INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Aerodynamics and lumped-masses combined with delay lines for modeling vertical and anterior-posterior phase differences in pathological vocal fold vibration

Carlo Drioli[1], *Philipp Aichinger*[2]

[1] Department of Mathematics, Computer Science and Physics, University of Udine, Italy.
[2] Medical University of Vienna, Department of Otorhinolaryngology, Division of Phoniatrics-Logopedics, Vienna, Austria.

carlo.drioli@uniud.it, philipp.aichinger@meduniwien.ac.at

Abstract

We discuss the representation of anterior-posterior (A-P) phase differences in vocal cord oscillations through a numerical biomechanical model involving lumped elements as well as distributed elements, i.e., delay lines. A dynamic glottal source model is illustrated in which the fold displacement along the vertical and the longitudinal dimensions is explicitly modeled by numerical waveguide components representing the propagation on the fold cover tissue. In contrast to other models of the same class, in which the reproduction of longitudinal phase differences are intrinsically impossible (e.g., in two-mass models) or not easy to control explicitely (e.g., in 3D 16-mass and multi-mass models in general), the one proposed here provides direct control over the amount of phase delay between folds oscillations at the posterior and anterior side of the glottis, while keeping the dynamic model simple and computationally efficient. The model is assessed by addressing the reproduction of typical oscillatory patterns observed in high-speed videoendoscopic data, in which A-P phase differences are observed. Experimental results are provided which demonstrate the ability of the approach to effectively reproduce different oscillatory patterns of the vocal folds.

Index Terms: High-speed video analysis, vocal folds dynamical modeling, voice quality characterization, voice disorders.

1. Introduction

Voice quality characterization is pivotal to the clinical care and science of voice disorders, because it aids the indication, selection, evaluation and optimization of medical treatment techniques. These techniques include voice therapy conducted by speech-language pathologist or logopedists, or phonosurgery, which includes surgery that aims at improving voice quality. A lack of understanding the relation between voice quality and vocal fold vibration patterns is observed. Existing models of the relations between different types of voice quality, glottal area waveform (GAW) patterns and auditory percepts are neither totally reliable nor exhaustive.

The dimensions of auditory voice quality, which were investigated in the past, include the degree of hoarseness, roughness, breathiness, pressedness, diplophonia, fry, tremor, creakiness, and harshness. Correlations of the degree of hoarseness with closure incompleteness, asymmetry, frequency differences, and closure inconsistency were found [1]. Roughness was found to be related with irregularity of glottal closure and periodicity differences [2], asymmetry, closure incompleteness, closure inconsistency, frequency difference [1], presence of mucus, left-right phase differences, and short-term frequency and amplitude modulation [3]. Breathiness and pressedness were often investigated in the context of a breathiness - normal - pressed continuum. Breathiness was found to be associated with closure incompleteness [4, 1, 3, 5], with higher values of the open quotient [6, 7, 8, 9, 10], with average glottal area in paralysed patients [11], with cycle shape differences in terms of Nyquist plot patterns [7], with GAW skewness [6], irregularity [3], increased ratio of cyclic GAW minima to maxima [12], perturbation of the GAW [7], zipperlike opening and closing of the vocal folds (”zipperlikeness”) [13], asymmetry, closure inconsistency, amplitude differences, and mucosal wave differences [1], as well as with phonatory onset latency and a decreased speed of adduction at voice onset [5]. In contrast, pressedness was found to be associated with decreases of left-right and anterior-posterior phase differences [14], decreased open quotients [7, 10, 8], completeness of glottal closure [5], a decreased ratio of cyclic GAW minima to maxima [15], with GAW cycle shape [16], with decreased ”zipperlikeness” [13], as well as with phonatory onset latency and speed of adduction [5]. Diplophonia was found to be related to irregular glottal closure and periodicity differences [2], frequency bimodality of vocal fold vibration [17], a repetitive pattern of glottal closures [18], bicyclicity [19], as well as with metacycles, phase differences, and closure incompleteness [20]. Fry was found to be related to metacycles, i.e., with 2 or 3 pulses followed by a long closed phase [2], or 2 small openings and closings of the vocal folds in rapid succession followed by a relatively long closed phase [20]. In a case study, tremor was found to coincide with slow variations in GAW peak amplitude [18]. Many of these GAW patterns are related to A-P phase differences, which motivates our modelling approach.

