. 10th International Workshop*. Firenze: Firenze University Press; 2017:115–118.