Algorithmic surface extraction from MRI data:
modelling the human vocal tract

D. Aalto[1,2], J. Helle[3], A. Huhtala[4], A. Kivelä[4], J. Malinen[4], J. Saunavaara[5], T. Ronkka[6]

[1]*Inst. of Behavioural Sciences (SigMe Group), University of Helsinki, Finland*

[2]*Dept. of Signal Processing and Acoustics, Aalto University, P.O. BOX 13000, FI-00076 Aalto, Finland*

[3]*Media Factory, Aalto University, P.O. BOX 31000, FI-00076 Aalto, Finland*

[4]*Dept. of Mathematics and Systems Analysis, Aalto University, P.O. BOX 11100, FI-00076 Aalto, Finland*

[5]*Dept. of Radiology, Medical Imaging Centre of Southwest Finland, University of Turku, Finland*

[6]*Aalto Design Factory, P.O. BOX 17700, FI-00076 Aalto, Finland*

daniel.aalto@iki.fi,{jukka.t.helle, antti.huhtala, atle.kivela, jarmo.malinen, teemu.ronkka}@aalto.fi, jani.saunavaara@tyks.fi

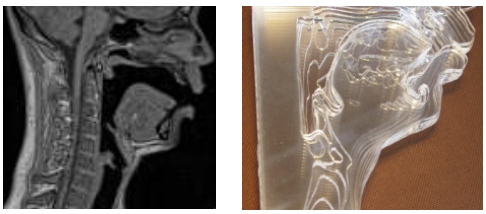
Keywords: MRI, 3D image processing, automatic surface extraction, FEM meshing, physical modelling.

Abstract: An algorithmic approach for 3D voxel MRI data vectorisation is proposed for the high-resolution imaging of the human vocal tract. Because the amount of manual work in data processing is minimised, very large data sets can be processed. The obtained data is used for both mathematical as well as physical modelling of speech and the vocal tract biophysics. A possible future application area is the production of scaffolds for tissue engineering in vocal tract area.

1 INTRODUCTION

We present techniques, algorithms, and results for automatic surface detection, vectorisation, and post-processing of grayscale voxel format output files, produced by Magnetic Resonance Imaging (MRI). There are several multi-purpose software solutions for processing medical images into vectorised surface models with emphasis on cardiovascular structures; see, e.g., MIMICS by Materialise (http://www.materialise.com) that was used in (Takemoto et al., 2010) for vocal tract (VT) extraction. However, when numerous test subjects and large data sets are concerned, the amount of manual labour must be minimised in the data processing. This is the main motivation for developing custom software for VT feature extraction.

Imaging of the human VT is required for accurate *computational modelling* of speech, based on true VT anatomic geometries. Such MRI-generated geometrics can be used for producing *physical models* of the VT in hard plastics or even in elastic materials. Printouts such as shown in Fig. 1(b) have applications in speech acoustics experiments, in designing and testing instrumentation as well as in model-based measurements of flow mechanical phenomena (e.g., conditions such as obstructive sleep apnea or exercise-induced asthma) that are impossible to study non-invasively in test subjects.



(a) (b)

Figure 1: (a) A mid-sagittal section of a male subject while pronouncing vowel [œ]. (b) A plastic printout of the same vowel geometry, consisting of 24 sagittal planes of which 14 is shown here.

Computational models

For the acoustic theory of speech, we consider modelling paradigms based on the *wave equation* and its low-frequency approximation, *Webster's horn equation*. These are respectively given by

$$\phi_{tt} = c^2 \Delta \phi \tag{1a}$$

$$\psi_{tt} = \frac{c^2}{A(s)} \frac{\partial}{\partial s} \left(A(s) \frac{\partial \psi}{\partial s} \right). \tag{1b}$$

As is well-known and explained in (Lukkari and Malinen, 2011), both of these approaches require

high-resolution geometric representation of the VT that must be obtained from test subjects by medical imaging. MRI techniques are preferred because ionizing radiation should be avoided in particular for healthy test subjects. The wave equation requires the detailed *geometry of the air column* in the VT extending from the glottis to the mouth opening and denoted by $\Omega \subset \mathcal{R}^3$ with its boundary $\partial\Omega$ and the air-tissue interface $\Gamma \subset \partial\Omega$. Webster's model is requires the *acoustic centre line* and the *area function* (denoted by $A(\cdot)$ in (1b)) that can be extracted from Ω.

