Interspeech 2024
1-5 September 2024, Kos, Greece

Measurement and simulation of pressure losses due to airflow in vocal tract models

Peter Birkholz, Patrick Häsner

Institute of Acoustics and Speech Communication, TU Dresden, Germany
peter.birkholz@tu-dresden.de

Abstract

We propose a unified model for viscous and kinetic energy losses in a discrete tube model of the vocal system including the glottis. In this model, a lossless Bernoulli flow is assumed at each transition between two tube sections if the downstream section has a smaller diameter than the upstream section, and otherwise the recovery of a fixed fraction of the dynamic pressure. For viscous losses, we propose a general equation according to which the pressure drop within a tube section is inversely proportional to a certain power of its cross-sectional area. The parameters of the model were adjusted to reproduce the results of measurements with physical replicas of the glottis and the vocal tract. The best agreement with the experimental data was achieved when 29% of the dynamic pressure were recovered at tube expansions, and when the viscous losses were proportional to the tube area to the power of -2.9. These results may improve articulatory speech synthesis.

Index Terms: one-dimensional flow model, flow resistance

1. Introduction

Articulatory speech synthesizers mostly use one-dimensional aerodynamic-acoustic simulations of the vocal tract, because they are much faster than two-dimensional or three-dimensional simulations and physically valid up to frequencies of about 4 kHz [1, 2, 3, 4]. An important element for realistic simulations is the modeling of energy losses. For voiced speech sounds they determine the bandwidths of the formants, and for unvoiced sounds they determine the airflow rate for a certain subglottal pressure, which in turn affects the power spectrum of turbulent noise sources [5].

Currently, the energy loss at the glottis is usually modeled differently than losses further downstream in the vocal tract. With regard to the glottis, the transglottal pressure drop ΔP is often modeled as

$$\Delta P = \underbrace{\Delta x \frac{12\mu}{lw^3} U}_{\Delta P_\mathrm{v}} + \underbrace{(k_\mathrm{ent} - k_\mathrm{exit}) \frac{\varrho}{2A^2} U^2}_{\Delta P_\mathrm{k}} \qquad (1)$$

based on a study by van den Berg et al. [6]. The term ΔP_v represents the viscous pressure loss of a fully developed laminar flow through a rectangular duct (the glottis), where Δx is the length of the duct (axial glottal length), l is the long side of the rectangle (anterior-posterior glottal length), w is the short side of the rectangle (glottal width), μ is the viscosity of the air, and U is the volume velocity through the glottis. The area of the glottis is $A = lw$, hence the viscous pressure loss can be rewritten as $\Delta P_\mathrm{v} = \Delta x 12\mu l^2/A^3$. When we assume that the glottal area changes mainly due to changes of its width w, and the length l remains constant, then $\Delta P_\mathrm{v} \propto A^{-3}$. Fulcher et al. [7] proposed a modification to the viscous term, which better fitted their own data from flow experiments with a physical replica of the glottis, namely

$$\Delta P_\mathrm{v} = \Delta x \frac{12\mu}{lw^3} \left(\frac{w}{w_\mathrm{ref}} \right)^\lambda U, \qquad (2)$$

where $w_\mathrm{ref} = 0.01$ cm and $\lambda = 0.41$. With this modification the viscous pressure loss becomes proportional to $A^{-2.59}$.

The term ΔP_k in Eq. (1) represents the loss of kinetic energy, where ϱ is the density of the air, and k_ent and k_exit describe the effects responsible for energy losses at the glottal entrance and the effect of pressure recovery at the glottal exit, respectively. The case $k_\mathrm{ent} = 1$ and $k_\mathrm{exit} = 0$ would correspond to a lossless conversion of static to dynamic pressure at the entrance of the glottis according to Bernoulli's equation, and a loss of all dynamic pressure at the exit of the glottis. Potential energy losses during the conversion from static to dynamic pressure can be considered by setting $k_\mathrm{ent} > 1$. Furthermore, a certain percentage of dynamic pressure is usually recovered at the glottal exit, which can be considered by setting $k_\mathrm{exit} > 0$. The coefficients are frequently set to the values obtained by van den Berg et al. [6], namely $k_\mathrm{ent} = 1.375$ and $k_\mathrm{exit} = 0.5$. However, there is no consensus on these values. For example, Ishizaka et al. [8] proposed to calculate k_exit based on the conservation of momentum at the glottal exit, which makes k_exit a function of the cross-sectional area of the glottis and the area immediately downstream from the glottis. The resulting coefficient is typically in the order of 0.05 to 0.4. With respect to the entrance coefficient, Pelorson et al. [9, 10] argue that $k_\mathrm{ent} = 1$ instead of 1.375 because the human glottis has a smooth entrance at which no significant flow separation should occur.

