Considering Lip Geometry in One-Dimensional Tube Models of the Vocal Tract

Peter Birkholz(✉) and Elisabeth Venus

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

Abstract. One-dimensional tube models are an effective representation of the vocal tract for acoustic simulations. However, the conversion of a 3D vocal tract shape into such a 1D tube model raises the question of how to account for the lips, because between the corners of the mouth and the most anterior points of the lips, the cross sections of the vocal tract are open at the sides and hence not well-defined. Here it was examined to what extent simplified tube models of the vocal tract with notches as representations of the lips are acoustically similar to corresponding unnotched models with reduced lengths at the lips end, both with and without teeth. To this end, 3D-printed models of /a, ae, e/ and schwa with different notches and reduced lengths were created. For these, the formant frequencies were measured and analyzed. The results indicate that notched resonators are acoustically most similar to their unnotched counterparts when the length of the unnotched tubes is anteriorly reduced by 50% of the notch depth. However, depending on the formant, vowel, and notch depth, the optimal length reduction can vary between 20–90%.

Keywords: Articulatory speech synthesis · Lip horn · Vocal tract termination

1 Introduction

Despite advanced three-dimensional articulatory models of the vocal tract (e.g., Engwall 2003; Birkholz 2013; Stavness et al. 2012), vocal tract acoustics are mostly simulated in terms of one-dimensional (1D) acoustic tube models that represent the vocal tract as a series of abutting cylindrical tube sections (e.g., Birkholz and Jackèl 2004). The 1D tube models assume plane wave propagation along the vocal tract midline and are much faster to compute than full 3D acoustic simulations. However, the lip region represents a serious difficulty for the 1D approach because the cross-sections of the vocal tract anterior to the corners of the lips are not closed at the sides and hence not defined. This raises the general question of how the "lip horn", i.e., the triangular-shaped space between the corners of the mouth and the most-anterior points of the lips, should be represented in 1D tube models from an acoustic point of view.

From measurements with one subject, Badin et al. (1994) found that the lip horn can be roughly approximated by a single uniform tube section with a length of 11 mm and an area equal to the intra-labial area (in the frontal plane). Lindblom et al. (2007) investigated whether the formant patterns of straight cylindrical tubes with notches at

Q. Fang et al. (Eds.): ISSP 2017, LNAI 10733, pp. 78–86, 2018.
https://doi.org/10.1007/978-3-030-00126-1_8

the "lip end" can be generally reproduced with unnotched tubes of the same diameter and reduced lengths. They found that this was possible when the length of the unnotched tube was the length of the notched tube reduced by about half the notch depth. The optimal length reduction depended somewhat on the tube radius, the notch depth, and the formant index. Lindblom et al. (2010, 2011) furthermore examined physical models with more realistic lip geometries and found that the acoustic effect of the lip horn can be approximated by a length increment that is applied to the last section of the "oral" tube (the tube running from the glottis to the corners of the lips). The length increment is calculated as the "anterior front cavity volume" (the volume anterior to the mouth corners) divided by the cross-sectional area of the most anterior section of the oral tube. However, exactly how the anterior front cavity volume is defined remained unanswered.

In the present study we built upon the investigation by Lindblom et al. (2007), i.e., we examined to what extent it is possible to obtain the formant frequencies of notched tubes with corresponding unnotched tubes of reduced length. We extended the study of Lindblom et al. in two ways:

1. We analyzed not only schwa-like tubes with a constant cross-section, but in addition three two-tube resonators representing the vowels /a, ae, e/.
2. All four resonators were analyzed with and without a row of teeth.

2 Method

2.1 Creation of the Physical Tube Models

In this study we constructed and 3D-printed simplified resonators for the vowels /a, ae, e/ and schwa in nine variants each (one basic variant and eight modified versions). While the basic resonator for schwa was approximated by a single straight cylindrical tube with a constant cross-sectional area of 6 cm^2, /a/, /ae/ and /e/ were approximated by two uniform cylindrical tubes each. The geometries of the basic resonators for /a/ and /ae/ were adopted from Flanagan (1965). The basic geometry of /e/ was based on the /i/ given by Flanagan (1965), but with the cross-sectional area of the anterior tube increased from 1 cm^2 to 2.15 cm^2 to account for the lower tongue position in /e/. We decided to use this /e/ instead of /i/ because the notches for the lips would have unnaturally acute angles for the simplified /i/ resonators. The geometries of all four basic resonators are shown in Fig. 1.