Among the different techniques available today for voice analysis and diagnosis, video data acquisition and processing is recognized as an essential tool for voice quality assessment and medical diagnosis. Recent research discussing connections between biomechanical modeling of the folds and high-speed videoendoscopic or videokymographic techniques can be found in [21, 22, 23]. Voice source analysis through numerical models of the vocal folds oscillatory patterns is nowday a mature research field, and reliable glottal models of different accuracy and complexity are available that mimic the underlying dynamics of the folds [24, 25, 26, 27, 28]. These models were originally intended to provide waveforms of phonatory acoustic emission. Validation via comparing them with high-speed video recordings provides an intermediate means of a reality check of these models.

In general, biomechanical models of the folds based on lumped elements (masses, springs and dampers) cannot repro-

http://dx.doi.org/10.21437/Interspeech.2019-2338

duce longitudinal phase differences (as the two-mass model) or do not offer the possibility of controlling the phase difference in an easy and direct way (as, for example, in the 3D 16-mass and multi-mass models in general, in which phase difference is fixed or depends on a large number of parameters). Conversely, we propose here an approach to fold edge modeling that allows direct control over the amount of phase delay between folds oscillations at the posterior and anterior side of the glottis. This edge displacement model is combined with a low-dimensional lumped-element scheme, which allows to keep the dynamic model simple and computationally efficient.

The proposed model provides in principle a tool for both the simulation of different phonation modalities involving longitudinal phase differences, and the analysis and interpretation of videoendoscopic data for voice quality classification and medical diagnosis.

2. Biomechanical Model

The biomechanical vocal folds model is a body-cover model in which the posterior inferior edge of each fold is represented by a single mass-spring system with stiffness k, damping r and mass m. Possible L-R asymmetry is taken into account by including two different single-mass systems, one for each fold (superscripts l and r will be used to indicate respectively left and right fold, α to indicate either left or right). The driving pressure P_m, and the resulting force F_m acting on the folds, are computed from the flow U_g and the inferior glottal area Ag_i using Newton's and Bernoulli's laws:

$$\begin{cases} m^\alpha \ddot{x}^\alpha(t) + r^\alpha \dot{x}^\alpha(t) + k^\alpha x^\alpha(t) = F_m^\alpha(t) \\ F_m^\alpha(t) = P_m(t) \cdot S_m^\alpha \\ P_m(t) = P_l - \frac{1}{2}\rho \frac{U_g(t)^2}{Ag_i(t)^2} \end{cases} \quad (1)$$

for $\alpha \in \{l, r\}$, and where S_m is the equivalent fold surface on which the pressure is exerted. The orientation of force F_m^l is perpendicular to S_m^l and oriented to the left, whilst F_m^r is perpendicular to S_m^r and oriented to the right.

The vertical phase difference of the vibration of the cord edges, which is essential for the modeling of self-sustained oscillations, is represented by a distributed element introducing a delay of the displacement induced by its propagation on the cover of the fold along the vertical axis. The propagation of the displacement along the thickness of the folds (sagittal axis) can be thus represented by a propagation line introducing a delay τ_{sag}.