The wave equation and related VT models have been used for studying speech production acoustics for a long time; see, e.g., (Lu et al., 1993; Švancara et al., 2004; Hannukainen et al., 2007; Aalto et al., 2012). Incorporating soft tissues into the models would provide tools for planning and evaluating oral and maxillofacial surgery; see (Dedouch et al., 2002; Švancara and Horáček, 2006). Such comprehensive models are nowadays feasible because of fast and inexpensive computers as well as high-resolution MRI machines.

Physical models

Physical models have a long history in speech research since 19th century and even earlier. For flow mechanical studies, physical modelling is extensively used (see, e.g., (Hirtum et al., 2011; Horáček et al., 2011)) even though Computational Fluid Dynamics (CFD) has replaced it in many cases (see (Takemoto et al., 2010; Šidlof et al., 2012) and the numerous references therein). In addition to speech-related phenomena, the flow mechanical treatment (either by CFD or by physical models) of the VT is required when modelling pathologies such as obstructive sleep apnea or exercise-induced asthma.

In contrast to linear acoustics described in previous section, the flow mechanics in the VT is computationally very demanding unless the model is radically simplified as in (Aalto et al., 2009; Aalto et al., 2011a). Such simplified models ignore compressibility and turbulence, and the boundary dynamics (i.e., the glottis) are realised by low-order mechanical systems. CFD models involving Navier–Stokes equations with elastic boundaries have been developed but they are computationally resource intensive and sometimes unable to account for details such as the true 3D VT geometry or the full closure of the glottis; see (Šidlof et al., 2012, Sections 2.2 and 3.3).

We conclude that there is essential motivation for producing detailed 3D printouts of the VT geometry in various phonetic and/or anatomic configurations. Access to physical models may even be more desirable than just having a computational model when an experimental arrangement must be constructed with existing sensors, actuators, and other instrumentation.

Processing of MRI data from human VT

Due to the very large amount of data, manual processing of the medical images is not feasible. There is a wide literature in image processing and feature extraction methods that are applied in medical imaging, e.g., (Gonzalez and Woods, 2001; Criminisi et al., 2011). However, practical applications (such as imaging the human VT) require custom procedures that must be specifically refined for the particular application. We concentrate on automatic extraction of the VT geometry from MRI images: that is, the tissue-air interface Γ as well as the mouth and the velar port openings that contribute to the full boundary of Ω.

We present methods for automatic removal of artefacts that result from "dry" osseous structures (including teeth) which cannot be separated from air in our MRI experiments. Separately produced teeth models are ultimately to be merged in MRI-generated geometries but this is not discussed in the present work. In order to remove the artefacts due to MRI transparent teeth, we use mathematical morphology and registration algorithms offered by the PCL library to provide alignment information. This is crucial since *solid geometries* are required for models like Eq. (1a) and printouts as shown in Fig. 1(b).

We conclude the paper by discussing further processing of the obtained solid geometry Ω: the extraction of intersection area functions in Section 4 and producing 3D printouts in Section 5.

2 DATA ACQUISITION IN MRI

Measurements are performed on a Siemens Magnetom Avanto 1.5T scanner (Siemens Medical Solutions, Erlangen, Germany). A 12-element Head Matrix Coil and a 4-element Neck Matrix Coil are used to cover the vocal and nasal tracts from the lips and nostrils to the beginning of the trachea. The coil configuration allows the use of Generalized Autocalibrating Partially Parallel Acquisition (GRAPPA) technique to accelerate acquisition (with acceleration factor 2 in all scans). 3D VIBE (Volumetric Interpolated Breath-hold Examination) MRI sequence (Rofsky et al., 1999) is used as it allows for the rapid 3D acquisition required for the experiments. Sequence parameters have been optimized in order to maximise the image resolution and the contrast as well as minimise the acquisition time as explained in (Aalto et al., 2011b, Section 2.3).

The MRI sequence output is stored as a DICOM file that comprises of 44 sagittal plane images: each contains 128×128 pixels that are of size $d = 1.8$mm in the data used for this paper[1]. The alignment of the planar images is carried out using the location data produced by the MRI machine. The bitmaps are stacked to form a 3D matrix of voxel data that represents the VT through grayscale values.