For each vowel, we constructed nine variants as exemplified for the /a/-resonator in Fig. 2. Besides the basic geometry for each vowel, there were four variants where the resonator length was reduced by 1 cm, 2 cm, 3 cm, and 4 cm at the anterior end. The other four variants correspond to the basic resonator with notches of 1 cm, 2 cm, 3 cm, and 4 cm depth at the anterior end. All models contained a small hole of 9 mm diameter at the glottal end to mount a measurement microphone during the acoustic measurements.

Finally, we constructed a simplified row of teeth for each vowel geometry that could be inserted into or removed from the resonators. Based on anatomical data from

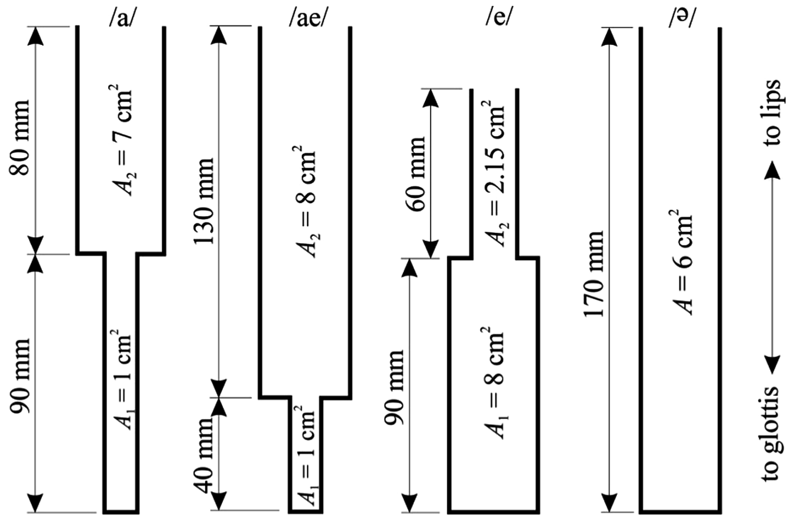

Fig. 1. Dimensions of the four basic vowel resonators.

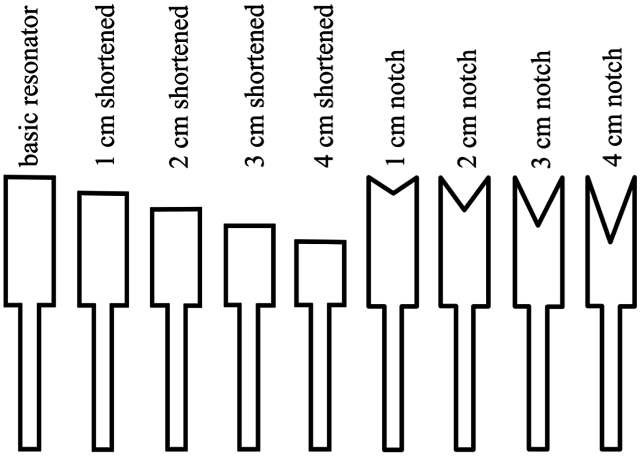

Fig. 2. The nine variants of resonators for /a/.

Slaj et al. (2010), the teeth were modeled as a curved strip where the parabola $y = 0.043 \cdot x^2 - 2.981$ defined the palatal boundary of the teeth row in the axial plane with x and y being the coordinates on the left-right and anterior-posterior axes in mm, respectively, and where $(x, y) = (0, 0)$ is the contact point between the left and right central incisors (see Fig. 3). The teeth strips were 2 mm thick and 1 cm high at the incisors, and inserted at a distance of 2 mm from the most anterior points of the tubes. Figure 3 shows the model variant with the 4 cm notch (maximally spread lips) and inserted teeth, for each vowel.