Inspired by the modeling approach in [26], and coherently with previous research addrssing glottal modeling by means of low-dimensional, lumped-element schemes [29], we locate the source of the displacement at a given point of the fold surface, and propagate the motion along the surface. Let $x^{i\text{-}p}$ be the posterior displacement of the fold at the entrance of the glottis (inferior, posterior edge), and $x^{s\text{-}p}$ the displacement at the exit (superior, posterior edge). Given an impact model f_X that reproduces the distortions on the folds displacement due to the mutual impact, and adds the offsets $x_0^l = -x_0$ and $x_0^r = x_0$ representing the resting position of the left and right fold respectively[1], their displacement can be computed as:

$$\begin{cases} x^{i\text{-}p,\alpha}(t) = f_X(x^\alpha(t), x_0^\alpha) \\ \qquad\qquad = \begin{cases} x^\alpha(t) + x_0^\alpha & \text{if } x^l(t) < x^r(t) + 2x_0 \\ (x^l(t) + x^r(t))/2 & \text{otherwise} \end{cases} \\ x^{s\text{-}p,\alpha}(t) = f_X(x^\alpha(t - \tau_{sag}), x_0^\alpha) \end{cases} \quad (2)$$

The longitudinal A-P phase difference of the displacement is also modeled here by a distributed element introducing a delay of the displacement, from the posterior end (P) to the anterior end (A), along the horizontal (longitudinal) axis. The propagation delay value τ_{lon} is selected according to the desired A-P phase difference. We will allow different L-R longitudinal phase differences (i.e., $\tau_{lon}^l \neq \tau_{lon}^r$) to model the asymmetric behavior observed in recorded high-speed video data. The displacements of the anterior edges are named $x^{i\text{-}a}$ and $x^{s\text{-}a}$, and are computed by:

$$\begin{cases} x^{i\text{-}a,\alpha}(t) = x^{i\text{-}p,\alpha}(t - \tau_{lon}^\alpha) \\ x^{s\text{-}a,\alpha}(t) = x^{s\text{-}p,\alpha}(t - \tau_{lon}^\alpha) \end{cases} \quad (3)$$

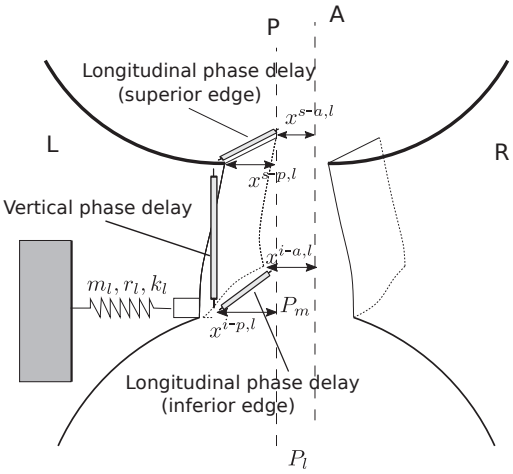

Figure 1: *Schematic view of the model: the vertical (inferior-superior) and longitudinal (anterior-posterior) phase differences of the fold displacement are modeled through two propagation lines. From bottom to top, P_l is the lung pressure, P_m is the driving pressure acting on the vocal folds; m, k, and r represent respectively the mass, stiffness, and damping of the fold, x variables indicate displacements at inferior/superior edges, at anterior and posterior commissures.*

Finally, a flow model converts the glottal area given by the fold displacements into the airflow at the entrance of the vocal tract. If L is the length of the glottis, the A-P axis is sliced in N_L sections, each one of length $\delta_L = L/N_L$. Due to projection, the glottal area at slice j is computed as the minimum cross-sectional area between the area a_j^i at inferior vocal fold edge and the area a_j^s at the superior vocal fold edge. The total glottal area is then computed as the sum of all areas along the A-P dimension, and the flow is assumed proportional to the total glottal area:

[1] Displacements are considered negative if on the left of the medial axis, positive otherwise.

$$\begin{cases} x_j^{i,\alpha}(t) = x^{i\text{-}p}(t - \frac{j\tau_{lon}^{\alpha}}{N_L}), j = 1 \dots N_L \\ x_j^{s,\alpha}(t) = x^{s\text{-}p}(t - \frac{j\tau_{lon}}{N_L}), j = 1 \dots N_L \\ a_j^i(t) = \delta_L(x_j^{i,l}(t) + x_j^{i,r}(t)) \\ a_j^s(t) = \delta_L(x_j^{s,l}(t) + x_j^{s,r}(t)) \\ U_g(t) = \sqrt{\frac{2P_l}{\rho}} \sum_{j=1}^{N_L} \min\{a_j^i(t), a_j^s(t)\} \end{cases} \quad (4)$$

where ρ is the air density, and P_l is the lung pressure.

