Proceedings of the 26th Nordic Seminar on Computational Mechanics
A. Logg, K.A. Mardal (Eds.)

Resonances and mode shapes of the human vocal tract during vowel production

Atle Kivelä[1], Juha Kuortti[1], and Jarmo Malinen[1]

(1)Aalto University, School of Science, Department of Mathematics and Systems Analysis, P.O. BOX 11100, FI-00076 AALTO, {atle.kivela, juha.kuortti, jarmo.malinen}@aalto.fi

Summary. We show that higher vowel formants may consist of multiple acoustic resonances. We compute the Helmholtz resonances of the vocal tract (VT) using FEM. The resulting resonance and pressure information is then compared to formants extracted from recorded vowel phonation. It is observed that the pure longitudinal modes appear in the sequence of all acoustic modes in an order that depends on the vowel even in a single test subject.

Key words: vocal tract, resonance, formants, FEM, MRI

Introduction

It is well known that the two or three lowest vowel formants $F_1, F_2, \ldots$ (in increasing order of frequency) are sufficient to distinguish between vowels in most cases. The formants, determined from sound samples, correspond closely to the acoustic resonance frequencies $R_1, R_2, \ldots$ of the human vocal tract (VT). The relevance of the higher formants and resonances (i.e., $n \geq 4$) has been a subject of study for a long time. The most recent findings are due to computational modelling of the VT acoustics in 3D geometries that have been obtained using X-ray computer tomography (CT) [7, 8].

The purpose of this article is to model VT acoustics using a 3D Helmholtz resonance model and Finite Element Method (FEM) for its numerical solution. Resonances obtained from the Helmholtz model are compared to (i) the resonances obtained from Webster's horn model with one spatial dimension, and (ii) the spectral envelopes extracted from sound samples from the same test subject. We observe that the Finnish vowel [œ] has transversal components in resonances R_4, R_5, R_6, and these form a cluster that gets experimentally recognised as just a single formant F_4. In this data, the lowest transversal resonance takes always place in piriform sinuses and not in the mouth cavity, see [6]. We remark that the positions of the longitudinal resonance frequencies in the sequence of all resonance frequencies vary, depending on the vowel geometry even in the same subject (see Fig. 1).

The computational geometries for this article have been obtained using Magnetic Resonance Imaging (MRI). Using MRI is not as straightforward as CT since osseous structures (such as maxillae and teeth) do not show up in the images. Further, the MRI machine produces a lot of acoustic noise which makes simultaneous speech recording more difficult [2, 3]. However, MRI can be used to make a large number of measurements from the same subject not having a medical condition that would justify the use of ionising radiation.

Methods

We use 3D geometries of the VT and a MATLAB-based FEM solver for the Helmholtz equation (1). The raw MRI data is first vectorised as described in [1]. In addition to Eq. (1), more simple acoustic resonance model Eq. (3) is used. The latter model accounts only for the longitudinal acoustic resonances. Hence, identification of pure longitudinal acoustic resonances from Eq. (1) can be carried out by comparison, without inspecting 3D eigenfunctions of Eq. (1) individually. Finally, Linear Predictive Coding (LPC) is used to determine spectral envelopes from recorded sound samples, and their peaks are compared to the computed resonances $R_1, R_2, \ldots$.

Helmholtz equation

We solve the acoustic resonances by first solving the Helmholtz problem

$$\begin{cases} \lambda^2\Phi_\lambda = c^2\Delta\Phi_\lambda \quad \text{on} \quad \Omega, \quad \Phi_\lambda = 0 \quad \text{on} \quad \Gamma_1, \\ \frac{\partial\Phi_\lambda}{\partial\nu} = 0 \quad \text{on} \quad \Gamma_2, \quad \text{and} \quad \lambda\Phi_\lambda + c\frac{\partial\Phi_\lambda}{\partial\nu} = 0 \quad \text{on} \quad \Gamma_3 \end{cases} \tag{1}$$

where the speed of sound is denoted by c, $\Omega \in \mathbb{R}^3$ is the interior volume of the VT whose boundary $\partial\Omega = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$ consists of the mouth opening Γ_1, the VT tissue walls Γ_2, and the (virtual) control surface Γ_3 right above the glottis. The exterior normal derivative is denoted by $\frac{\partial}{\partial\nu}$ and the solution Φ_λ is the velocity potential, giving the acoustic pressure by $\rho\lambda\Phi_\lambda$ where ρ is the density of air.

