ACOUSTIC COMPARISON OF PHYSICAL VOCAL TRACT MODELS WITH HARD AND SOFT WALLS

P. Birkholz, P. Häsner, S. Kürbis

Institute of Acoustics and Speech Communication, TU Dresden, Germany

ABSTRACT

This study explored how the frequencies and bandwidths of the acoustic resonances of physical tube models of the vocal tract differ when they have hard versus soft walls. For each of 10 tube shapes representing different vowels, two physical models were made: one with rigid plastic walls, and one with soft silicone walls. For all models, the acoustic transfer functions were measured and the bandwidths and frequencies of the first three resonances were determined. For the models with soft walls, the resonance frequencies and bandwidths below 2 kHz were consistently higher than for the models with hard walls. These results confirm the general predictions of mathematical lumped-element models for the mechanical impedance of the vocal tract walls. The resonance bandwidths of the silicone models were well within the range of 50–90 Hz measured for humans, while they were too small for the rigid tubes below 2 kHz.

Index Terms— Vocal tract walls, transfer function measurement, acoustic resonances

1. INTRODUCTION

Physical models of the vocal tract and the vocal folds are valuable tools for speech research because they allow the study of acoustic and aerodynamic effects of precisely and individually controlled model parameters. These models are used, for example, to study source-filter interaction [1, 2], formant tuning in singing at high pitches [3], the acoustic effect of side cavities of the vocal tract [4], or the effect of subglottal or supraglottal conditions on phonation [5, 6].

The used physical models of vocal folds are often quite advanced and mimic the shape, the layered structure and the softness of human vocal folds [7, 8, 2]. In contrast, the vocal *tract* is often simplified as a resonator with hard walls. The complexity of the vocal tract models ranges from straight tubes with a constant cross-sectional area [5], to static tubes with area functions that mimic specific vowels [9, 2, 10], to articulated models with an adaptable shape (e.g., robotic vocal tract models like [11, 12]). While some of the robotic

vocal tract models have a soft deformable tongue [11], the remaining vocal tract usually still has hard walls.

Obviously, the fleshy surfaces of the human vocal tract are not rigid and thus have a finite mechanical impedance. The acoustic effects of the soft walls have been analyzed in a number of studies [13, 14, 15, 16], but mainly in terms of mathematical models that represent the wall impedance as a lumped mass–compliance–viscous loss combination (see Sec. 2). Such models predict that soft walls increase the frequencies and the bandwidths of the resonances at low frequencies (mainly of the first resonance), and that these effects quickly diminish towards higher frequencies.

However, to our knowledge, the predicted effects have not been systematically shown for *physical* vocal tract models yet. In this study, we made physical tube models for 10 different vowels – each in one variant with rigid plastic walls and in one variant with soft silicone walls. The acoustic transfer functions of all models were measured and the frequencies and bandwidths of the resonances were analyzed. The acoustic characteristics due to soft walls were compared with the general predictions of the mathematical models and with bandwidth measurements of the human vocal tract.

2. COMPUTATIONAL ANALYSIS

While the mechanical impedance of the vocal tract walls could in principle have any frequency characteristic, Ishizaka et al. [17] inferred from measurements with human tissue that it behaves locally much like a lumped–element spring–mass–damper system at low frequencies. Accordingly, the mechanical wall impedance *per-unit-area* $\overline{Z}$ can be written as