In the vocal tract above the glottis, viscous losses are often modeled as distributed pressure losses using Poiseuille's equation, which describes the pressure drop across a cylindrical tube section of length Δx and cross-sectional area A under the assumption of a fully developed laminar flow [11, 12]:

$$\Delta P_\mathrm{v} = \Delta x \frac{8\mu\pi}{A^2} U. \qquad (3)$$

In contrast to ΔP_v in Eqs. (1) and (2), here $\Delta P_\mathrm{v} \propto A^{-2}$.

A supraglottal *kinetic* energy loss is usually only considered locally at the site of a critical constriction with the area A_c (e.g. in fricatives) as

$$\Delta P_\mathrm{k} = k_\mathrm{c} \cdot \frac{\varrho}{2A_\mathrm{c}^2} U^2, \qquad (4)$$

where k_c is an empirical coefficient [13, 1]. However, using a lumped loss at constrictions can cause acoustic artifacts in

10.21437/Interspeech.2024-1970

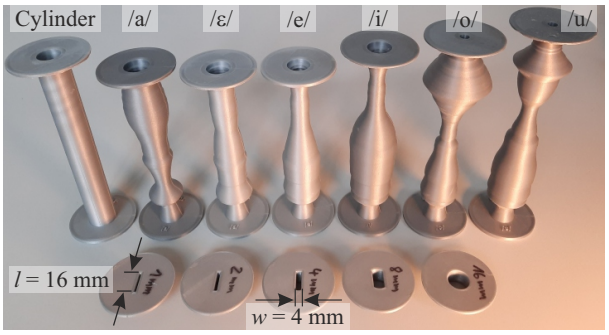

Figure 1: *Back row: Tube models of the vocal tract. The mouth opening is at the upper end. Front row, from left to right: Glottis models with glottal widths of 1, 2, 4, 8, and 16 mm.*

aerodynamic-acoustic simulations of running speech, where the constriction sites change dynamically [14].

The above overview shows that there are several ways of modeling the viscous and kinetic energy losses in the vocal tract. However, there is little experimental evidence for one-dimensional flow models of the vocal tract [15], especially when they include *both* the glottis and the supraglottal cavities. The goal of the present study was to propose a *unified* approach to modeling viscous and kinetic energy losses in the vocal tract, which is suitable to be included in one-dimensional models for articulatory speech synthesis. In this model, not only the viscous losses but also the kinetic energy losses are distributed (as opposed to localized losses in previous studies). The model has three parameters that were adjusted to reproduce flow data obtained with physical replicas of the vocal tract including the glottis.

2. Method

2.1. Measurements with physical tube models

To obtain reference data for the proposed theoretical model (see Sec. 2.2), we performed flow experiments with physical replicas of the glottis and the vocal tract. Glottis models with different glottal areas were combined with tube models of the vocal tract for different vowels, and the pressure drop across these models was measured for multiple airflow rates.

A photograph of all vocal tract and glottis models is shown in Figure 1. They were designed with the software Autodesk Inventor (version 2019) and 3D-printed on an Ultimaker 5S printer using the material polylactic acid. The vocal tract models comprise a cylinder as well as straight axisymmetric tubes with area functions of the vowels /a, ɛ, e, i, o, u/. The area functions were exported from the articulatory speech synthesizer VocalTractLab 2.3 (`www.vocaltractlab.de`) based on the respective vocal tract shapes of the standard speaker model. The tube models were chosen because they cover a range of conditions with different numbers of constrictions and constrictions of different lengths. For example, the models for /e/ and /i/ have a single relatively long constriction in the anterior part, while the models for /o/ and /u/ have two shorter constrictions in the middle and at the anterior end. The smallest cross-sectional areas of these models are in the range of 0.25 to 0.36 cm^2. The diameter at the glottal end is 1.7 cm for all models, which tapers to a diameter of about 0.68 cm (area of 0.36 cm^2) at the upper end of the epilaryngeal tube. The cylinder model has a length of 17 cm and a uniform diameter of 1.7 cm. Further details about these models can be found in [16].