The variational formulation of Eq. (1) can be turned into a quadratic eigenvalue problem as shown in [4]. The matrices for the eigenvalue problem are constructed using FEM with piecewise linear shape functions and tetrahedral meshes with approximately 10^5 elements. We obtain the stiffness matrix $\mathbf{K}$, the mass matrix $\mathbf{M}$, and the matrix $\mathbf{P}$ which presents the absorbing glottis boundary condition in Eq. (1). The discretised eigenvalue problem can be written as

$$\lambda^2\mathbf{K}\mathbf{x}(\lambda) + \lambda c\mathbf{P}\mathbf{x}(\lambda) + c^2\mathbf{M}\mathbf{x}(\lambda) = 0. \tag{2}$$

The imaginary parts of the smallest eigenvalues give resonance frequencies $R_1, R_2, \ldots$ (in increasing order of frequency), and the longitudinal resonances are identified from these by comparing them to the resonances obtained from (3).

Resonance version of Webster's horn model

Using the 3D wave equation is impractical for modelling acoustics of the VT in high temporal and spatial resolution. The classical Webster's horn model is computationally much lighter, and in some special geometries even analytic solutions are possible [9]. For Webster's model, a tubular domain is defined by a parameterised centreline and corresponding cross-sectional areas $A(s)$ as shown in Fig. 2. The resonances of Webster's horn model can then be obtained from the eigenvalue problem

$$\lambda^2\psi_\lambda = \frac{c^2\Sigma(s)^2}{A(s)}\frac{\partial}{\partial s}\left(A(s)\frac{\partial\psi_\lambda}{\partial s}\right) \quad \text{for} \quad s \in [0, \ell], \tag{3}$$

where ℓ is the length of the VT, and the function $\Sigma(s)$ has been introduced in [5] for taking into account the centreline curvature. The Webster's velocity potential ψ_λ corresponds to the 3D velocity potential given by (1). We solve Eq. (3) numerically by using piecewise linear elements and end point conditions at $s = 0, \ell$ corresponding to the boundary conditions in Eq. (1).

Eqs. (1)–(3) are solved numerically using sufficiently many elements so that all computed resonances can be regarded as accurate.

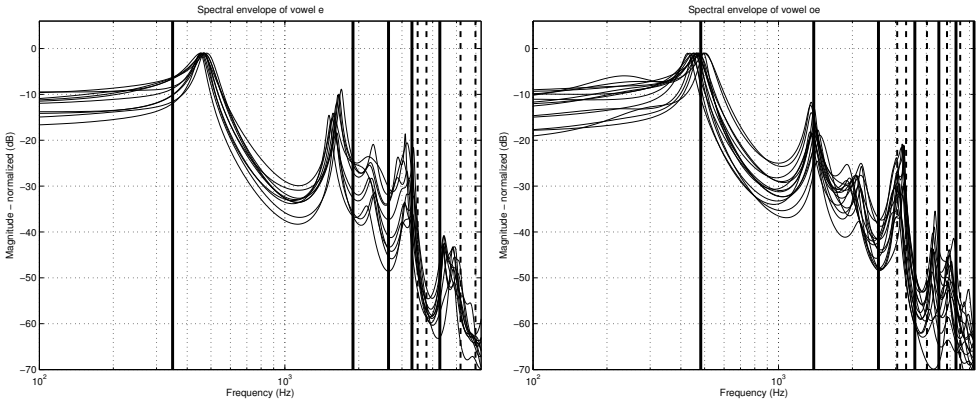

Figure 1. The spectral envelopes of Finnish vowels [œ,e] recorded in an anechoic chamber. These curves differ only slightly from those recorded during the MRI [3]. The solid vertical lines represent longitudinal resonances while the dashed lines represent transversal resonances, both computed from Eq. (1).

LPC

Spoken vowels can be identified based on wide bandwidth acoustic energy concentrations around discrete frequencies in the power spectrum of the measured signal. Such energy peak locations $F_1, F_2, \ldots$ (in increasing order of frequency) are called formants, and they are related to the frequencies $R_1, R_2 \ldots$ computed from Eq. (1). In measured signals, formants can discriminated from harmonic overtones of the fundamental glottal frequency by the fact that formants, indeed, have much wider bandwidth. Thus, smoothing of the power spectrum allows the formants to be usually extracted with good accuracy.

The most popular method of spectral smoothing in speech science is LPC. LPC is a forward predictor FIR filter whose frequency response gives an estimate for the spectral envelope of the signal; the poles of the filter can be regarded as a practical definition of formant locations. LPC is mathematically equivalent to fitting a low-order rational function $R(s)$ to the power spectrum function on the imaginary axis. Thus, plotting the values of $|R(i\omega)|$ for real ω yields the desired smoothing of the power spectrum, and formant frequencies $F_1, F_2, \ldots$ correspond then to the imaginary parts of the poles of $R(s)$.