$$\overline{Z} = B + j\omega M + K/(j\omega), \tag{1}$$

where ω is the angular frequency and B, M, and K are the per-unit-area values of the mechanical resistance, mass, and stiffness, respectively. This approach has been widely adopted to model the vocal tract walls, usually assuming a uniform distribution of the impedance along the tract [13, 15, 18]. Since the stiffness term in Eq. (1) becomes very small compared to the other terms for frequencies around the first resonance and above [19], it is often neglected so that Eq. (1) becomes $\overline{Z} = B + j\omega M$.
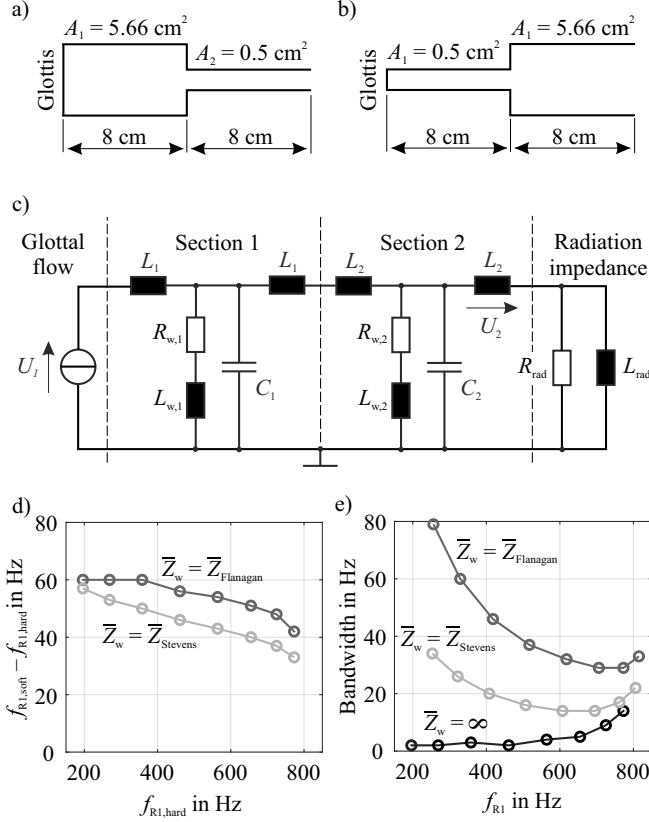
Fig. 1. *a, b) Simplified tube models with two sections for the lowest f_{R1} and the highest f_{R1}, respectively. c) Acoustic network model for the tube models in a, b). d) Shift of f_{R1} due to soft walls for two different values for the per-unit-area wall impedance (see text). e) Bandwidths of the first resonance for different values of the per-unit-area wall impedance.*

In the following, we analyze the effect of this wall impedance model on the frequency f_{R1} and bandwidth B_{R1} of the first resonance using a two-tube approximation of the vocal tract. Thereby, the tract is approximated in terms of two concatenated cylindrical tube sections with different cross-sectional areas A_1 and A_2, a constant length $l_1 = l_2 = 8$ cm for both sections, and an infinite glottal impedance. To obtain a wide range of f_{R1} values, we consider 8 model configurations that vary from an /i/-like shape (Fig. 1 a) to an /a/-like shape (Fig. 1 b), namely $(A_1, A_2) \in \{(5.66, 0.5), (4, 0.71), (2.83, 1), (2, 1.41), (1.41, 2), (1, 2.83), (0.71, 4), (0.5, 5.66)\}$ (cm^2).

For low frequencies[1] these tube models can be represented by the lumped-element acoustic network in Fig 1c. Here, $L_i = \varrho l_i/(2A_i)$ and $C_i = A_i l_i/(\varrho c^2)$ $(i = 1, 2)$ represent the acoustic mass and compliance of the air in the tube sections, $L_{\text{rad}} = 8\varrho/(3\pi\sqrt{\pi A_2})$ and $R_{\text{rad}} =$

---

[1]At the highest considered frequency of 800 Hz, the wavelength is still $5.5\times$ the section length of 8 cm.

$128\varrho c/(9\pi^2 A_2)$ constitute the radiation impedance, and $\varrho = 0.00114$ g/cm^3 is the air density and $c = 35000$ cm/s is the sound velocity [20]. The components $R_{\text{w},i}$ and $L_{\text{w},i}$ constitute the (acoustic) wall impedance $Z_{\text{w},i}$, which is related to the mechanical per-unit-area wall impedance as $Z_{\text{w},i} = \overline{Z}/(S_i \cdot l_i)$, where S_i is the perimeter and $S_i \cdot l_i$ is the surface area of tube section i. Losses due to viscous frication at the tube wall have been omitted because of their small magnitude at low frequencies [13].