The discretization of the equations (1)-(4) leads to a discrete-time system that is numerically solved to obtain an estimate of the glottal flow $U_g(nT_s)$ and of the folds displacements $x_j^{i,\alpha}(nT_s)$ and $x_j^{s,\alpha}(nT_s)$, at discrete time n, with T_s being the sampling interval and j being the slice index.

The model simulations are operated at an audio sampling frequency $F_s = (1/T_s) = 22050$ in our experiments, and the oscillatory patterns are visualized as if the folds were observed from above. In this way, the oscillation patterns can be visually compared to high-speed video data.

In this preliminary investigation, we aimed at demostrating that some common oscillatory patterns of the folds observed in the video data can be qualitatively reproduced by the model, without attempting at obtaining an accurate superposition, i.e. we do not aim to achieve copy-synthesis at this time. Thus, no precise period synchronization was pursued, and an approximate match of the period estimated from the video data was obtained by tuning the m and k parameters of the folds manually by trial and error (note that the oscillation frequency of the mass-spring system in isolation is $f_0 = 1/2\pi\sqrt{k/m}$, but often the closed-loop system fundamental frequency resulting by the interaction with the other elements of the model is slightly different).

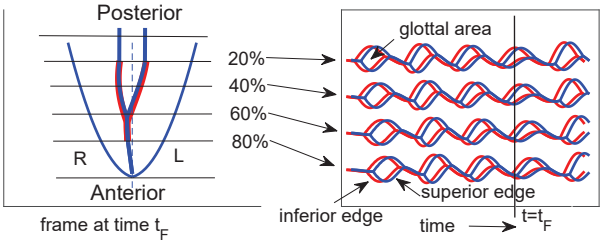

Figure 2: *Example of oscillatory pattern obtained by L-R mass unbalancing and symmetric A-P phase delay. Left panel: instantaneous frame representing the glottis as seen from above; Right panel: oscillatory patterns corresponding to time evolution of the folds displacement at four equally spaced locations along the posteroanterior axis. The vertical line in the right panel indicates the time corresponding to the instantaneous frame in the left panel. A time shift of the open intervals in the glottal area waveform, due to the progressive opening, can be observed.*

The direct control of longitudinal propagation allows to obtain complex vibratory patterns featuring asymmetric oscillation and arbitrary A-P phase differences. Figure 2 illustrates the simulation from a configuration in which the oscillating frequency of the left and right fold are unbalanced and set respectively to $f_0^l = 210$ Hz, $f_0^r = 180$ Hz, and the two longitudinal phase delays are equally set to $\tau_{lon}^l = \tau_{lon}^r = 1.8$ msec. Given that the resulting glottal waveform period is approximately 5 msec, the A-P phase difference is approximately 130 degree. The resulting oscillation is characterized by a zipper-like opening pattern due to the longitudinal phase delay, with paramedian collision due to the mass unbalancing of the folds.

3. Imitation of patterns observed in high-speed videoendoscopic data

The ability of the model to reproduce A-P phase differences characterizing several non−modal vibratory patterns data, is investigated. A selection of five recordings from the Laryngeal High-Speed Video Database of Pathological and Non-Pathological Voices described in [30] is used, in which several examples of patterns with A-P phase differences were observed: in sample S1, double pulsing with L-R symmetry is observed due to an A-P phase difference of approximately 180 degrees, resulting in a voice with a subharmonic spectral component; in sample S2, a time−invariant A-P phase difference is observed, as well as an L-R vocal folds unbalancing; sample S3 has a time-invariant L-R phase difference (fold mass unbalancing); in sample S4, a zipper-like motion due to a small A-P phase difference of approximately 20 degrees is observed; finally in sample S5, a time-invariant A-P phase difference is observed only on the left vocal fold (on the right in the video).