3 SURFACE MODELS FOR THE VOCAL TRACT

In traditional medical imaging, it is usually sufficient to produce visualisations suitable for inspection by trained and experienced radiologists. Hence, automatic feature extraction (without expert manual work) from the same image material is expected to be a difficult problem. On the other hand, solving, e.g., (1a) numerically requires a *solid triangular surface representation* of the wanted interface Γ from which *tetrahedral mesh* of Ω for FEM can be generated.

The air-tissue interface Γ can be extracted by, e.g., *edge detection operators* generalised from 2D image analysis to 3D, see, e.g., (Bomans et al., 1990). However, there is a more widely used approach in 3D image analysis: the gray values of voxel data are regarded as a smooth function $g : \widetilde{\Omega} \to [0,1]$ with $\Omega \subset \widetilde{\Omega} \subset \mathcal{R}^3$, and the *isosurface* $S_\alpha := \{x \in \widetilde{\Omega} : g(x) = \alpha\}$ is extracted for some value $\alpha \in (0,1)$. More precisely, the threshold value α is determined by (1) applying 2D edge detection operations to all of the sagittal 128×128 bitmaps, (2) detecting the threshold gray value from each such image, and (3) finally defining α as the average of these values. The `isosurface` function in MATLAB is used to extract the triangulated representation of $\Gamma = S_\alpha$, based on g and α. We conclude that the isosurface extraction method works very well for air-tissue surface extraction because of the steep grayscale gradient resulting from the carefully optimised MRI sequence parameters.

The algorithm progress can be outlined in the following six steps:

1. **Pre-processing**: The voxel data is smoothed to remove noise.
2. **Initial surface extraction** is carried out by the isosurface method as described above.
3. **Producing the artefact prior model**: "Undesired artefacts" such as vertebrae and maxillae, etc., are manually identified from the *initial surface* to produce an *artefact surface prior model*.
4. **Removing artefacts**: Artefact model is aligned with the initial surface using algorithms provided by PCL (Rusu and Cousins, 2011). Based on the obtained location information, the undesired artefacts are masked from the original voxel data.
5. **Final surface extraction**: The surface Γ is extracted from the artefact-free voxel data by the isosurface method as described above.
6. **Locating boundaries**: The glottis, the velar port, and the mouth opening positions are located from Γ. These openings are covered by triangulated surfaces, and they are joined with Γ to produce a triangulated surface model for the full boundary $\partial\Omega$.

[1]The pixel number is always the same when using this sequence but the pixel size varies according to the physical dimension of the test subject.

Remarks on artefact prior modelling

The prior models for the artefacts are produced by first extracting a surface mesh for a single vowel geometry. Maxillae, vertebrae, and other unwanted parts of the mesh are then manually selected to obtain two artefact prior models, one for each jaw. It is unlikely that this procedure can be automatised since it requires understanding of context. Varying positions of the artefacts is mostly due to the movement of the temporomandibular joint. Hence, using two separate artefact model geometries makes the model alignment by PCL (Step 4 above) more accurate. The variable relative positions of vertebrae are not so significant, and they can be considered as a rigid structure in alignment. Using this approach, the single artefact model can be used for masking unwanted features in all vowel geometries, albeit from the same patient.

Creating artefact models remains the most laborious piece of manual work. It takes about one hour to model the artefacts for one test subject. Many tools can be used to extract the artefact models, such as MeshLab (http://meshlab.sourceforge.net) and Blender (http://www.blender.org). The main reason for using the PCL library for alignment of various geometries is that the MRI machine and the experimental arrangements cannot create sufficient location information for anatomic structures in the data.

Imaging teeth and aligning them with MRI-produced geometries remains an open problem.

Remarks on the mesh

As the final result, we obtain triangulated surface models of VT geometries where undesired artefacts (maxillae, vertebrae) as well as teeth and the nasal tract have been excluded. The side lengths of the

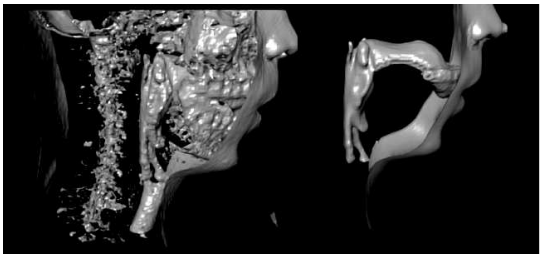


Figure 2: Shaded visualisation of the extracted isosurface before and after the artefact masking. The face is not part of the final computational geometry boundary $\partial\Omega$.

surface triangles are bounded above by $d\sqrt{3}$ where $d = 1.8$mm is the voxel resolution when using the MRI data as described above in Section 2. As reported in (Aalto et al., 2012, Section 4), the geometric error of the surface mesh is of order 0.5mm except in those parts of the model where, e.g., MRI-transparent teeth cause a crude error.

