28th Nordic Seminar on Computational Mechanics
NSCM-28
A. Berezovski, K.Tamm, T.Peets (Eds.)

MOUTH IMPEDANCE OPTIMISATION FOR VOCAL TRACT RESONANCES OF VOWELS

JUHA KUORTTI, JOONAS KIVI, JARMO MALINEN, AND ANTTI OJALAMMI

Aalto University, School of Science,
Department of Mathematics and Systems Analysis,
Otakaari 1, P.O BOX 11000, FI-00076 AALTO,
{juha.kuortti, joonas.kivi, jarmo.malinen, antti.ojalammi}@aalto.fi;
http://speech.math.aalto.fi

Key words: Webster's equation, vocal tract, speech, MRI, FEM

Summary. The vocal tract acoustic resonances can be computed numerically from anatomic geometries (obtained by MRI) using a FEM-based 3D Helmholtz or 1D Webster solver. The numerical results differ from values measured from speech partly due to contribution of the exterior space acoustics. We experiment with a constant acoustic impedance at mouth opening, related to lossy Sommerfeld's radiation condition that is optimised to minimise the resonance discrepancy between computations and measurements. It is observed that by optimisation, the average discrepancy in the three lowest resonances drops from previously obtained 2.5 semitones to 0.9 semitones. Moreover, the optimal impedance have positive real part, and their absolute values corresponds to the mouth opening area in an expected way.

1 INTRODUCTION

It is possible to model speech in high resolution using vocal tract (VT) anatomic configurations from magnetic resonance imaging (MRI). Such computational models have potential applications in, e.g., planning and evaluating oral and maxillofacial surgery.[1,2] One hallmark of high precision is the following: the model should be able to replicate the spectral envelope peaks (known as *vowel formants* $F_1, F_2, \ldots$ in phonetics) of vowel utterance that have been recorded from the same test subject and, if possible, simultaneously during the MRI experiment that produces the VT geometry for computational acoustics.

However, the geometry of the VT is not the only acoustic component that affects the formant frequencies but the acoustic environment plays a role as well. This is particularly significant in speech inside a constrained space that, e.g., the MRI head and neck coils unavoidably are. If the exterior space acoustic is ignored, a vowel and frequency dependent discrepancy appears between the computed VT resonances and their counterparts measured from simultaneously recorded speech.[3] In our earlier experiments that were based on the same data as this study, the average discrepancy for the three lowest formant frequencies F_1, F_2, and F_3 was estimated at 2.5 semitones but this value, of course, contains contributions from other error sources as well.[2]

The purpose of this article is to experiment with the mathematically simplest model for the exterior space acoustics: imposing a *constant acoustic impedance* $\theta = \nu + i\eta$, $\nu, \eta \in \mathbb{R}$, (i.e., an complex-valued impedance that is constant on all frequencies) at the mouth opening as a boundary condition. We use the resonance model of the VT given in Eq. (1) below, based on the generalised Webster's model.[4] This numerically efficient model makes use of the intersectional areas $A(\cdot)$ of the VT, and it does not take into account non-longitudinal standing waves at all. Hence, it can be used only for the three lowest formants that all lie under 4 kHz, whereas treating the higher cross-modes would require a Helmholtz solver in 3D.[5,6] Since most of the vowel information is contained in the two lowest formants F_1 and F_2, Webster's model can, however, be used to obtain phonetically relevant information.

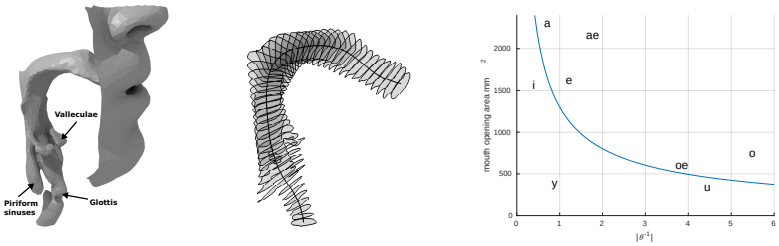

Figure 1: Left: A surface model of the VT and face for vowel [œ]. Middle: The slicing of the VT used for obtaining the area function $A(\cdot)$ and other parameters required for Eq. (1). Right: The mouth opening area $A(\ell)$ as a function of optimal impedance $|\theta_{opt}|$ for all Finnish vowels [ɑ, e, i, o, u, y, æ, œ].