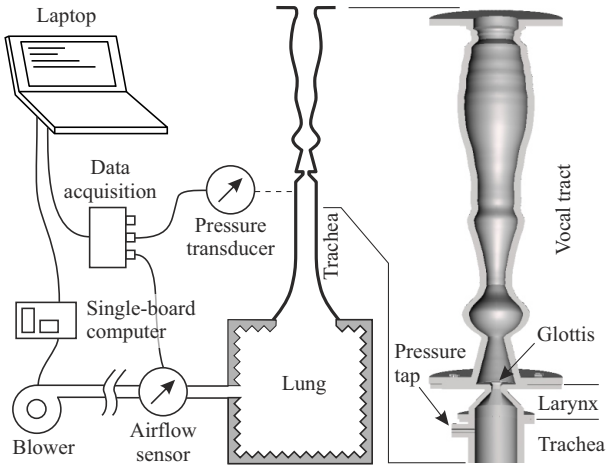

Figure 2: *Measurement setup. The right side shows a cutaway view of the vocal tract for /a/, the glottis replica ("Larynx"), and the upper part of the trachea.*

The glottis models consist of three regions: a short cylindrical part with a diameter of 1.8 cm that represents the upper part of the trachea, a 7 mm long section that represents the conus elasticus where the cross-section tapers towards that of the actual glottis, which is represented as a rectangular duct (with slightly rounded edges) with a length of 3 mm (glottal depth), a long side (glottal length) of 16 mm and a short side of the glottal width w. We made five glottis models with glottal widths of 1, 2, 4, 8, and 16 mm. For the width of 16 mm, the glottal area became a full circle with a diameter of 16 mm, as shown in Figure 1. Glottal widths of $w = 1$ mm and $w = 2$ mm roughly represent intermediate and maximal areas during phonation, $w = 4$ mm and $w = 8$ mm correspond to glottal areas used for voiceless fricatives [5], and $w = 16$ mm corresponds to an area of 201 mm^2 typical for breathing [17]. The area functions and 3D-printable STL files of all models are provided in the supplemental material[1].

The glottis and vocal tract replicas were used in the measurement setup shown in Figure 2. Here, air was fed from a radial blower (U71HX-024KX-6 by micronel, Tagelswangen, Switzerland) through a 2 m long tube into a box representing the lungs and from there through a model of the trachea. The glottis and vocal tract models were mounted on the upper end of the trachea, as shown in the cutaway view at the right side of the figure. A pressure transducer (DMU4, Kalinsky Sensor Elektronik, Erfurt, Germany) was used to capture the subglottal pressure at the upper end of the trachea, and an airflow sensor (AWM720-P1, Honeywell, Charlotte, North Carolina) was used to capture the flow rate at the entrance to the "lung box". The sensor values were digitized by a data acquisition device (DataTranslation 9837C, Norton, MA) with a sampling rate of 48 kHz and 24 bits quantization per channel and sent to a laptop. The laptop was running a custom software to display the data and control the power of the radial blower.

For each combination of a glottis model and a vocal tract model, the blower power was adjusted sequentially to generate 7 fluid power levels: 100, 200, 400, 700, 1100, 1600, and

[1] `https://www.vocaltractlab.de/index.php?page=birkholz-supplements`

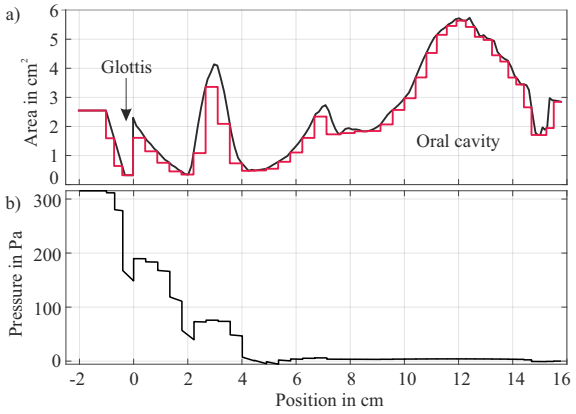

Figure 3: *a) Continuous (black) and discrete (red) area functions for the glottis model (glottal width of 2 mm) concatenated with the tube model for /a/. b) Simulated distribution of the static pressure for a flow rate of 500 cm³/s.*

2200 mW. At each power level, the pressure and airflow rate were measured. The measurements were carried out three times on different days and averaged over the repetitions. Since the results for the glottal widths of 8 mm and 16 mm were nearly identical (less than 5% deviation for all measurements), the data with the 16 mm glottis were excluded from further analysis.