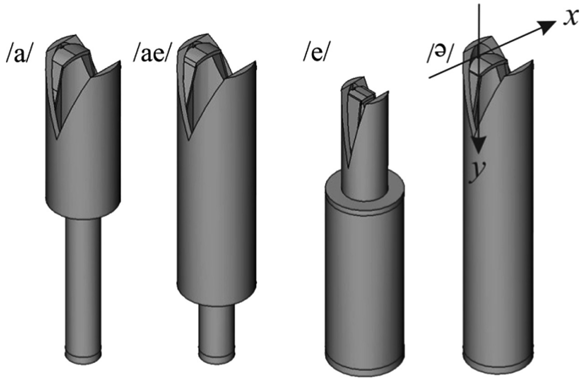

Fig. 3. Resonators with 4 cm notch depth and inserted teeth. The *x* and *y* axes define the coordinate system in which the shape of the row of teeth was defined.

All 4 · 9 = 36 resonators and the teeth rows were manufactured with a 3D-printer (type Ultimaker 2) using the material polylactide (PLA), i.e., the resonators had hard walls. The schwa-resonators were printed in one piece each, and the resonators for /a/, /ae/, and /e/ were printed in two pieces each (“big tube” and “little tube”) and then glued together. The wall thickness of all models was 3 mm, and the walls were 100% filled with PLA (100% infill). In previous experiments we found that less than 100% infill made the model walls less stiff, which may cause undesirable resonances in the acoustic transfer functions due to vibrations of the walls.

2.2 Measurement of Formant Frequencies

For each resonator, we measured the volume velocity transfer functions between the glottis and the lips both with and without teeth. The transfer functions were measured with the recently proposed method by Fleischer et al. (2018) that avoids the difficulties of constructing a volume velocity source for glottal excitation. It extends an idea of Kitamura et al. (2009) that is based on the principle of acoustic reciprocity. The method excites the resonances in a given model with a loudspeaker (VISATON speaker, type FR 10-8 Ω, cone diameter 10 cm) about 30 cm in front of the model that emits a broadband sine sweep with a power band of 50–10.000 Hz into the lip opening of the model. At the same time, the sound pressure $P_g(\omega)$ is measured at the glottal end inside the model using a 1/4” measurement microphone (type MK301E/MV310, www.microtechgefell.de) inserted through a hole at the (otherwise closed) glottal end. After this, the lip opening of the resonator is closed with modeling clay, and a second measurement of pressure $P_m(\omega)$ with the same sweep excitation is performed with a microphone centered around 3 mm in front of the closed lips. According to Fleischer et al. (2018), the ratio $P_g(\omega)/P_m(\omega)$ is exactly the volume velocity transfer function between the glottis and the lips, that is typically used to characterize the acoustics of vowels. The spectral resolution of the transfer functions was 1 Hz, and the first four formant frequencies were determined as the peaks of the magnitude spectrum.

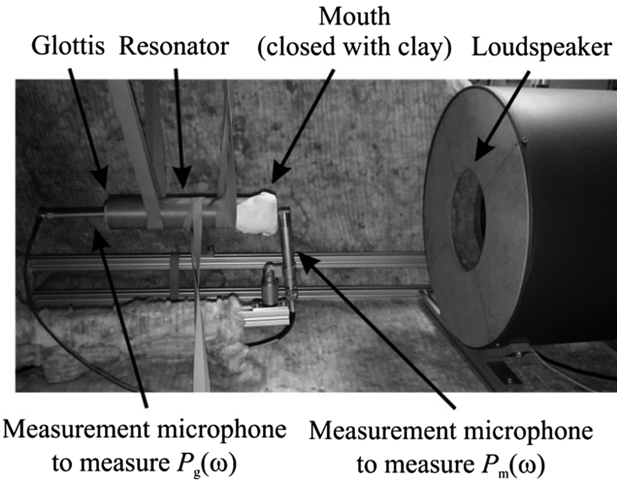

Fig. 4. Setup for the measurement of the acoustic transfer functions, here with a schwa resonator with the lip opening closed with clay to measure $P_m(\omega)$.

The measurement setup is shown in Fig. 4, where the lips have been closed for the measurement of $P_m(\omega)$. All measurements were performed in an anechoic chamber.

Figure 5 shows three of the measured transfer functions for /ae/-resonators with 0 cm, 2 cm, and 4 cm deep notches (without teeth). Apparently, the formant frequencies increase monotonically as the notch depth is increased from 0 cm to 4 cm. Hence, the effect of increasing the notch depth in a tube is similar to the effect that we expect when the length of the tube is reduced (without making a notch).