For the acoustic analysis of this network, we consider three cases for the wall impedance: hard walls ($\overline{Z} = \infty$), and two specific settings for soft walls proposed by Flanagan et al. [13] and Stevens [15]:

$$\overline{Z}_{\text{Flanagan}} = (1600 + j\omega \cdot 1.5)\,\text{g s}^{-1}\,\text{cm}^{-2}, \quad (2)$$

$$\overline{Z}_{\text{Stevens}} = (1000 + j\omega \cdot 2.0)\,\text{g s}^{-1}\,\text{cm}^{-2}. \quad (3)$$

For all three cases and for all model configurations, the volume velocity transfer function $U_2(\omega)/U_1(\omega)$ (see Fig. 1 c) was determined using the transmission-line matrix method [20]. For each of these transfer functions, the frequency and 3-dB bandwith of the first resonance was determined from the magnitude spectrum.

The results are shown in Fig. 1 d and e. With soft walls, f_{R1} is increased by 30–60 Hz compared to the case with hard walls. This increase is mainly determined by the reactance of $\overline{Z}$ with a stronger frequency increase for a smaller reactance. Furthermore, the soft walls lead to an increased bandwidth of the first resonance. This bandwidth increase is mainly determined by the resistive component of $\overline{Z}$ and decreases with increasing f_{R1}. These results agree well with analyses based on similar acoustic circuits [13, 15].

3. PHYSICAL VOCAL TRACT MODELS

To analyze the acoustic effects of soft walls for *physical* tube models, we created, for each of 10 inner vocal tract shapes, one resonator with hard walls and one with soft walls. All resonators were designed as straight tubes with circular cross-sections. Nine of the tube shapes were based on the area functions of the German vowels /a, e, i, o, u, ɛ, ø, y, ə/ defined in the software VocalTractLab 2.3 (www.vocaltractlab.de). The 10th shape was simply a cylinder with a length of 17 cm and a diameter of 1.7 cm.

To make the resonators with rigid walls, 3D-printable tube models with a wall thickness of 3 mm were designed for each of the 10 area functions using the CAD software Autodesk Inventor (version 2019). At the ends of the tubes, flanges with a diameter of 57 mm were added. All ten models were 3D-printed with an Ultimaker 3 printer using the plastic material PLA (polylactic acid) with an infill ratio of 100 %.

The resonators with soft walls were made of silicone (TFC type 13 by Trollfactory) with the shore hardness 00 ("super soft"). The inner shape (i.e. the air-filled cavity

shape) of the models was identical to their respective counterparts with hard walls, but the thickness of the soft walls was set to 10 mm, which corresponds roughly to the thickness of the cheeks. The models were cast in two halves and glued together with silicone glue (type SIL-POXY/0 by KauPo) after curing. The casting molds were designed as two parts (Fig. 2 a) with Autodesk Inventor and 3D-printed with PLA.

Fig. 2 b shows the finished silicone and the plastic tube models for the vowels /a/, /i/, and /u/. The area functions and the 3D-printable STL files of the tube models and th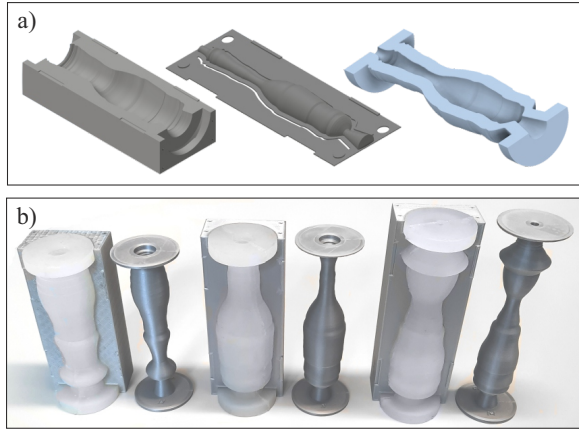e casting molds are available as supplemental material at `https://www.vocaltractlab.de/index.php?page=birkholz-supplements`.

