CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN

12th International Conference on Voice Physiology and Biomechanics, December, 2-4, 2020, Grenoble, France

Towards an open-access database of 3D shapes of the vocal tract and their aero-acoustical properties

Mario Fleischer[1*] & Peter Birkholz[2]

[1]Department of Audiology and Phoniatrics, Charité – Universitätsmedizin Berlin, Berlin, Germany,
[2]Institute of Acoustics and Speech Communication, Technische Universität Dresden, Dresden, Germany

*mario.fleischer@charite.de

Introduction

For a comprehensive and profound understanding of the origin and transmission of acoustic energy in the human vocal system, representative three-dimensional geometric descriptions of the vocal tract (VT) shape are essential. There is a range of potential applications for the provided data. For example, they could be used

- to create articulatory models of the vocal tract or individual articulators based on the provided vocal tract shapes.
- to validate computational acoustic models of the vocal tract, especially simplified (and fast) 1D and 2D models, using the provided vocal tract geometries along with the transfer functions as reference.
- to validate computational aeroacoustic models that simulate noise generation in the vocal tract, using the provided vocal tract geometries along with the turbulence noise measurements as reference.
- to assess the acoustic effects of certain geometric features or side cavities of the vocal tract, like the piriform fossae, the interdental spaces, or the vallecula, based on the 3D-printable models.
- as teaching tools to demonstrate the relationship between vocal tract shape and acoustics, based on the 3D-printable models. With suitable physical mechanisms for the voiced excitation, e.g., a reed pipe as described by Arai [1], the 3D-printed models can be used to synthesize different vowels.
- in combination with other MRI or CT datasets, to study questions of morphology and anatomic development, gender differences, or inter-speaker anatomic or articulatory variability of the vocal tract.

Methods

We determined the air-filled cavities of the VT of one male and one female speaker by using 3D Magnetic Resonance Imaging. The subjects were asked to produce 22 German speech sounds each (16 vowels and 6 consonants). For all resulting 44 models, the teeth of the subjects were added to the raw data before segmentation.

The data here were extensively processed and evaluated to make them accessible to non-experts on volumetric MRI processing.

Based on the processed 3D images, we built a complete set of volume meshes needed for numerical analyses and half-shell 3D-printable models.

All vocal tract models were acoustically characterized in terms of their transfer functions (0-10 kHz) that were determined both numerically with the 3D Finite Element Method, and experimentally by means of an external source excitation paradigm [2]. Furthermore, for the analyses of fricatives and in terms of human recognition of the acoustic outcome, all 3D-printed models were physically excited (synthesized) by using an artificial flow source at the glottis [3]. The processing steps and the derived analyses are summarized in Figure 1.

Finally, to ascertain the perceptual validity of the vocal tract models, a perception experiment was conducted where listeners were asked to identify the phonemes from the sounds that were synthesized with the 3D-printed vocal tract models.