After a solid surface representation of the geometry has been obtained, the mesh generator TetGen (Si, 2011) is used to create a tetrahedral mesh of Ω. After mesh generation FEM can be used to solve (1a) in Ω.

4 AREA FUNCTION AND CENTRE LINE EXTRACTION

The data for Webster's horn model (1b) is derived from the full geometry Ω of the vocal tract, and it consists of the areas $A(s)$ of the intersections $\Gamma(s)$ of Ω with normal planes of a *nominal centre line* $\gamma(s)$ of Ω as shown in Fig. 3. The variable s denotes the distance along the centre line from the glottis end of Ω. Clearly, such centre lines are not *uniquely* defined on geometric grounds even though Ω is of tubular form.

We present a procedure that generates families of $\gamma(s)$'s which are normal to a common set of intersection surfaces $\{\Gamma(s)\}$. Having a family of centre lines is desirable because (1b) depends implicitly on the choice of $\gamma(s)$ (that is, via the arc length parametrisation of the area function $A(\cdot)$), and the choice of the particular $\hat{\gamma}(s)$ from the family can be used to tune the resonances of the simplified model (1b) to match those of more accurate model (1a). We note that the same area function data can be used as well for *Kelly–Lochbaum VT models* (Kelly and Lochbaum, 1962) and for producing the VT transfer functions for *source-filter* -based models (Fant, 1960).

To produce the nominal centre line $\gamma(s)$ in Fig. 3, we find the real nonnegative solution of the Poisson

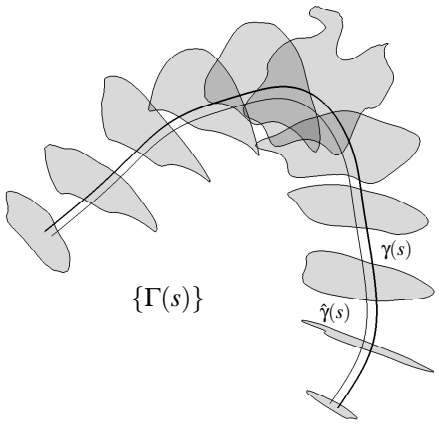


Figure 3: A subset of $\{\Gamma(s)\}$ and two centre lines $\gamma(s)$ and $\hat{\gamma}(s)$ corresponding to the chosen set of area functions.

problem

$$\begin{aligned} &\Delta u = 1 \text{ on } \Omega, \quad u = 0 \text{ on VT walls,} \\ &\frac{\partial u}{\partial \nu} = 0 \text{ at mouth and glottis} \end{aligned} \tag{2}$$

by FEM using the tetrahedral mesh described above. The form of such u can be understood as the steady state temperature distribution with isolation at mouth and glottis and heat sink at tissue boundaries. Hence, by finding and combining the highest points of u gives us the arc-length parametrised curve $\gamma(\cdot)$ that starts from the glottis end, ends up at the mouth end, and keeps strictly away from the walls of VT. The normal planar sections $\Gamma(s_j)$ for sampling points s_j, $j = 1,\ldots,J$, are produced from vectors $\gamma(s_j)$, $\gamma'(s_j)$, and Ω by the usual geometric operations, and the intersection areas $A(s_j)$ by `polyarea` function in MATLAB. Other centre line functions $\hat{\gamma}(\cdot)$ are produced as cubic splines that are normal to all $\{\Gamma(s_j)\}_{j=1,\ldots,J}$. These objects are shown in Fig. 3 for vowel geometry [œ] and $J = 11$.

We remark that the same approach could be used to produce branched centre lines, representing a more complicated topology (e.g., the VT with the nasal tract). The simple Webster's horn model, however, doesn't consider this kind of branching. For pioneering work in area function extraction by MRI, see (Story et al., 1996).

5 3D PRINTOUTS OF THE VOCAL TRACT

We have made preliminary experiments to produce plastic models of the VT in 1:1 scale. We intend to

use these models for acoustic and flow measurements in order to augment and validate the numerical results from mathematical models.