2.2. Theoretical model

Our approach to model viscous and kinetic losses in a uniform way assumes that the vocal tract (including the glottis and subglottis) is represented as a sequence of short abutting cylindrical tube sections [18]. The lengths and cross-sectional areas of the tube sections are given in terms of a discrete area function. As an example, the red curve in Figure 3a shows the discrete area function that was derived from the combined continuous area function (black curve) of the physical replicas of the glottis and the vocal tract for /a/. As proposed in [19], the areas of the discrete sections assume the minimum of the continuous area function in the regions of the sections.

As shown in Section 1, previous approaches to model the *viscous pressure loss* ΔP_v over a short tube section of length Δx have in common that they are proportional to the airflow rate U, to Δx and to a (negative) power of the cross-sectional area. However, the proportionality factor and the area exponent differ between the equations. Here we propose the following general equation for the viscous loss over a tube section i,

$$\Delta P_{\mathrm{v},i} = l_i R_\mathrm{ref} \cdot (A_\mathrm{ref}/A_i)^\alpha \, U, \tag{5}$$

where l_i and A_i are the length and cross-sectional area of the tube section, and R_ref, A_ref and α are free parameters. A_ref was fixed to 1 cm^2, so that R_ref can be interpreted as the flow resistance (ratio of pressure drop to volume flow) per unit length for an area of 1 cm^2, and α models the variation of the resistance with the area. Both R_ref and α were adjusted to fit the simulations to our experimental data (see Section 2.3).

Inspired by [14], *kinetic energy losses* are distributed along the entire vocal system and depend on whether the tube is locally contracting or expanding. Specifically, the pressure change at each transition between two tube sections i and $i+1$ is assumed to be proportional to that of a lossless Bernoulli flow

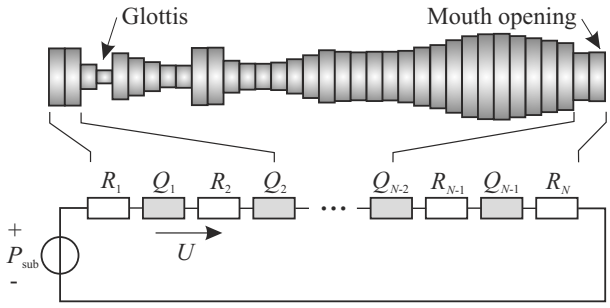

Figure 4: *Discrete tube model and fluidic network.*

between these sections, namely

$$\Delta P_{\mathrm{k},i} = \begin{cases} k_\mathrm{c} \cdot (\varrho/2)(1/A_{i+1}^2 - 1/A_i^2)U^2 & \text{for } A_{i+1} < A_i \\ k_\mathrm{e} \cdot (\varrho/2)(1/A_{i+1}^2 - 1/A_i^2)U^2 & \text{otherwise,} \end{cases} \tag{6}$$

where the proportionality factors k_c and k_e account for potential losses at tube contractions and expansions, respectively. With this approach, the total kinetic pressure change across a tube contraction or expansion that extends over multiple tube sections (sum of the involved $\Delta P_{\mathrm{k},i}$) depends only on the initial and final cross-sectional areas of the expansion or contraction. This makes the results independent of the discretization step size of the area function.

A summation of all $\Delta P_{\mathrm{k},i}$ for the *whole* tube furthermore shows that the overall kinetic pressure loss is essentially a function of the *difference* $(k_\mathrm{c} - k_\mathrm{e})$ of the coefficients, so that their individual values cannot be independently determined from our experimental data. Hence, following [9] we set $k_\mathrm{c} = 1$ in this study, assuming a lossless conversion from static to dynamic pressure at tube contractions. This leaves k_e as the only parameter to be determined for the kinetic energy losses.