In this study, we optimise the (normalised) acoustic impedance θ so that the lowest resonance frequencies $R_1(\theta), R_2(\theta)$, and $R_3(\theta)$ from Eq. (1) match the target formants F_1, F_2, and F_3 that have been extracted from speech signals. Solving the spectral inversion problem yields different optimal $\theta = \theta_{opt}$ for each vowel. Using such θ_{opt} for each vowel separately, the average formant discrepancy drops to 0.9 semitones as reported in Tables 1. Recalling the parallel coupling law of impedances, we expect $|\theta_{opt}|$ to be inversely proportional to the mouth opening area $A(\ell)$. Also this is confirmed reasonably well as shown in Fig. 1.

2 WEBSTER'S RESONANCE EQUATION

We use the generalised Webster's horn model[4] for modelling the VT acoustics in each of the Finnish vowel configurations [ɑ, e, i, o, u, y, æ, œ]. The resonances of the model can be computed from the eigenvalue problem

$$\begin{aligned} &\left(\frac{{\lambda_\theta}^2}{c^2\Sigma(s)^2} + \frac{2\pi\alpha W(s)\lambda_\theta}{A(s)}\right)\psi_{\lambda_\theta} = \frac{1}{A(s)}\frac{\partial}{\partial s}\left(A(s)\frac{\partial\psi_{\lambda_\theta}}{\partial s}\right) \quad \text{on} \quad [0,\ell], \\ &\left(\frac{\lambda_\theta}{c}\right)\psi_{\lambda_\theta}(0) - \frac{\partial\psi_{\lambda_\theta}}{\partial s}(0) = 0, \quad \text{and} \quad \left(\frac{\lambda_\theta}{c}\right)\psi_{\lambda_\theta}(\ell) + \theta\frac{\partial\psi_{\lambda_\theta}}{\partial s}(\ell) = 0. \end{aligned} \tag{1}$$

Here $A(\cdot)$ denotes the cross-sectional areas of the VT volume from MRI as shown in Fig. 1, parameterised by the arch length of its centreline with length ℓ.[7] The

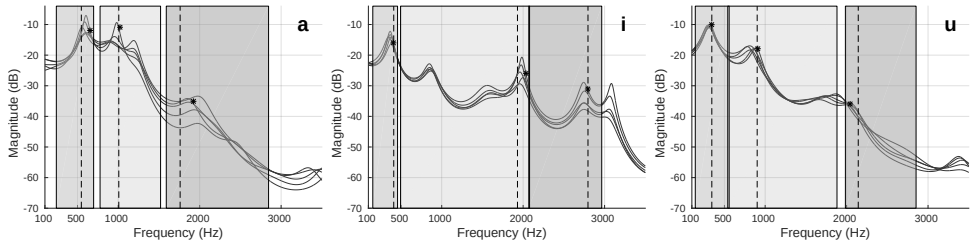

Figure 2: The measured spectra of Finnish vowels [ɑ, i, u] and their formant peaks, marked by asterisks. The dashed lines are the optimised resonances $R_j(\theta_{opt})$ for $j = 1, 2, 3$, from Webster's model whose range (over all feasible θ) is also indicated.

parameters $\Sigma(\cdot)$ and $W(\cdot)$ relate to the curvature of the VT, $c = 343$ m/s is the speed of sound in air, and $\alpha = 7.6{\cdot}10^{-7}$ s/m is related to the energy dissipation into tissues.[4] We use the constant complex impedance θ at mouth (i.e., $s = \ell$) as a tunable model for the exterior acoustics, inspired by a lossy version of Sommerfeld's radiation condition. The lowest resonance frequencies $R_j(\theta)$ for $j = 1, 2, 3$, are obtained from the imaginary parts of the smallest eigenvalues $\lambda_\theta = \lambda_\theta(j)$ of Eq. (1) after normalisation by 2π.

3 FORMANT EXTRACTION

In speech signals, formants can discriminated from harmonic overtones of the glottal frequency f_0 since formants have much wider bandwidth, and they can be extracted from a spectral envelope.[2,8] The spectral envelope and associated peaks can be obtained by solving the Yule–Walker equations $\hat{\mathbf{R}}_{xx}\hat{\mathbf{a}} = \hat{\mathbf{r}}_x$ where $\hat{\mathbf{R}}_{xx}$ is the autocorrelation estimate of the speech signal. The coefficient vector $\hat{\mathbf{a}}$ defines a forward predictor polynomial $A(z) = a_0 + \sum_k \hat{a}_k z^{-k}$ whose zeroes define the poles of an all-pole IIR filter $H(z) = A(z)^{-1}$. By plotting $10\log(|H(i\omega)|)$ for $\omega \in [0, \pi]$, we obtain the spectral envelopes in Fig. 2 where peaks correspond to formant frequency estimates F_j for $j = 1, 2, 3$.