Using Blender (http://www.blender.org), the surface model of the vocal tract was extruded 4mm outward along its surface normals to create the VT wall. The data was then transferred from Blender to Axon 2 by Bits from Bytes, Ltd (http://www.bitsfrombytes.com), in STL format to generate the G-code for the 3D printer.

As in (Takemoto et al., 2010), only hard plastic printouts have been considered so far. Our first trials were performed using a 3D printer 3DTouch by Bits from Bytes, Ltd. The printing time for the full VT in 1:1 scale in PLA plastic was in excess of 12 hours when using a layer height of 0.25mm. Due to the inconvenient geometric shape of the VT, polyvinyl alcohol (PVA) was used as a support material. The long printing time combined with the tendency of PVA to clog the extruder nozzle resulted in a disappointing print quality and a success rate of less than 25%. Should 3D printing be considered as a method for producing a large number of plastic models of the VT, stereolithography as in (Takemoto et al., 2010) or selective laser sintering could produce higher quality models than fused deposition modelling with a far greater success rate.

The prototype model shown in Fig. 1(b) was produced by cutting sagittal intersection contours from 3 mm thick acrylic plate. We used Legend 36ext by Epilog Laser (http://www.epiloglaser.com) which is a 2D CO_2-laser cutter with $P = 60W$. The cutting of all 24 sheets took under 2 hours. Stacking the sheets gives an approximation of the air column extending from below the vocal folds to the mouth opening, excluding the nasal cavity.

The cutting angle of the sheets is always 90° in the model of Fig. 1(b), and this results in significant "stepping" of the VT boundary surface. The stepping can be reduced either by using thinner acrylic plate or, preferably, by varying the cutting angle so as to make the adjacent sheets fit to each other without steps. The variable cutting angle requires using, e.g., a CNC mill or a more advanced cutter than Legend 36ext that has only two degrees of freedom. It is straightforward to produce the G-code file for a CNC mill (including the variable angle data) from the surface geometries $\partial\Omega$ that have been obtained as explained above. Compared to direct 3D printing, one advantage of the cutting or milling method is that there is more choice in the model material.

Optically transparent models (for Particle Image Velocimetry studies as in (Horáček et al., 2011)) cannot be created simply by stacking transparent acrylic sheets but moulds for casting such models can. The same can be said about using elastic materials, too.

We conclude that the human VT is a challenging geometric object for fast prototyping and further experimentation with various technical solutions is required. Our aim is to produce a large number of inexpensive printouts in a wide range of different configurations of the VT, without sacrificing the high quality of vectorised MRI geometries Ω described in above sections.

6 CONCLUSIONS

We have considered automatic surface detection and vectorisation of MR images of the head and neck area, especially concentrating on the vocal tract. In contrast to the work described in (Takemoto et al., 2010, Section II.B), a very large number of MR images are required for model validation as well as medical applications. Hence, the amount of manual work must be minimised without sacrificing the data quality. Apart from some tasks related to artefact detection (see Step 3 in Section 3) and remaining open problems with efficient teeth modelling, the data processing can be carried out by a fully computerised procedure described in this work.

The vectorised data was used in three related purposes: producing detailed geometries for numerical acoustics and flow mechanical computations (Section 3), simplifications for more traditional speech modelling paradigms (Section 4), and using the same data for anatomically accurate physical models for experimental study (Section 5) – this includes the construction of "talking head" sound sources. Clearly, developing algorithms for any one of these application areas gives valuable tools for the other two as well.

In addition to modelling purposes, physical printouts of vocal tract geometries may have direct applications in reconstructive surgery and tissue engineering. Tissue grafts are produced by seeding and attachment of human cells into a *scaffold*. Scaffolds must satisfy many material requirements due to biology (Sachlos and Czernuszka, 2003) as well as have the correct geometric shape, too.

The current version of the software described in this paper can be obtained from the authors by request.

ACKNOWLEDGEMENTS

The authors were supported by the Finnish Academy grant Lastu 135005, European Union grant Simple4All, Aalto Starting Grant, and Åbo Akademi Institute of Mathematics.

REFERENCES

Aalto, A., Aalto, D., Malinen, J., and Vainio, M. (2011a). Interaction of vocal fold and vocal tract oscillations. In *Proceedings of the 24th Nordic Seminar on Computational Mechanics*.