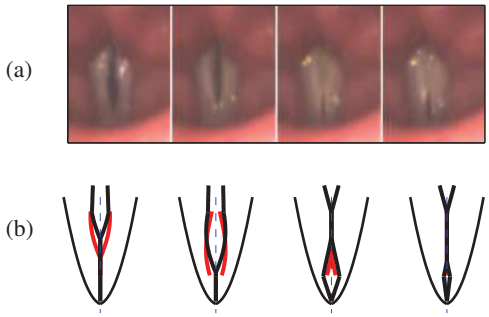

Figure 3: *Selection of frames within a vibratory cycle observed in HSV S1 (a), and a selection of frames from the model simulation (b). Double pulsing is observed due to an A-P phase difference of approximately 180 degrees. The two frames on the left show the main opening of the glottis, associated with a large glottal pulse, during which the outmost anterior part of the glottis is closed, while the rest is open. Conversely, the two frames on the right show the smaller of the two pulses, during which only the outmost anterior part of the glottis is open. The model simulation captures these observations qualitatively, involving a large A-P phase difference of 180 degrees.*

Figures 3 to 6 show selected frames of the five recordings on the top, and model simulations on the bottom. For each sample four frames are shown, which are representative landmark frames obtained from one cycle of vibration. Thus, the four frames repeat in a cyclic way, but only one cycle is shown here.

The simulation shown in Figure 3 aims at reproducing the double pulsing with L-R symmetry. The left and right fold oscillating frequency were set both to $f_0^l = f_0^r = 210$ Hz, and the two longitudinal phase delays were equally set to $\tau_{lon}^l = \tau_{lon}^r = 2.5$ msec, resulting in an A-P phase delay of approximately 180 degree. The resulting oscillation is characterized by two pulses within one period, during which only one out of the outmost anterior part or posterior part of the glottis is open.

In Figure 4, in both the recorded and the simulated panel,

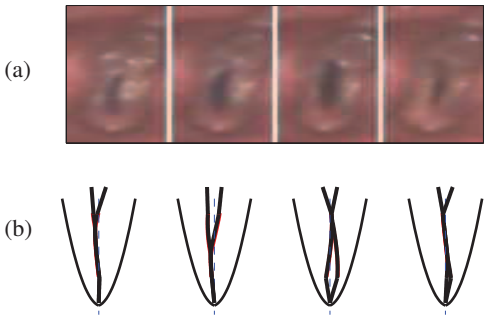

Figure 4: *Selection of frames within one cycle from HSV S2 (a), and a selection from the model simulation (b). A complicated vibration pattern is observed due to L-R mechanical differences, as well as A-P phase differences.*

a complicated vibration pattern due to L-R mechanical differences, as well as A-P phase differences, is observed. The left and right fold oscillating frequency were set to respectively $f_0^l = 210$ Hz, and $f_0^l = 180$ Hz, and the two longitudinal phase delays were set to $\tau_{lon}^l = \tau_{lon}^r = 1.8$ msec, resulting in a A-P phase delay of approximately 130 degrees.