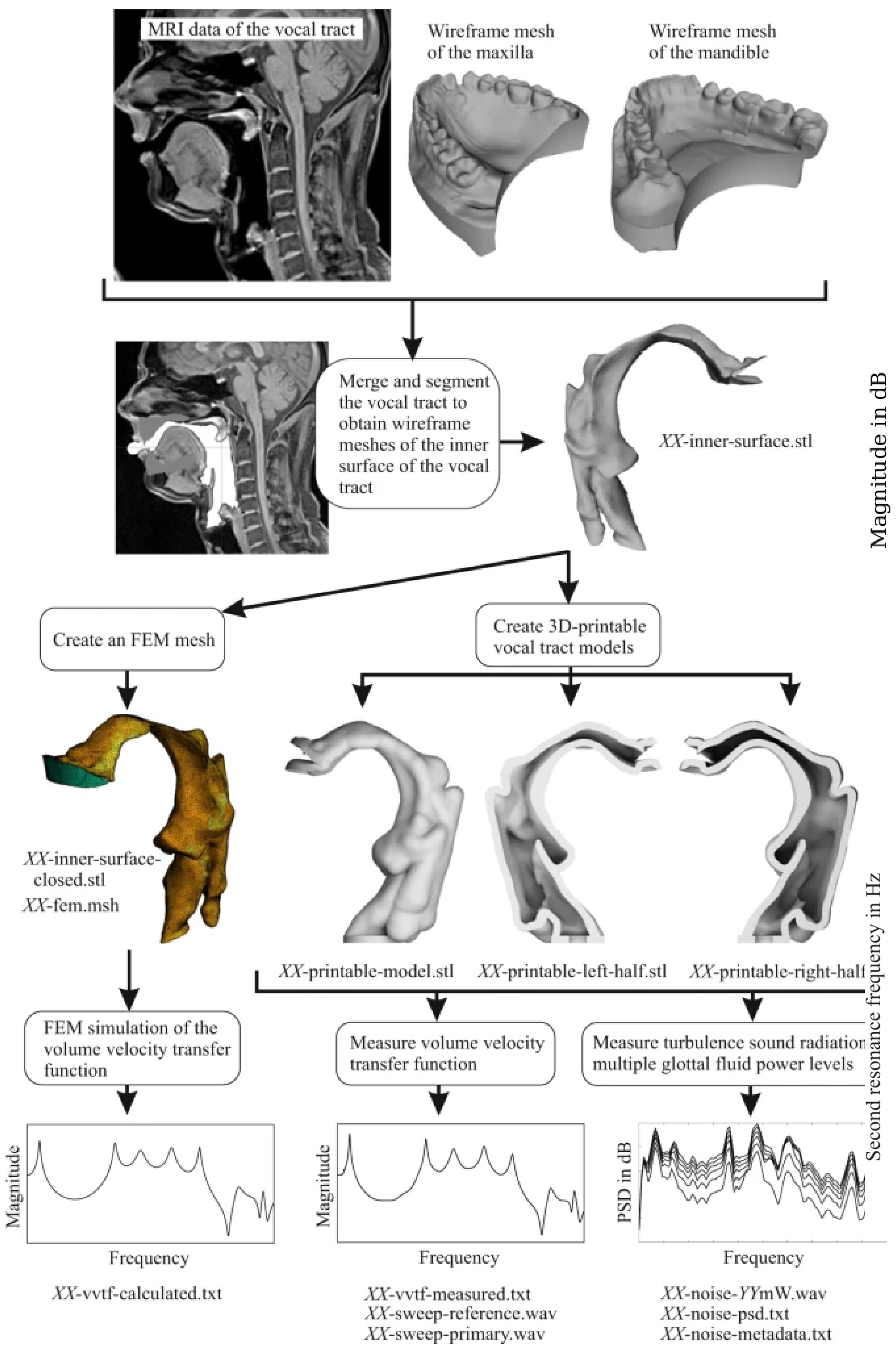

Figure 1: Overview of the data acquisition and processing. The shown images and spectra are for the tense vowel /eː/ of speaker 1. XX in the file names is a placeholder for the speech sound labels.

Results and Discussion

We found an excellent agreement (with an average difference of below 1%) of the experimentally and numerically determined frequency values of the first three vocal tract resonances (f_{R1}, f_{R2} & f_{R3}, Figure 2). Moreover, the first two resonances of the vowels were plausibly arranged in the f_{R2}-f_{R1}-plane and clearly separated from each other (Figure 3).

That indicates that all vowels were produced differently from each other, despite the non-ideal recording conditions in the MRI scanner and the need to artificially sustain the vowels for 12 seconds.

These theoretical findings are supported by the perception experiment. For the male subject, the recognition rate was 68.3% (78.6%) for the model (reference) stimuli of the tense vowels, 53.0% (60.7%) for the model (reference) stimuli of the lax vowels, and 64.4% (82.6%) for the model (reference) stimuli of the fricatives. For the female subject, the recognition rate was 57.0% (77.4%) for the model (reference) stimuli of the tense vowels, 47.3% (71.7%) for the model (reference) stimuli of the lax vowels, and 35.6% (65.8%) for the model (reference) stimuli of the fricatives.

We made all raw data, models and results freely available on an open-access platform to the scientific community. The final state of this project, including findings of flow-induced sound experiments, can be found in [4].