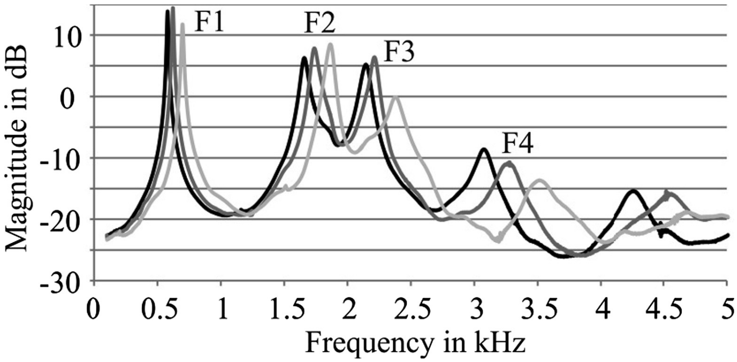

Fig. 5. Measured transfer functions for /ae/ without teeth with 0 cm (black), 2 cm (gray), and 4 cm (light gray) deep notches.

2.3 Calculation of Reduction Factors

For each of the first four formants of the 16 individual notched resonators (4 vowels × 4 notch depths) we determined, based on the acoustic data, what the length reduction of the corresponding basic (unnotched) resonator would need to be in order to produce the same formant frequency. This procedure is illustrated in Fig. 6, which shows the formant frequencies of the /a/-resonators for all four shortening lengths and notch depths, i.e., for all nine tube variants for /a/. The Figure shows that the formant frequencies of the notched resonators (solid lines) are consistently lower than the frequencies of the unnotched resonators with a shortening length equal to the notch depth. For any of the first four formants and any of the four notch depths (1 cm, 2 cm, 3 cm, 4 cm) of the notched resonators, there was one shortening length Δl of the corresponding unnotched resonator that produced the same formant frequency. The gray arrows in Fig. 6 illustrate the calculation of Δl for F_4 of the /a/-resonator with a 3 cm notch. Here, the corresponding shortening length is $\Delta l \approx 1.5$ cm (formant frequencies for shortening lengths between the discrete values of 0 cm, 1 cm, 2 cm, 3 cm, and 4 cm were linearly interpolated). In this way, we calculated a "reduction factor" $k = \Delta l/d$ for each vowel, notch depth d, and formant, both with and without inserted teeth, i.e., $4 \times 4 \times 4 \times 2 = 128$ factors in total. For the example in Fig. 6, the reduction factor is $k \approx 1.5$ cm/3 cm = 0.5. Hence, the notch depth of a notched

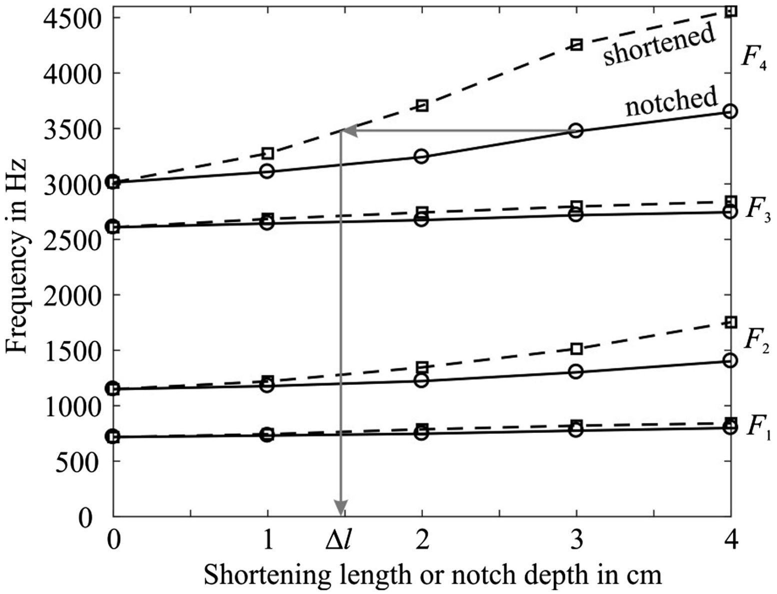

Fig. 6. Measured formant frequencies for the vowel /a/ (without teeth) with the different notch depths (solid lines) and shortened lengths (dashed lines). The gray arrows indicate as an example that for F_4, the resonator with a 3 cm notch is equivalent to the corresponding unnotched resonator shortened by about 1.5 cm.

resonator multiplied by the reduction factor yields the length by which the original unnotched resonator has to be shortened to produce the same formant frequency.