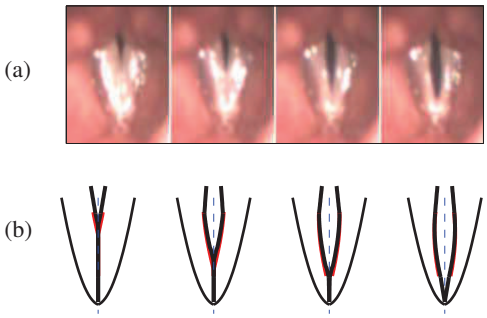

Figure 5: *Selection of frames from HSV S4 (a), and model simulation (b). A zipper-like vibration pattern is observed due to a A-P phase difference equal on both folds.*

In Figure 5, a zipper-like posterior to anterior opening pattern, was simulated by L-R symmetric A-P phase differences. The left and right folds' oscillating frequencies were equally set to $f_0^l = f_0^r = 210$ Hz, and the two longitudinal phase delays were equally set to $\tau_{lon}^l = \tau_{lon}^r = 1.5$ msec, resulting in an A-P phase delay of approximately 108 degrees. Note that in this case, the model was not able to reproduce the closing pattern of the folds, which was a zipper-like A-P motion. For the time being, the model desing allows in fact to reproduce only posterior toward anterior motion patter delays. This is an aspect of the model that will be further investigated.

In Figure 6, an asymmetric A-P phase delay is reproduced by setting $\tau_{lon}^l = 1.5$ msec, and $\tau_{lon}^r = 0$, providing a A-P phase delay of approximately 108 degrees only on the left fold.

4. Conclusions

We discussed the modeling of longitudinal A-P phase differences in vocal folds oscillations by means of a lumped-elements vocal fold model, which relies on distributed elements (delay lines) to represent the longitudinal propagation of the fold displacement. In contrast to conventional biomechanical modeling

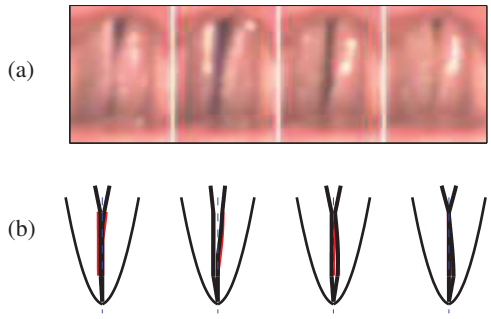

Figure 6: *Selection of frames from HSV S5 (a), and from model simulation (b). The vocal fold shown to the left vibrates with a smaller amplitude than the vocal fold to the right. This is modelled by using a larger mass for reducing vibration amplitude. For this vocal fold, the A-P phase difference is set to 0, whereas for the other one shown to the right, a strong A-P phase shift is used.*

of the folds, in which the main objective is the reproduction of the glottal flow waveform, here we explicitely address the reproduction of the oscillatory patterns observed in high-speed video recordings of the folds, including vertical and longitudinal phase differences and left-right fold mass unbalancing. The model was assessed qualitatively, by empirically tuning its parameters to replicate some typical oscillatory patterns observed in high speed videoendoscopic data, in which A-P phase differences, zipper-like opening and closing patterns, paramedian collision due to the mass unbalancing, are observed.

In its present form, the model has a number of limitations, e.g., the model is able to reproduce closing patterns characterised by a zipper-like posterior-to-anterior motion, but not closing patterns characterised by a zipper-like anterior-to posterior motion. Also, it is possible to control the phase differences, but not other differences such as cycle shape. These are aspects of the model that will be further investigated in the future.

From the point of view of physical correctness of the model, improving it may require to add more degrees of freedom in the mechanics, a more realistic flow description (in particular by taking into account viscosity during the closure of the glottis) and acoustical coupling with the vocal tract. We also stress that the anterior-posterior differences are modeled only morphologically and do not contribute (or at least contribute only partially to the flow computation) to the spontaneous dynamics of the model. These can be seen as rough simplifications if the main interest is in the understanding and accurate simulation of the folds dynamics, but can be an advantage in terms of controllability and stability of the model.

5. Acknowledgment

This work was supported by the Austrian Science Fund (FWF): KLI 722-B30.

The authors want to thank Jean Schoentgen for a fruitful discussion.

6. References

[1] S. Niimi and M. Miyaji, "Vocal fold vibration and voice quality," *Folia Phoniatr. Logop.*, vol. 52, pp. 32–38, 2000.