Aalto, A., Alku, P., and Malinen, J. (2009). A LF-pulse from a simple glottal flow model. In *Proceedings of the 6th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA2009)*, pages 199–202.

Aalto, D., Aaltonen, O., Happonen, R.-P., Malinen, J., Palo, P., Parkkola, R., Saunavaara, J., and Vainio, M. (2011b). Recording speech sound and articulation in MRI. In *Proceedings of BIODEVICES 2011*, pages 168–173.

Aalto, D., Huhtala, A., Kivelä, A., Malinen, J., Palo, P., Saunavaara, J., and Vainio, M. (2012). How far are vowel formants from computed vocal tract resonances? Preprint (downloadable from arXiv), 13 pp.

Bomans, M., Hohne, K.-H., Tiede, U., and Riemer, M. (1990). 3-d segmentation of MR images of the head for 3-D display. *IEEE Transactions on Medical Imaging*, 9(2):177 –183.

Criminisi, A., Shotton, J., and Konukoglu, E. (2011). Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Technical Report MSR-TR-2011-114, Microsoft Research.

Dedouch, K., Horáček, J., Vampola, T., and Černý, L. (2002). Finite element modelling of a male vocal tract with consideration of cleft palate. In *Forum Acusticum*.

Fant, G. (1960). *Acoustic Theory of Speech Production*. Mouton, The Hague.

Gonzalez, R. C. and Woods, R. E. (2001). *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2nd edition.

Hannukainen, A., Lukkari, T., Malinen, J., and Palo, P. (2007). Vowel formants from the wave equation. *J. Acoust. Soc. Am. Express Letters*, 122(1):EL1–EL7.

Hirtum, A. V., Pelorson, X., and Estienne, O. (2011). Experimental validation of flow models for a rigid vocal tract replica. *J. Acoust. Soc. Am.*, 130(4):2128–2138.

Horáček, J., Uruba, V., Radolf, V., Veselý, J., and Bula, V. (2011). Aitflow visualization in a model of human glottis near the self-oscillating vocal folds model. *Applied and Computational Mechanics*, 5:21–28.

Kelly, J. and Lochbaum, C. (1962). Speech synthesis. In *Proceedings of the 4th International Congress on Acoustics*, pages Paper G42: 1–4.

Lu, C., Nakai, T., and Suzuki, H. (1993). Finite element simulation of sound transmission in vocal tract. *J. Acoust. Soc. Jpn. (E)*, 92:2577 – 2585.

Lukkari, T. and Malinen, J. (2011). Webster's equation with curvature and dissipation. Preprint (downloadable from arXiv), 22 pp. + 5 pp. appendix.

Rofsky, N., Lee, V., Laub, G., Pollack, M., Krinsky, G., Thomasson, D., Ambrosino, M., and Weinreb, J. (1999). Abdominal MR imaging with a volumetric interpolated breath-hold examination. *Radiology*, 212(3):876 – 884.

Rusu, R. B. and Cousins, S. (2011). 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*.

Sachlos, E. and Czernuszka, J. T. (2003). Making tissue engineering scaffolds work. review: the application of solid freeform fabrication technology to the production of tissue engineering scaffolds. *Eur Cell Mater*, 5:29–39; discussion 39–40.

Si, H. (2011). TetGen: A quality tetrahedral mesh generator and Three-Dimensional Delaunay triangulator. `http://tetgen.berlios.de/`. Accessed Feb. 13th, 2012.

Story, B., Titze, I., and Hoffman, E. (1996). Vocal area functions from magnetic resonance imaging. *J. Acoust. Soc. Am.*, 100(1):537–554.

Takemoto, H., Mokhtari, P., and Kitamura, T. (2010). Acoustic analysis of the vocal tract during vowel productions by finite-difference time-domain method. *J. Acoust. Soc. Am.*, 128(6):3724–3738.

Šidlof, P., Horáček, J., and Řidký, V. (2012). Parallel CFD simulation of flow in a 3d model of vibrating human vocal folds. *Computers and Fluids*.

Švancara, P. and Horáček, J. (2006). Numerical modelling of effect of tonsillectomy on production of Czech vowels. *Acta Acustica united with Acustica*, 92:681 – 688.

Švancara, P., Horáček, J., and Pešek, L. (2004). Numerical modelling of production of Czech wovel /a/ based on FE model of the vocal tract. In *Proceedings of International Conference on Voice Physiology and Biomechanics*.