Fig. 2. *a) The casting mold (left, without walls at the ends) and the negative (middle) to create one half of a silicone tube model (right). b) The silicone and the plastic tube models for the vowels /a/, /i/, and /u/ (from left to right). The silicone models are leaning against the casting mold to stand upright.*
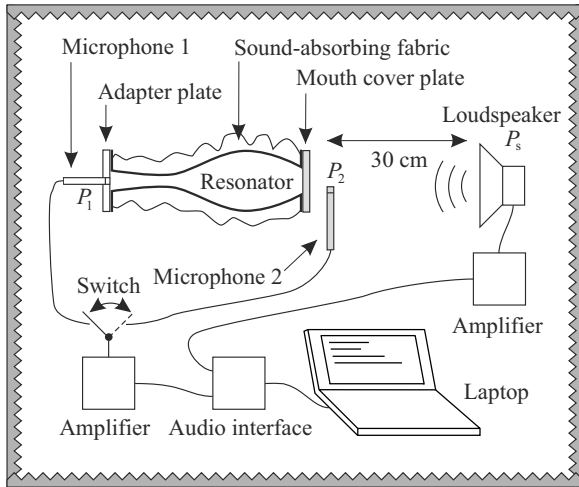

Fig. 3. *Setup to measure the volume velocity transfer function of the resonators in an anechoic chamber.*

4. MEASUREMENTS

For all physical resonators, the volume velocity transfer function (VVTF) from the glottis to the lips (corresponding to $U_2(\omega)/U_1(\omega)$ in Fig. 1 c) was measured with the method by Fleischer et al. [21], which is based on the principle of reciprocity. The setup is schematically shown in Fig. 3, where the tube model to measure is placed about 30 cm in front of a loudspeaker, which is directed towards the "mouth opening" of the model. The actual measurement is performed in two steps. In the first step, the loudspeaker emits a broadband sine sweep signal (100 Hz to 10 kHz in our case) while the sweep response P_1 is measured with a microphone inside the model at the closed glottal end (the gray mouth cover and microphone 2 in Fig. 3 are removed in this step). In the second step, the "mouth" is closed with a cover plate, the loudspeaker emits the same sweep signal as before, and a second microphone measures the sweep response P_2 immediately in front of the cover plate. The VVTF can then be calculated from these sound pressure signals as $H(\omega) = P_1(\omega)/P_2(\omega)$. More details of this method are given in [21, 10] and in the video `https://youtu.be/9AoRS9X2BNY`.

Fig. 4 shows the measured transfer functions of all resonators (black curves for the hard walls, and gray curves for the soft walls), which show the expected resonance patterns for vowels. For /i/, /o/, and /u/ there were a few deviations from a perfect all-pole spectrum between 1–2.5 kHz, which we cannot fully explain yet. However, apart from f_{R3} for /o/ and /u/, the frequencies and the 3-dB-bandwidths of the first three resonances could be easily measured for all models. For illustration, the black vertical lines in Fig. 4 mark the resonance frequencies for the models with hard walls.

5. RESULTS AND DISCUSSION

Fig. 5 a shows the shift of the resonance frequencies $f_{R1/2/3}$ due to soft walls as a function of the resonance frequency of the hard models. At the lowest resonance frequencies, the difference amounts to about 150 Hz and then drops to around 25 Hz at 1.5 kHz, before it starts to scatter more strongly above 1.5 kHz. These measurements generally confirm the pattern predicted by the calculations for f_{R1} in Sec. 2, but the *calculated* shifts of the first resonance were less severe with a maximum of 60 Hz and the relative slope of the decrease was less steep. The stronger frequency shifts of the physical models could be explained with the thickness of the silicone walls of 1 cm. In the mathematical model for the wall impedance, this would correspond to a wall mass per unit area of about 1 g/cm^2 (assuming a density of 1 g/cm^3 of the silicone), which is less than the 1.5 g/cm^2 and 2 g/cm^2 in Eqs. (2) and (3). As it was found in Sec. 2 and discussed in [15], a smaller wall mass (reactive component of the mechanical wall impedance) leads to a stronger shift of the resonances.

