COMPARISON OF ACOUSTIC PARAMETERS OF INHALATIONS VS. EXHALATIONS WITH 3D-PRINTED VOCAL TRACT MODELS

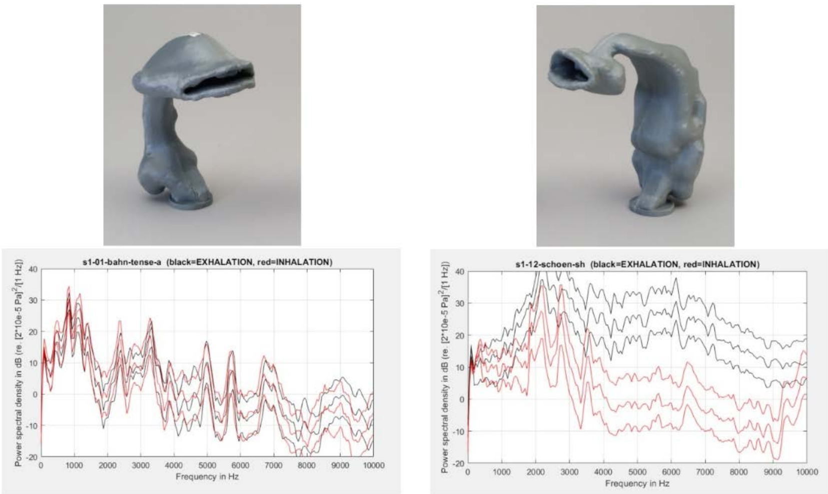
Raphael Werner[1], Susanne Fuchs[2], Jürgen Trouvain[1], Steffen Kürbis[3], Bernd Möbius[1], Peter Birkholz[3]
[1]*Saarland University, Saarbrücken, Germany*
[2]*Leibniz Centre General Linguistics (ZAS), Berlin, Germany*
[3]*TU Dresden, Dresden, Germany*

Our research focuses on the acoustic characteristics of inhalation noises in speech and the underlying articulatory mechanisms. Previously, we have shown (Werner et al., 2021) two main similarities of breathing noise to selected speech sounds: a) enhanced amplitudes at frequencies corresponding to low vowel formants and b) spectral characteristics of voiceless obstruents with a back place of articulation. This comparison is limited in that speech sounds are generally egressive, while inhalation noise is produced with an ingressive airstream. However, the airstream direction might be crucial for the production of noise, downstream of a constriction. To better understand the impact of airstream direction on acoustic properties, we used vocal tract models, producing four vowels and four fricatives. This allows us to analyze the spectra of the radiated sounds concerning a change of airflow direction while keeping the oral configuration constant.

For this, 3D-printed vocal tract models, excluding the nasal cavity, were used that were based on MRI data of a male and female German speaker's vocal tract producing /iː, aː, uː, ɘ, x, ç, ʃ, s/ (Birkholz et al., 2020) (see Fig. 1). To imitate in- and exhalations they were supplied with static airflow through the glottis at three fluid power levels in two airflow directions. Overall, we thus had 96 recordings (8 vocal tract configurations × 2 directions × 2 model speakers × 3 power levels). Although analyzing inhalations using LPC-based formant tracking looked promising (Werner et al., 2021), here we used the averaged power spectral density of the sounds produced over 10 s, as the assumption for a voiced source is not met here. To characterize and compare the sound spectra we calculated coefficients of the Discrete Cosine Transforms (DCT) 0-3. DCT0 corresponds to its mean amplitude, DCT1 to its slope, DCT2 to its curvature, and DCT3 to the amplitudes of the higher frequencies (Jannedy & Weirich, 2017). For each DCT coefficient, we fitted a linear mixed effects model with direction and vocal tract setting, and their interactions, as predictors. The models also included random intercepts for speaker and power level. We used lme4 (Bates et al., 2015) for model fitting and emmeans (Lenth, 2021) for pairwise post-hoc comparisons between in- and exhalations for each configuration. Since we found no significant interaction for DCT3, we used an additive model there.

None of the models returned a main effect for direction. For DCT0, the post-hoc comparisons showed a significant direction contrast for /iː, ç, ʃ, s/ with significantly higher DCT0 values in exhalation for all four. These sounds are produced with high tongue positions that lead to a concentrated airstream hitting the incisors. This obstacle source amplifies the signal in exhalation, but is much weaker in inhalations. For DCT1, inhalation and exhalation were significantly different in /ʃ, s/, and for DCT2 only in /ʃ/.

These results suggest that reversing the airstream direction with a noisy source has segment-specific effects on the spectrum's mean amplitude, slope, curvature, and higher frequencies. Rather than a general effect of direction, differences are found for sibilants, especially /ʃ/, and for mean intensity in settings involving high tongue positions.



Top: two of the 3D-printed vocal tracts corresponding to a male speaker producing the sounds /aː/ (left) and /ʃ/ (right); bottom: power spectral densities (0–10 kHz) of the radiated sounds generated with a static airflow through the glottis at different fluid power levels for the model representing /aː/ (left) and /ʃ/ (right). Exhalations are shown in black, inhalations in red.

References

Bates, D., Mächler, M., Bokler, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), pp. 1–48.

Birkholz, P., Kürbis, S., Stone, S., Häsner, P., Blandin, R., & Fleischer, M. (2020). Printable 3D vocal tract shapes from MRI data and their acoustic and aerodynamic properties. *Scientific Data, 7*(1), pp. 1–16.

Jannedy, S., & Weirich, M. (2017). Spectral moments vs discrete cosine transformation coefficients: Evaluation of acoustic measures distinguishing two merging German fricatives. *The Journal of the Acoustical Society of America, 142*(1), pp. 395–405.

Lenth, R.V. (2021). *emmeans: Estimated Marginal means, aka Least-Squares Means*, r package version 1.6.1.

Werner, R., Fuchs, S., Trouvain, J., & Möbius, B. (2021). Inhalations in Speech: Acoustic and Phys-iological Characteristics. *Interspeech 2021*, pp. 3186–3190.