With these considerations, the pressure changes in the vocal tract can be represented in terms of the fluidic network shown in Figure 4. Here, P_sub is the subglottal pressure, N is the number of tube sections, R_n are *linear* resistors that cause a pressure change proportional to U according to Eq. (5), and Q_n are *non-linear* resistors that cause a pressure change proportional to U^2 according to Eq. (6). Note that $Q_N = 0$, because there is no pressure recovery at the tube exit [20]. The network allows to determine the complete pressure distribution along the tube axis for steady flows and can be readily combined with the transmission-line circuit representation of the vocal tract to model both aerodynamics and acoustics [18, 14].

2.3. Determination of model parameters

To find values for the three model parameters R_ref, α and k_e that minimize the deviation between measurements and simulations, we performed a complete grid search in the intervals $R_\mathrm{ref} \in [800; 2.6 \cdot 10^6]$ Pa·s/m^4, $\alpha \in [2; 3]$, and $k_\mathrm{e} \in [0; 0.8]$. For each setting of the parameters, the subglottal pressure was simulated for all airflow rates and tube models used in the experiments in Section 2.1 based on the discretized area functions. The overall deviation between simulations and measurements for a certain parameter setting was then calculated as the *root mean square percentage error* E [21] between simulated and measured subglottal pressure values across all tube models and flow rates.

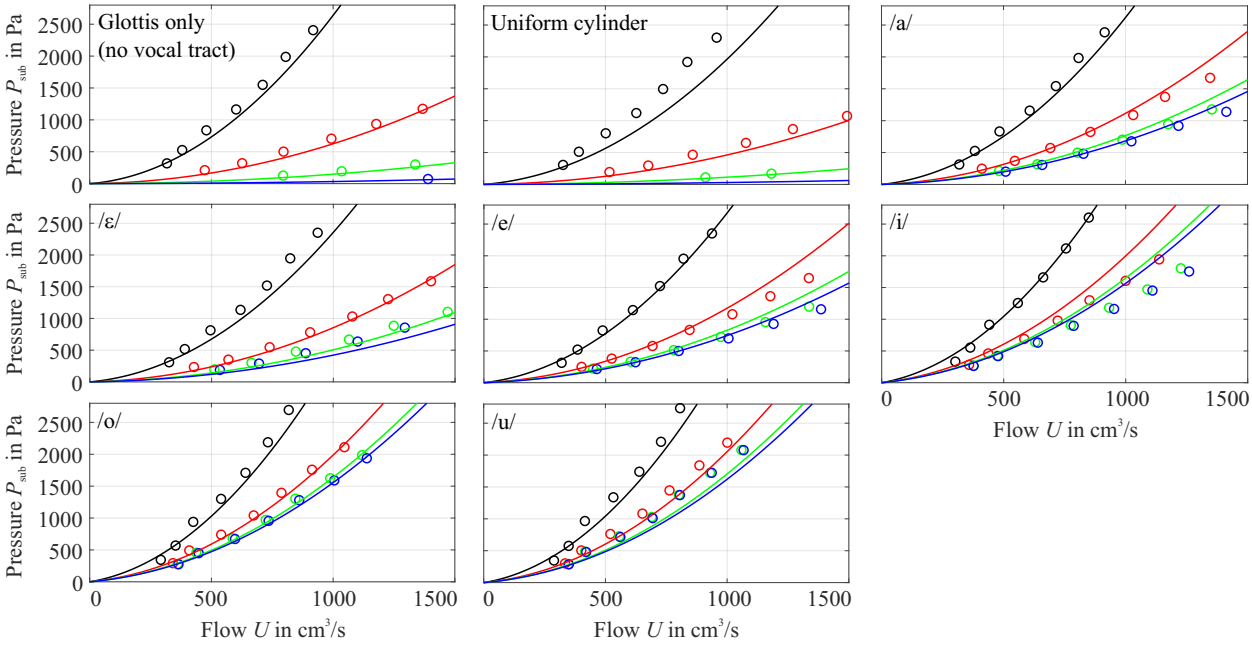

Figure 5: *Measured flow-pressure points (circles) and simulated flow-pressure curves with optimized parameters. The black, red, green, and blue curves were obtained for glottal widths of 1 mm, 2 mm, 4 mm, and 8 mm, respectively.*