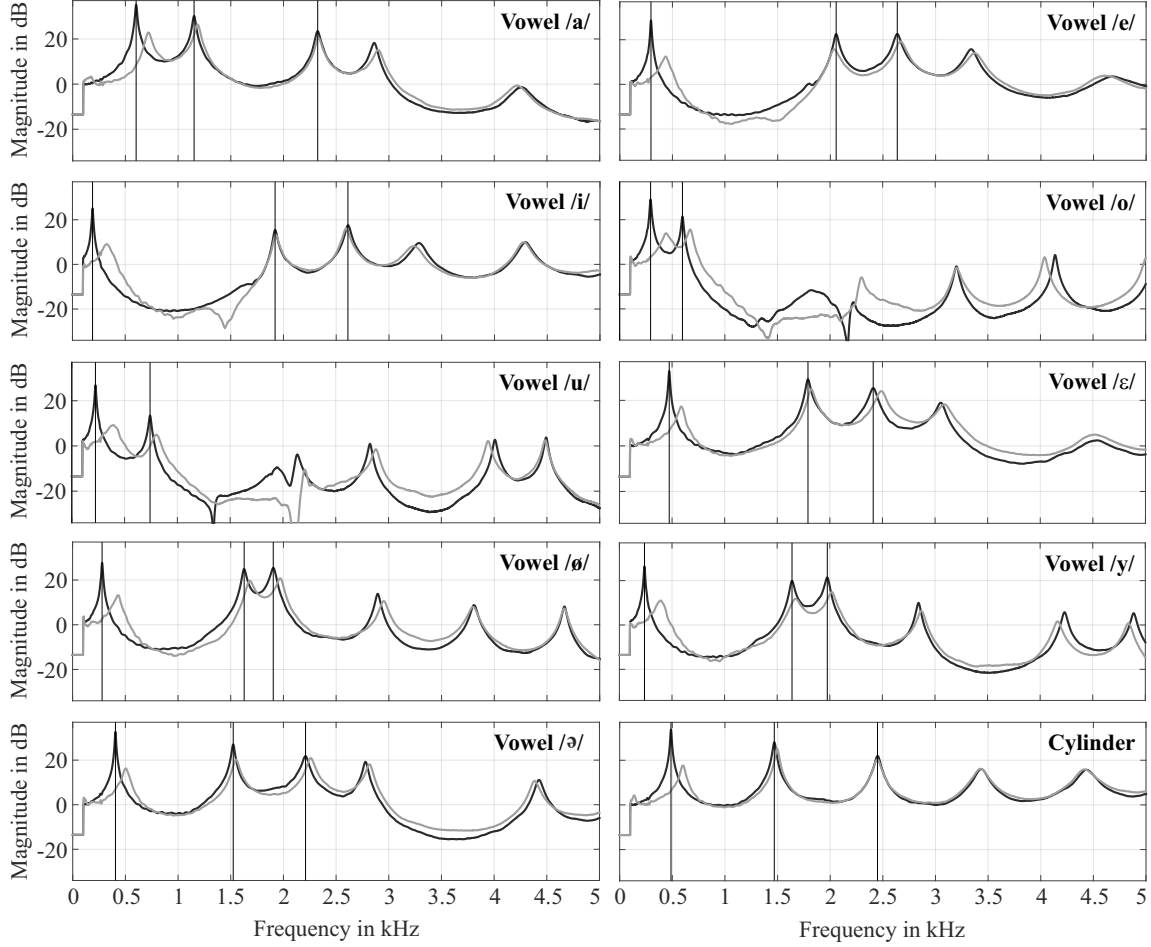The measured *bandwidths* $B_{R1/2/3}$ shown in Fig. 5 b are