[2] M. Miyaji, Y. Iwamoto, M. Oda, and S. Niimi, "Relation between voice quality and pathological vibratory patterns using high-speed digital imaging." *Nihon Jibiinkoka Gakkai Kaiho*, vol. 102, no. 3, pp. 354–367, 1999.

[3] I. M. Verdonck-de Leeuw, J. M. Festen, and H. F. Mahieu, "Deviant vocal fold vibration as observed during videokymography: the effect on voice quality." *J Voice*, vol. 15, no. 3, pp. 313–322, 2001. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11575628 http://dx.doi.org/10.1016/S0892-1997(01)00033-9

[4] S. Kiritani, K. Honda, H. Imagawa, and H. Hirose, "Simultaneous high-speed digital recording of vocal fold vibration and speech signal," in *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 11, 1986, pp. 11–15.

[5] T. Shiba and D. Chhetri, "Dynamics of Phonatory Posturing at Phonation Onset," *Laryngoscope*, vol. 126, no. 8, pp. 1837–1843, 2016.

[6] J. Kreiman, Y. Shue, G. Chen, M. Iseli, B. R. Gerratt, J. Neubauer, and A. Alwan, "Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal area waveform shape in sustained phonation," *J. Acoust. Soc. Am.*, vol. 132, no. 4, pp. 2625–2632, 2012.

[7] K. Ahmad, Y. Yan, and D. M. Bless, "Vocal fold vibratory characteristics in normal female speakers from high-speed digital imaging," *J. Voice*, vol. 26, no. 2, pp. 239–253, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.jvoice.2011.02.001

[8] G. Chen, J. Kreiman, B. R. Gerratt, J. Neubauer, Y. Shue, and A. Alwan, "Development of a glottal area index that integrates glottal gap size and open quotient." *J. Acoust. Soc. Am.*, vol. 133, no. 3, pp. 1656–66, 2013.

[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," *J. Voice*, vol. 30, no. 2, pp. 205–214, 2016.

[10] H. Yokonishi, H. Imagawa, K. I. Sakakibara, A. Yamauchi, T. Nito, T. Yamasoba, and N. Tayama, "Relationship of Various Open Quotients with Acoustic Property, Phonation Types, Fundamental Frequency, and Intensity," *J. Voice*, vol. 30, no. 2, pp. 145–157, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.jvoice.2015.01.009

[11] J. P. Jeannon, P. N. Carding, and J. A. Wilson, "Vocim analysis of laryngeal images: Is breathiness related to the glottic area?" *Clin. Otolaryngol. Allied Sci.*, vol. 23, no. 4, pp. 351–353, 1998.

[12] G. Chen, J. Kreiman, Y. Shue, and A. Alwan, "Acoustic correlates of glottal gaps," in *Proc. Int. Conf. Spoken Lang. Process. (Interspeech)*, 2012, pp. 1600–1603.

[13] R. Orlikoff, M. E. Golla, and D. Deliyski, "Analysis of longitudinal phase differences in vocal-fold vibration using synchronous high-speed videoendoscopy and electroglottography," *J. Voice*, vol. 26, no. 6, pp. 816.e13–816.e20, 2012.

[14] H. S. Bonilha, D. Deliyski, and T. Gerlach, "Phase asymmetries in normophonic speakers: visual judgments and objective findings," *Am. J. Speech-Language Pathol.*, vol. 17, no. 4, pp. 367–376, 2008.

[15] G. Chen, J. Kreiman, Y. Shue, and A. Alwan, "Acoustic correlates of glottal gaps," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, pp. 2673–2676, 2011.