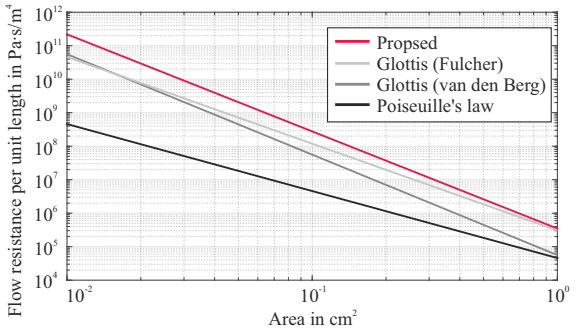

Figure 6: *Variation of the flow resistance per unit length as a function of the cross-sectional area for different equations.*

3. Results and discussion

The parameter values that minimize the deviation between the measured and simulated subglottal pressure values across all models and flow rates are $R_{\text{ref}} = 347,922$ Pa·s/m^4, $\alpha = 2.9$ and $k_{\text{e}} = 0.29$ with an error of $E = 14.9\%$. Figure 5 shows the measured flow-pressure pairs as circles along with the simulated flow-pressure curves. Different glottal widths are represented by different colors (black for 1 mm, red for 2 mm, green for 4 mm, and blue for 8 mm), while each panel represents one vocal tract model (including the absence of a vocal tract in the top left panel). The simulations approximate the measured data points reasonably well.

With regard to the kinetic energy losses, the difference of the contraction and expansion factors is $k_{\text{c}} - k_{\text{e}} = 0.71$ in our model. This value is somewhat lower than 0.875 as determined by van den Berg for the glottis [6], and lower than the value of 1.0 assumed by Birkholz et al. [14]. Hence, the overall kinetic energy losses in the proposed model are lower than in these earlier models.

With regard to viscous losses, we obtained an area exponent of 2.9, which is close to the value of 3 assumed in many studies for the glottis, and higher than the value of 2.59 obtained by Fulcher et al. [7] or the value of 2 in Poiseuille's law. Figure 6 compares the flow resistance as a function of the tube area between the proposed Eq. (5) with the optimized parameters (red curve), Eq. (2) proposed for the glottis by Fulcher et al. (light gray curve), the viscous resistance for the glottis in Eq. (1) (gray curve), and Poiseuille's law (black curve). This shows that Poiseuille's law, which is often used to model viscous losses in the vocal tract, strongly underestimates the losses for small cross-sectional areas. This in turn leads to an underestimation of formant bandwidths in acoustic simulations [22].

The fluidic network shown in Figure 4 and the optimized model parameters allow to calculate the pressure distribution along the tube axis for any given area function and airflow rate. As an example, Figure 3b shows the pressure distribution for the glottis with a width of 2 mm connected to the vocal tract for /a/ at a flow rate of 500 cm^3/s. For the three tube contractions and expansions in the posterior part, the associated pressure drops and (partial) recoveries can be clearly seen. In future studies, it would be interesting to compare such pressure profiles with distributed pressure measurements along the whole tube axis.

Finally, we would like to compare the error of 14.9% between measurements and simulations for the optimized model parameters with other choices for the model parameters. If we model viscous losses with Poiseuille's law and set $k_{\text{e}} = 0$ for tube expansions as in [14], the error increases to 20.9%. When the viscous loss according to Poiseuille's law is combined with $k_{\text{e}} = 0.29$, the error becomes 28.0%, and when the optimized viscous resistance is combined with $k_{\text{e}} = 0$, the error becomes 31.2%. This shows that the parameters for both viscous and kinetic losses must be set correctly to achieve optimum agreement with the experimental data.

4. Acknowledgements

We thank Mario Fleischer for useful discussions about this study.

5. References

[1] A. J. S. Teixeira, R. Martinez, L. N. Silva, L. M. T. Jesus, J. C. Principe, and F. A. C. Vaz, "Simulation of human speech production applied to the study and synthesis of European Portuguese," *EURASIP Journal on Applied Signal Processing*, vol. 9, pp. 1435–1448, 2005.

[2] S. Fels, J. E. Lloyd, K. Van Den Doel, F. Vogt, I. Stavness, and E. Vatikiotis-Bateson, "Developing physically-based, dynamic vocal tract models using ArtiSynth," in *Proc. of the 7th International Seminar on Speech Production (ISSP 2006)*, Ubatuba, Brazil, 2006, pp. 419–426.