Fig. 4. *Measured volume velocity transfer functions of the 10 vowel resonators with hard walls (black) and soft walls (gray).*
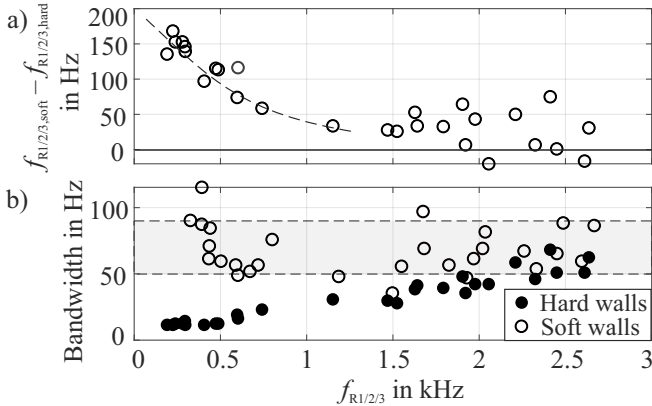

Fig. 5. *a) Shift of $f_{R1/2/3}$ due to soft walls. The abscissa shows the resonance frequencies of the hard-walled resonators. b) Bandwidths $B_{R1/2/3}$ of the resonators with hard walls (black circles) and soft walls (white circles). The shaded area is the bandwidth range of humans [18].*

consistently smaller for the hard models than for the soft models up to 2 kHz. The data up to 800 Hz agree well with the computations for B_{R1} with $\overline{Z} = \infty$ and $\overline{Z} = \overline{Z}_{Flanagan}$ (see Fig. 1 e). Furthermore, the bandwidths of the soft resonators fell well into the range of 50–90 Hz that was reported for the human vocal tract [18], while the bandwidths for the hard models stayed below this range up to 2 kHz.

6. CONCLUSIONS

The results suggest that a lumped-element representation of the wall impedance in computational models of the vocal tract is a valid approximation for physical resonators, at least up to 800 Hz, where the wall impedance has a pronounced effect on the frequencies and bandwidths. With regard to robotic vocal tract models, the used soft silicone material has suitable properties to reproduce the bandwidths measured for the human vocal tract. In future work, it would be interesting to find out to what extent the acoustic effects of soft walls are perceivable (in both physical and computational models) and how they affect the naturalness of the generated speech.

7. REFERENCES

[1] Kishin Migimatsu and Isao T Tokuda, "Experimental study on nonlinear source–filter interaction using synthetic vocal fold models," *The Journal of the Acoustical Society of America*, vol. 146, no. 2, pp. 983–997, 2019.

[2] Peter Birkholz, Falk Gabriel, Steffen Kürbis, and Matthias Echternach, "How the peak glottal area affects linear predictive coding-based formant estimates of vowels," *The Journal of the Acoustical Society of America*, vol. 146, no. 1, pp. 223–232, 2019.

[3] Matthias Echternach, Peter Birkholz, Louisa Traser, Tabea V Flügge, Robert Kamberger, Fabian Burk, Michael Burdumy, and Bernhard Richter, "Articulation and vocal tract acoustics at soprano subject's high fundamental frequencies," *The Journal of the Acoustical Society of America*, vol. 137, no. 5, pp. 2586–2595, 2015.

[4] Bertrand Delvaux and David Howard, "A new method to explore the spectral impact of the piriform fossae on the singing voice: Benchmarking using MRI-based 3D-printed vocal tracts," *PLOS ONE*, vol. 9, no. 7, pp. 1–15, 07 2014.

[5] Stefan Kniesburges, Veronika Birk, Alexander Lodermeyer, Anne Schützenberger, Christopher Bohr, and Stefan Becker, "Effect of the ventricular folds in a synthetic larynx model," *Journal of Biomechanics*, vol. 55, pp. 128–133, 2017.