[16] K. Ahmad, Y. Yan, and D. Bless, "Vocal fold vibratory characteristics of healthy geriatric females - Analysis of high-speed digital images," *J. Voice*, vol. 26, no. 6, pp. 751–759, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.jvoice.2011.12.002

[17] P. Aichinger, I. Roesner, M. Leonhard, B. Schneider-Stickler, D.-M. Denk-Linnert, W. Bigenzahn, A. Fuchs, M. Hagmüller, and G. Kubin, "Comparison of an audio-based and a video-based approach for detecting diplophonia," *Biomed. Signal Proces.*, vol. 31, pp. 576–585, 2017.

[18] H. Larsson, S. Hertegård, P. Lindestad, and B. Hammarberg, "Vocal Fold Vibrations: High-Speed Imaging, Kymography, and Acoustic Analysis: A Preliminary Report," *Laryngoscope*, vol. 110, no. 12, pp. 2117–2122, 2000.

[19] Y. Yan, K. Izdebski, E. Damrose, and D. Bless, "Quantitative Analysis of Diplophonic Vocal Fold Vibratory Pattern from High-Speed Digital Imaging of Glottis," in *Sixth Int. Work. Model. Anal. Vocal Emiss. Biomed. Appl.*, vol. 6, 2009, pp. 145–148.

[20] R. Patel, L. Liu, N. Galatsanos, and D. M. Bless, "Differential vibratory characteristics of adductor spasmodic dysphonia and muscle tension dysphonia on high-speed digital imaging." *Ann. Otol. Rhinol. Laryngol.*, vol. 120, no. 1, pp. 21–32, 2011.

[21] A. P. Pinheiro, D. E. Stewart, C. D. Maciel, J. C. Pereira, and S. Oliveira, "Analysis of nonlinear dynamics of vocal folds using high-speed video observation and biomechanical modeling," *Digital Signal Processing*, vol. 22, no. 2, pp. 304 – 313, 2012.

[22] M. Döllinger, P. Gómez, R. R. Patel, C. Alexiou, C. Bohr, and A. Schützenberger, "Biomechanical simulation of vocal fold dynamics in adults based on laryngeal high-speed videoendoscopy," *PLOS ONE*, vol. 12, no. 11, pp. 1–26, 11 2017. [Online]. Available: https://doi.org/10.1371/journal.pone.0187486

[23] C. Drioli and G. L. Foresti, "Accurate glottal model parametrization by integrating audio and high-speed endoscopic video data," *Signal, Image and Video Processing*, vol. 9, pp. 1451–1459, 2015.

[24] K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," *The Bell Syst. Tech. J.*, vol. 51, no. 6, pp. 1233–1268, July-August 1972.

[25] T. Koizumi, S. Taniguchi, and S. Hiromitsu, "Two-mass models of the vocal cords for natural sounding voice synthesis," *J. Acoust. Soc. Am.*, vol. 82, no. 4, pp. 1179–1192, October 1987.

[26] I. R. Titze, "The physics of small-amplitude oscillations of the vocal folds," *J. Acoust. Soc. Am.*, vol. 83, no. 4, pp. 1536–1552, April 1988.

[27] X. Pelorson, A. Hirschberg, R. R. van Hassel, and A. P. J. Wijnands, "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model," *J. Acoust. Soc. Am.*, vol. 96, no. 6, pp. 3416–3431, December 1994.

[28] J. C. Lucero, J. Schoentgen, J. Haas, P. Luizard, and X. Pelorson, "Self-entrainment of the right and left vocal fold oscillators," *The Journal of the Acoustical Society of America*, vol. 137, no. 4, pp. 2036–2046, 2015.

[29] C. Drioli, "A flow waveform-matched low-dimensional glottal model based on physical knowledge," *J. Acoust. Soc. Am.*, vol. 117, no. 5, pp. 3184–3195, May 2005.

[30] P. Aichinger, I. Roesner, M. Leonhard, D.-M. Denk-Linnert, W. Bigenzahn, and B. Schneider-Stickler, "A database of laryngeal high-speed videos with simultaneous high-quality audio recordings of pathological and non-pathological voices," in *LREC*, 2016.