[3] P. Birkholz, "Modeling consonant-vowel coarticulation for articulatory speech synthesis," *PLOS ONE*, vol. 8, no. 4, p. e60603, 2013.

[4] R. Blandin, M. Arnela, S. Félix, J.-B. Doc, and P. Birkholz, "Efficient 3D acoustic simulation of the vocal tract by combining the multimodal method and finite elements," *IEEE Access*, vol. 10, pp. 69 922–69 938, 2022.

[5] K. N. Stevens, *Acoustic Phonetics*. The MIT Press, 1998.

[6] J. Van Den Berg, J. T. Zantema, and P. J. Doornenbal, "On the air resistance and the Bernoulli effect of the human larynx," *The Journal of the Acoustical Society of America*, vol. 29, no. 5, pp. 626–631, 1957.

[7] L. P. Fulcher, R. C. Scherer, and T. Powell, "Viscous effects in a static physical model of the uniform glottis," *The Journal of the Acoustical Society of America*, vol. 134, no. 2, pp. 1253–1260, 2013.

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

[9] X. Pelorson, A. Hirschberg, R. R. van Hassel, A. P. J. Wijnands, and Y. Auregan, "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. application to a modified two-mass model," *The Journal of the Acoustical Society of America*, vol. 96, no. 6, pp. 3416–3431, 1994.

[10] X. Pelorson, C. Vescovi, E. Castelli, A. Hirschberg, A. P. J. Wijnands, and H. M. A. Bailliet, "Description of the flow through in-vitro models of the glottis during phonation. application to voiced sound synthesis," *Acta Acustica*, vol. 82, pp. 358–361, 1996.

[11] S. Maeda, "A digital simulation method of the vocal-tract system," *Speech Communication*, vol. 1, pp. 199–229, 1982.

[12] B. Elie and Y. Laprie, "Extension of the single-matrix formulation of the vocal tract: Consideration of bilateral channels and connection of self-oscillating models of the vocal folds with a glottal chink," *Speech Communication*, vol. 82, pp. 85–96, 2016.

[13] M. M. Sondhi and J. Schroeter, "A hybrid time-frequency domain articulatory speech synthesizer," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 7, pp. 955–967, 1987.

[14] P. Birkholz, D. Jackèl, and B. J. Kröger, "Simulation of losses due to turbulence in the time-varying vocal system," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 4, pp. 1218–1226, 2007.

[15] A. V. Hirtum, X. Pelorson, O. Estienne, and H. Bailliet, "Experimental validation of flow models for a rigid vocal tract replica," *The Journal of the Acoustical Society of America*, vol. 130, no. 4, pp. 2128–2138, 2011.

[16] P. Birkholz, P. Häsner, and S. Kürbis, "Acoustic comparison of physical vocal tract models with hard and soft walls," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2022)*, Singapore, 2022, pp. 8242–8246.

[17] A. Scheinherr, L. Bailly, O. Boiron, A. Lagier, T. Legou, M. Pichelin, G. Caillibotte, and A. Giovanni, "Realistic glottal motion and airflow rate during human breathing," *Medical Engineering & Physics*, vol. 37, no. 9, pp. 829–839, 2015.

[18] J. L. Flanagan, *Speech Analysis, Synthesis and Perception*. Springer-Verlag, Berlin, 1965.

[19] P. Birkholz, "Enhanced area functions for noise source modeling in the vocal tract," in *Proc. of the 10th International Seminar on Speech Production (ISSP 2014)*, Cologne, Germany, 2014, pp. 37–40.

[20] A. C. Yunus and C. J.M., *Fluid Mechanics: Fundamentals and Applications (SI Units)*. McGraw Hill, 2006.

[21] M. V. Shcherbakov, A. Brebels, N. L. Shcherbakova, A. P. Tyukov, T. A. Janovsky, V. A. Kamaev *et al.*, "A survey of forecast error measures," *World Applied Sciences Journal*, vol. 24, no. 24, pp. 171–176, 2013.

[22] P. Birkholz, R. Blandin, and S. Kürbis, "Bandwidths of vocal tract resonances in physical models compared to transmission-line simulations," *The Journal of the Acoustical Society of America*, vol. 153, no. 6, pp. 3281–3281, 2023.