[6] Patrick Häsner, Andreas Prescher, and Peter Birkholz, "Effect of wavy trachea walls on the oscillation onset pressure of silicone vocal folds," *The Journal of the Acoustical Society of America*, vol. 149, no. 1, pp. 466–475, 2021.

[7] Brian A Pickup and Scott L Thomson, "Flow-induced vibratory response of idealized versus Magnetic Resonance Imaging-based synthetic vocal fold models," *The Journal of the Acoustical Society of America*, vol. 128, no. 3, pp. EL124–EL129, 2010.

[8] Y. Xuan and Z. Zhang, "Influence of embedded fibers and an epithelium layer on the glottal closure pattern in a physical vocal fold model," *Journal of Speech, Language, and Hearing Research*, vol. 57, no. 2, pp. 416–425, 2014.

[9] Takayuki Arai, "The replication of Chiba and Kajiyama's mechanical models of the human vocal cavity," *Journal of the Phonetic Society of Japan*, vol. 5, no. 2, pp. 31–38, 2001.

[10] Peter Birkholz, Steffen Kürbis, Simon Stone, Patrick Häsner, Rémi Blandin, and Mario Fleischer, "Printable 3D vocal tract shapes from MRI data and their acoustic and aerodynamic properties," *Scientific Data*, vol. 7, no. 1, pp. 1–16, 2020.

[11] Kotaro Fukui, Toshihiro Kusano, Yoshikazu Mukaeda, Yuto Suzuki, Atsuo Takanishi, and Masaaki Honda, "Speech robot mimicking human articulatory motion," in *Interspeech 2010*, Makuhari, Japan, 2010, pp. 1021–1024.

[12] Ian S. Howard, "Robotic actuation of a 2D mechanical vocal tract," in *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2017*, Jürgen Trouvain, Ingmar Steiner, and Bernd Möbius, Eds. 2017, pp. 25–32, TUDpress, Dresden.

[13] J. L. Flanagan, K. Ishizaka, and K. L. Shipley, "Synthesis of speech from a dynamic model of the vocal cords and vocal tract," *The Bell System Technical Journal*, vol. 54, no. 3, pp. 485–506, 1975.

[14] G. Fant, L. Nord, and P. Branderud, "A note on the vocal tract wall impedance," *STL-QPSR*, vol. 4, pp. 13–20, 1976.

[15] Kenneth N. Stevens, *Acoustic Phonetics*, The MIT Press, 1998.

[16] M. Fleischer, S. Pinkert, W. Mattheus, A. Mainka, and D. Mürbe, "Formant frequencies and bandwidths of the vocal tract transfer function are affected by the mechanical impedance of the vocal tract wall," *Biomechanics and Modeling in Mechanobiology*, vol. 14, no. 4, pp. 719–733, 2015.

[17] Kenzo Ishizaka, J. C. French, and James L. Flanagan, "Direct determination of vocal tract wall impedance," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 23, no. 4, pp. 370–373, 1975.

[18] Noel Hanna, John Smith, and Joe Wolfe, "Frequencies, bandwidths and magnitudes of vocal tract and surrounding tissue resonances, measured through the lips during phonation," *The Journal of the Acoustical Society of America*, vol. 139, no. 5, pp. 2924–2936, 2016.

[19] H. Wakita and G. Fant, "Toward a better vocal tract model," *STL-QPSR*, vol. 1, pp. 9–29, 1978.

[20] Peter Birkholz, *3D-Artikulatorische Sprachsynthese*, Logos Verlag Berlin, 2005.

[21] Mario Fleischer, Alexander Mainka, Steffen Kürbis, and Peter Birkholz, "How to precisely measure the volume velocity transfer function of physical vocal tract models by external excitation," *PLOS ONE*, vol. 13, no. 3, pp. 1–16, 03 2018.