The formant analysis of speech during MRI may result in extra poles for $H(z)$ due to exterior resonances within the MRI coils, residual acoustic noise from the MRI machine, and issues related to signal processing. Further, because speech during MRI contains a substantial amount of noise, specialised signal processing is necessary.[2,8]

	[ɑ]	[e]	[i]	[o]	[u]	[y]	[æ]	[œ]	abs. avg
D_1	-0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
D_2	0.0	0.0	-0.9	0.0	0.0	-0.7	1.8	2.2	0.7
D_3	-3.0	1.1	0.0	3.5	1.4	1.8	2.8	2.3	2.0

Table 1: Discrepancies D_j (in semitones) of vowel formants F_j and optimised resonances $R_j(\theta_{opt})$ for $j = 1, 2, 3$. The discrepancies are given by $D_j = 12\ln(R_j(\theta_{opt})/F_j)/\ln 2$. Except for [y], the vowel formants have been estimated from simultaneously recorded MRI/speech data pairs at the glottal frequency $f_0 = 104$ Hz. Due to artefacts in the signals recorded during MRI, the formants for [y] have been extracted from a recording in anechoic chamber from the same test subject.

4 IMPEDANCE OPTIMISATION AT MOUTH

We solve the forward problem Eq. (1) numerically by using 120 piecewise linear elements for a discrete number of $\theta \in [0.1, 10]$, resulting in values of $R_j(\theta)$ as explained in Section 2. The optimal $\theta = \theta_{opt}$ is chosen by minimising the total discrepancy $D_{tot}(\theta) := \sum_{j=1,2,3} |\ln(R_j(\theta)/F_j)|$ for each of the Finnish vowels [ɑ, e, i, o, u, y, æ, œ]. The resulting discrepancies are given in Table 1 for simultaneously recorded MRI/speech data pairs as targets F_j for $j = 1, 2, 3$.

The stages described in Sections 2 – 4 have been realised in MATLAB R2015a.

5 RESULTS AND CONCLUSIONS

The results in Table 1 are comparable with our earlier results [2] since the same MRI and speech data has been used in both studies. The average discrepancy in Table 1 is 0.9 semitones, representing an improvement of 1.5 semitones compared to the earlier results where exterior acoustics was modelled by the Dirichlet boundary condition at the mouth opening. For the two lowest formants, the average discrepancy is only 0.3 semitones. This is a quite satisfactory outcome, bearing in mind the extreme simplicity of the constant complex impedance model and the fact that natural test subject related variation is already in the class of 0.5 semitones.

REFERENCES

[1] Aalto, D. *et al.* Recording speech sound and articulation in MRI. In *Proceedings of BIODEVICES 2011*, 168–173 (2011).

[2] Aalto, D. *et al.* Large scale data acquisition of simultaneous MRI and speech. *Appl Acoust* **83**, 64–75 (2014).

[3] Kivelä, A. *Acoustics of the vocal tract: MR image segmentation for modelling.* Master's thesis, Aalto University School of Science, Department of Mathematics and Systems Analysis (2015).

[4] Lukkari, T. & Malinen, J. Webster's equation with curvature and dissipation. arXiv:1204.4075 (2013). Submitted.

[5] Hannukainen, A., Lukkari, T., Malinen, J. & Palo, P. Vowel formants from the wave equation. *J Acoust Soc Am* **122**, EL1–EL7 (2007).

[6] Kivelä, A., Kuortti, J. & Malinen, J. Resonances and mode shapes of the human vocal tract during vowel production. In *Proceedings of 26th Nordic Seminar on Computational Mechanics*, 112–115 (2013).

[7] Aalto, D. *et al.* Algorithmic surface extraction from MRI data: modelling the human vocal tract. In *Proceedings of BIODEVICES 2013*, 257–260 (2013).

[8] Kuortti, J. & Malinen, J. Post-processing speech recordings during MRI. arXiv:1509.05254 (2015).