3 Results

Figure 7 illustrates how the reduction factor varies as a function of the notch depth for the vowel /a/. Here, the average reduction factor is 0.55 and varies between 0.23 to 0.75 for a notch depth of 1 cm, and between 0.4 and 0.7 for a notch depth of 4 cm. Furthermore, the overall variation of the reduction factor is greater for the resonators with teeth than for the resonators without teeth. Similar pictures were obtained for the vowels /e/, /ae/, and schwa.

The same trends as shown in Fig. 7 can be observed when the 128 calculated reduction factors are sorted by formants and vowels as in Fig. 8. Here the average factor is also about 0.5. However, the factor varies across vowels and formants, and the variation is considerably greater for models that include teeth (0.2–0.9) as compared to models without teeth (0.3–0.7).

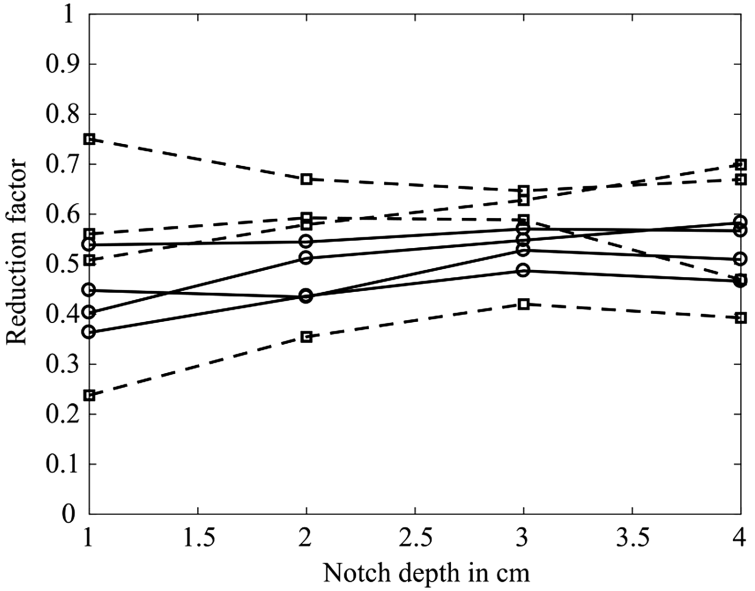

Fig. 7. Reduction factors for the /a/-resonators vs. notch depth. Solid lines: without teeth; dashed lines: with teeth. The four solid lines and the four dashed lines correspond to the four formants.

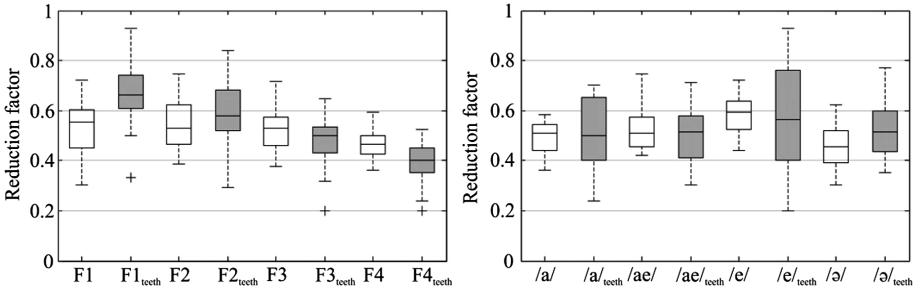

Fig. 8. Calculated length reduction factors for all measured conditions sorted by formants (left) and by vowels (right). Data for the resonators with teeth are shown as the gray boxplots.