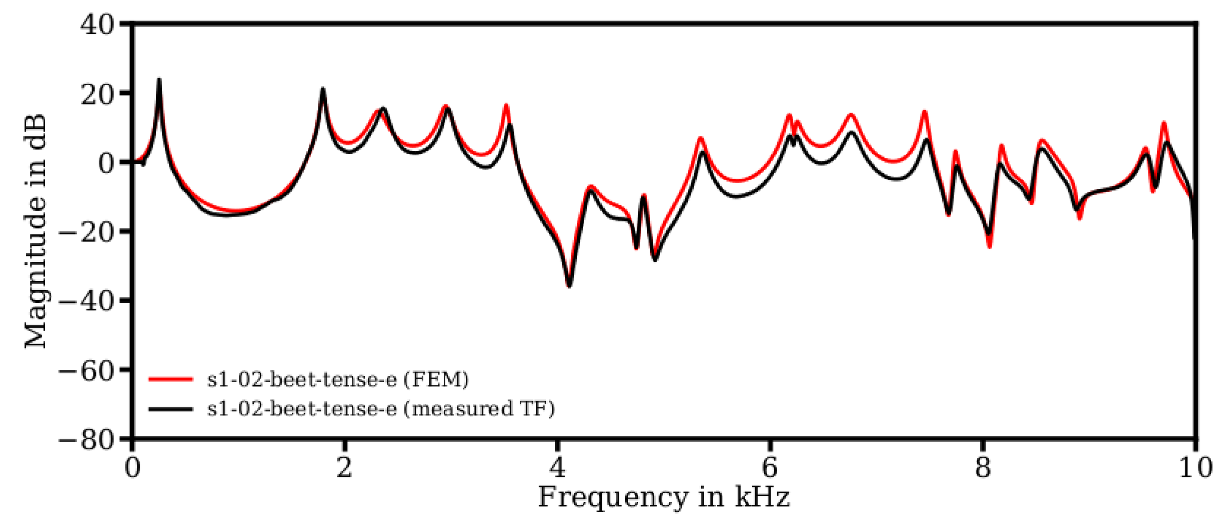

Figure 2: Experimentally and numerically determined volume velocity transfer function for the tense vowel /eː/ of speaker 1.

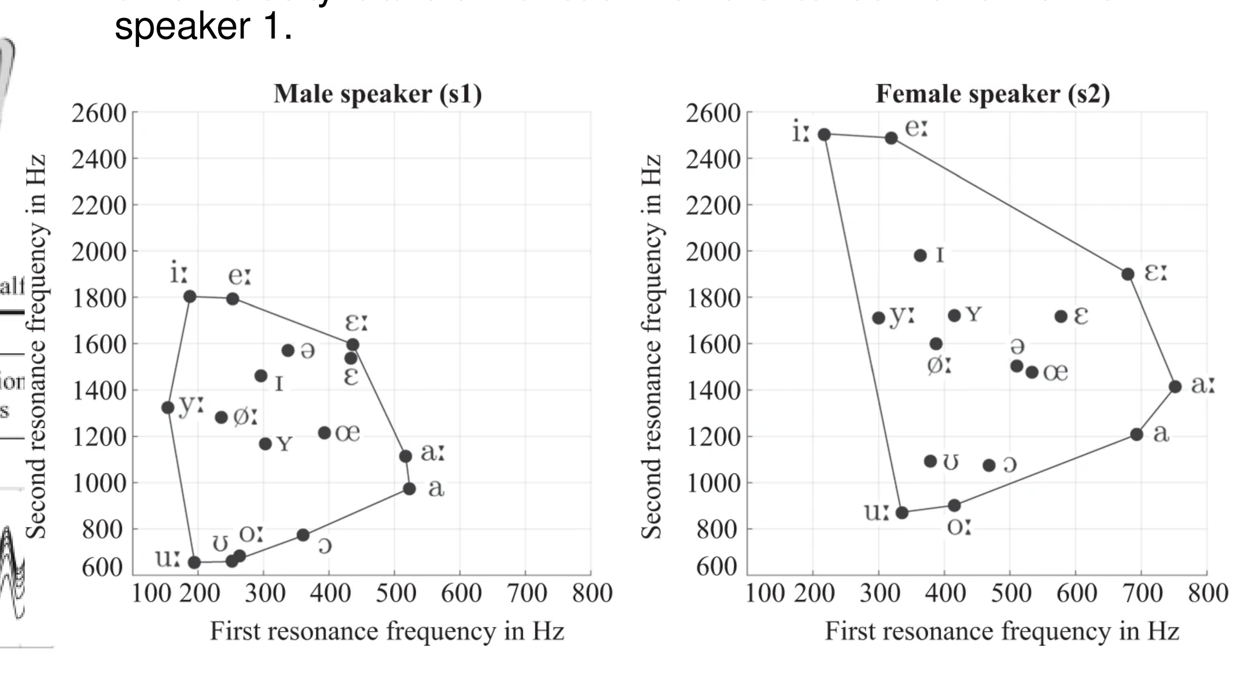

Figure 3: Measured resonance frequencies of the vowels.

References

[1] Arai T. Education in acoustics and speech science using vocal-tract models. The Journal of the Acoustical Society of America. 2012;131(3):2444–2454.

[2] Fleischer M, Mainka A, Kürbis S, Birkholz P. How to precisely measure the volume velocity transfer function of physical vocal tract models by external excitation. PLoS ONE. 2018;13(3):e0193708.

[3] Birkholz P, Gabriel F, Kürbis S, Echternach M. How the peak glottal area affects linear predictive coding-based formant estimates of vowels. The Journal of the Acoustical Society of America. 2019;146(1):223–232. doi:10.1121/1.5116137.

[4] Birkholz P, Kürbis S, Stone S, Häsner P, Blandin R, Fleischer M. Printable 3D vocal tract shapes from MRI data and their acoustic and aerodynamic properties. Scientific Data. 2020;7(1):255.