# Considering lip geometry in one-dimensional tube models of the vocal tract

Peter Birkholz, Elisabeth Venus
*Institute of Acoustics and Speech Communication, Technische Universität Dresden, Germany*
peter.birkholz@tu-dresden.de

Despite advanced three-dimensional articulatory models of the vocal tract (e.g., Birkholz, 2013; Stavness et al., 2012), vocal tract acoustics is mostly simulated in terms of one-dimensional (1D) acoustic tube models to massively reduce computing time. The lip region represents a major difficulty for this approach, because the cross-sections of the vocal tract in front of the corners of the mouth are not closed at the sides. This raises the general question how the 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. Lindblom et al. (2010) investigated whether the formant patterns of cylindrical tubes with notches at the "lip end" can be reproduced with unnotched tubes of the same diameter and of reduced lengths. They found that this was possible when the length of the unnotched tube was roughly the length of the notched tube reduced by about half the notch depth. The precise length reduction depended somewhat on the tube radius, the notch depth, and the formant. In the present study we extended this investigation in two ways: a) we analyzed not only tubes with a constant cross-section, but in addition three two-tube resonators representing the vowels /a, ɛ, e/. b) All four resonators were analyzed with and without a row of teeth.

To this end, each of the simplified tube geometries for the vowels /a, ɛ, e, ə/ given by Flanagan (1972) was physically produced using a 3D printer in nine variants: the original geometry as given by Flanagan, the original geometry with the length reduced at the anterior end by 1 cm, 2 cm, 3 cm, and 4 cm, and the original geometry with notches of 1 cm, 2 cm, 3 cm, and 4 cm depth at the anterior end. In addition, a removable teeth row was printed for each of the models. As an example, Figure 1a shows the four vowels with the maximal notch depth and with inserted teeth. For each model (both with and without teeth) the volume velocity transfer function was measured using the highly accurate method by Fleischer et al. (submitted), and the first four formant frequencies were obtained (see Figure 1b for examples). Based on these data, we determined for each of the first four formants of each of the 16 notched models (4 vowels x 4 notch depths), which length reduction $\Delta l$ of the original resonator geometry was required for its individual formant frequencies to equal those of the corresponding notched model. 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 x 4 x 4 x 2 = 128 factors in total. Formant frequencies for length reductions between the discrete values of 0 cm, 1 cm, 2 cm, 3 cm, and 4 cm of the physical models were linearly interpolated.

Figure 2 shows the calculated reduction factors sorted by formants and vowels. Across all conditions, the factor is on average approximately 0.5 (similar to Lindblom et al., 2010), i.e., a notched tube can be acoustically approximated by a corresponding unnotched tube that is shortened by 50% of the notch depth. 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). Hence, in contrast to the findings for tubes with a constant cross section and without teeth by Lindblom et al. (2010), the representation of the lip region in terms of a simple tube section with reduced length is a rather crude approximation for more realistic vocal tract models that include teeth. Possibly, a matched horn-shaped tube section might be a better 1D approximation of the actual 3D lip region.

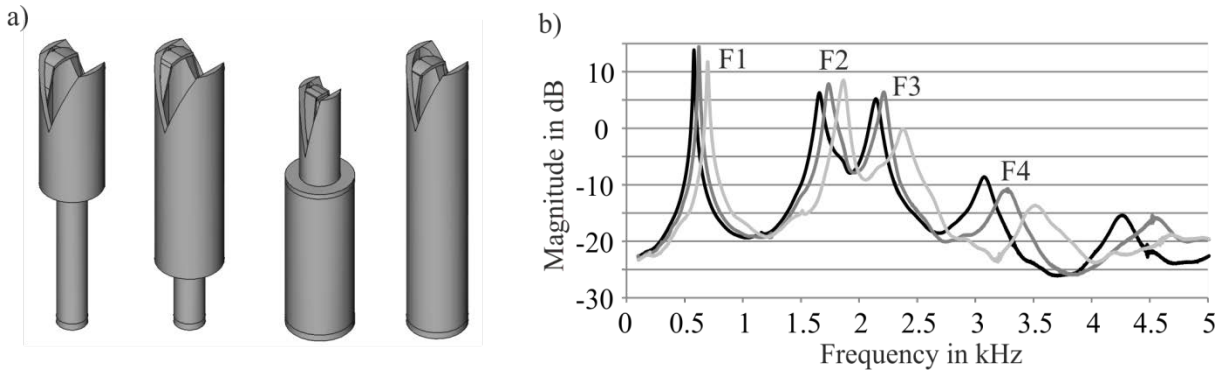


Figure 1: a) Vowel resonators with 4 cm notch depth and inserted teeth. b) Measured transfer functions for /ɛ/ without teeth with 0 (black), 20 (gray), and 40 mm (light gray) deep notches.

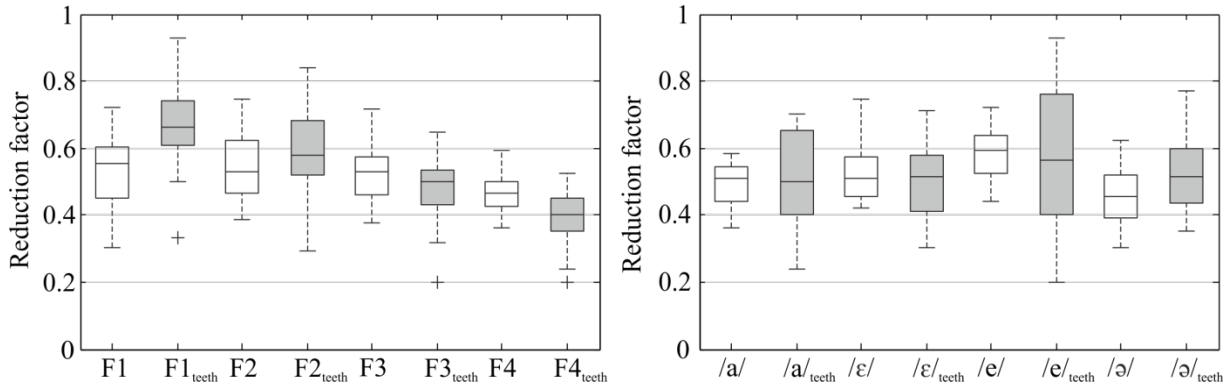


Figure 2: Calculated length reduction factors for all measured conditions sorted by formants (left) and by vowels (right). Data for the resonators *with* teeth are shown as the gray boxplots.

## References


Birkholz P (2013). Modeling consonant-vowel coarticulation for articulatory speech synthesis. PLoS ONE, 8(4): e60603

Flanagan J (1972). *Speech analysis synthesis and perception*. 2nd edition, Springer-Verlag Berlin

Fleischer M, Mainka A, Kürbis S, Birkholz P (submitted). How to accurately measure the transfer function of physical vocal tract models by external excitation. Submitted to *The Journal of the Acoustical Society of America*

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

Stavness I, Gick B, Derrick D, Fels S (2012). Biomechanical modeling of English /r/ variants. *The Journal of the Acoustical Society of America*, 131(5), EL355-EL360.