4 Discussion and Conclusion

Lindblom et al. (2007) showed that it is possible to reproduce the formant frequencies of a uniform cylindrical tube with a notch at the lip opening with a corresponding unnotched tube, whose length is reduced by about 50% of the notch depth. Here we examined to what extent this finding holds for other resonator shapes beyond a uniform cylindrical tube and for the case that a row of teeth is included in the models. Across all conditions (vowel, formant index, notch depth), the average reduction factor was about 0.5 (50%), in accordance with Lindblom et al. (2007). However, the variation of this factor is considerably greater for resonators with teeth compared to resonators without teeth. Hence, for more realistic vocal tract shapes, the error can become relatively high when the notched lip region is replaced by a cylindrical tube section with half the length of the notch. The greatest variation of reduction factors was obtained for the vowel /e/ with teeth. So it seems that the impact of the teeth on the variation of the reduction factor is greatest for where the cross-sectional areas of the anterior vocal tract are small.

An aspect that was not analyzed here is the effect of a notch on the bandwidths of the formants. A notched lip opening has a bigger radiating surface and hence a bigger radiation impedance than an unnotched lip opening for the same vowel. Therefore, notching can be expected to increase the bandwidths of the formants more as compared to an unnotched case for which the formant frequencies are the same. So, using a shortened unnotched tube to replace a notched lip opening might underestimate formant bandwidths. However, that needs to be explored in detail in future work.

In conclusion, our data suggest that substituting the notch region of a notched tube model by a cylindrical tube section with a length of half the notch depth provides formant frequencies that are roughly equal to those of the notched resonator. The alternative approximation of the lip horn by a fixed-length tube section of 11 mm, as suggested by Badin et al. (1994), will not be optimal, because our data show that the length of the substitute tube section should vary with the notch depth. In future projects, it might be interesting to explore whether the 3D lip region of a resonator can be more accurately represented as a horn-shaped tube section in 1D tube models (instead of a single cylindrical tube section) with an increasing cross-sectional area from the corners of the mouth to the most anterior points of the lips.

References

Badin, P., Motoki, K., Miki, N., Ritterhaus, D., Lallouache, M.T.: Some geometric and acoustic properties of the lip horn. J. Acoust. Soc. Jpn. (E) **15**(4), 243–253 (1994)

Birkholz, P.: Modeling consonant-vowel coarticulation for articulatory speech synthesis. PLoS ONE **8**(4), e60603 (2013). https://doi.org/10.1371/journal.pone.0060603

Birkholz, P., Jackèl, D.: Influence of temporal discretization schemes on formant frequencies and bandwidths in time domain simulations of the vocal tract system. In: Proceedings of the Interspeech 2004-ICSLP, Jeju, Korea, pp. 1125–1128 (2004)

Engwall, O.: Combining MRI, EMA and EPG measurements in a three-dimensional tongue model. Speech Commun. **41**, 303–329 (2003)

Flanagan, J.L.: Speech Analysis, Synthesis and Perception. Springer, Heidelberg (1965)

Fleischer, M., Mainka, A., Kürbis, S., Birkholz, P.: How to precisely measure the transfer function of physical vocal tract models by external excitation. PLoS ONE **13**(3), e0193708 (2018). https://doi.org/10.1371/journal.pone.0193708

Kitamura, T., Takemoto, H., Adachi, S., Honda, K.: Transfer functions of solid vocal-tract models constructed from ATR MRI database of Japanese vowel production. Acoust. Sci. Technol. **30**(4), 288–296 (2009)

Lindblom, B., Sundberg, J., Branderud, P., Djamshidpey, H.: On the acoustics of spread lips. Proc. Fonetik TMH-QPSR **50**(1), 13–16 (2007)

Lindblom, B., Sundberg, J., Branderud, P., Djamshidpey, H., Granqvist, S.: The Gunnar Fant legacy in the study of vowel acoustics. In: Proceedings of the 10ème Congrès Français d'Acoustique, Lyon, France (2010)

Lindblom, B., Sundberg, J., Branderud, P., Djamshidpey, H.: Articulatory modeling and front cavity acoustics. Proc. Fonetik TMH-QPSR **51**(1), 17–20 (2011)

Slaj, M., Spalj, S., Pavlin, D., Illes, D., Slaj, M.: Dental archforms in dentoalveolar Class I, II and III. Angle Orthod. **80**(5), 919–924 (2010)

Stavness, I., Gick, B., Derrick, D., Fels, S.: Biomechanical modeling of English /r/ variants. J. Acoust. Soc. Am. **131**(5), EL355–EL360 (2012)