Results

The LPC spectral envelopes and resonance frequencies $R_1, R_2, \ldots$ up to 4 kHz for Finnish vowels [œ, e] are shown in Fig. 1. Some of the resonances are identified as longitudinal by Webster's horn model, and they are plotted with solid lines in Fig. 1. The remaining resonances from Eq. (1) are regarded as transversal, and this behaviour can be observed from the acoustic pressure distributions shown in Fig. 2.

Results presented here and in [7, 8] confirm that higher formants may consist of a cloud of longitudinal and transversal resonances. This clustering phenomena appears to be significant from the fourth formant onwards. As expected, the first and second formants can be reliably identified with the resonances of Webster's horn model.

Conclusions

The resonances from Eq. (1) differ somewhat from those extracted from experimental data. Apart from the lowest formant F_1 (where the formant extraction from speech signal is most error-prone), the discrepancy is likely to be caused by the unrealistic acoustic impedance at mouth in Eq. (1) as discussed in [4].

The full classification of the resonance structure of the VT in various vowel configurations requires additional effort and a combined MRI/sound data set from several test subjects. An-

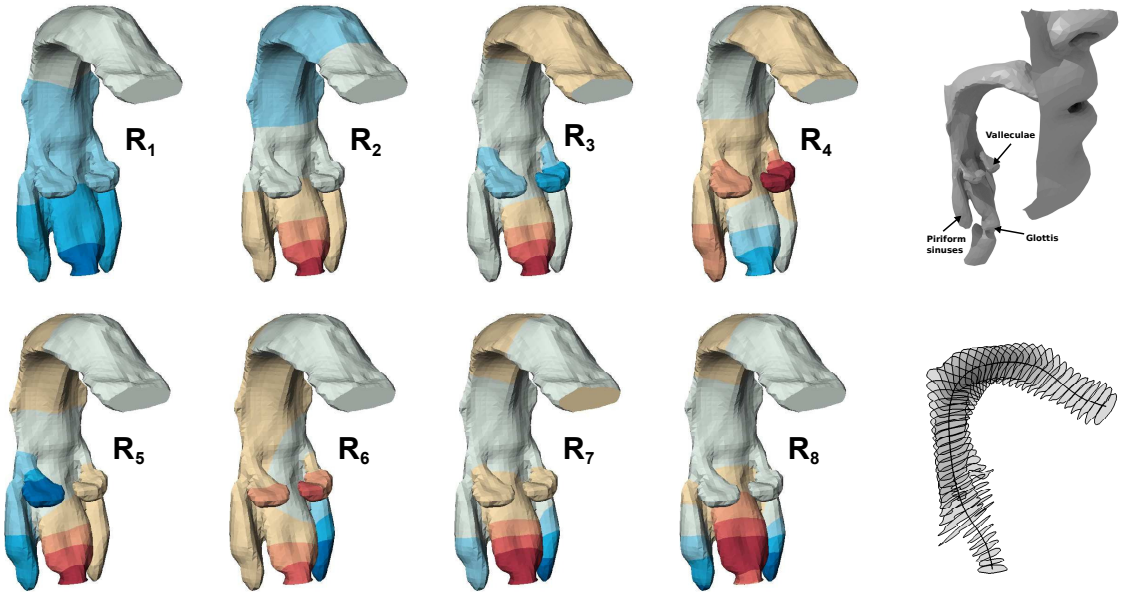

Figure 2. Modes of the first eight resonances for vowel [œ], the surface mesh with face, and an example of a generated area function.

other interesting challenge is the study of the singer's formant in sopranos and tenors where similar resonance clustering is expected to take place.

References

[1] D. Aalto, J. Helle, A. Huhtala, A. Kivelä, J. Malinen, J. Saunavaara, and T. Ronkka. Algorithmic surface extraction from MRI data: modelling the human vocal tract. In *BIODEVICES*, 2013.

[2] D. Aalto, J. Malinen, P. Palo, O. Aaltonen, M. Vainio, R.-P. Happonen, R. Parkkola, and J. Saunavaara. Recording speech sound and articulation in MRI. In *BIODEVICES*, 2011.

[3] D. Aalto, O. Aaltonen, R.-P. Happonen, P. Jääsaari, A. Kivelä, J. Kuortti, J.-M. Luukinen, J. Malinen, T. Murtola, R. Parkkola, J. Saunavaara, T. Soukka, and M. Vainio. Measurement of acoustic and anatomic changes in oral and maxillofacial surgery patients. 2013. arXiv:1309.2811.

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

[5] T. Lukkari and J. Malinen. Webster's equation with curvature and dissipation. arXiv 1204.4075, 2013 (submitted).

[6] J. Sundberg. Articulatory interpretation of the "singing formant". *J. Acoust. Soc. Am.*, 55(4):838–844, 1974.

[7] T. Vampola, J. Horáček, A.-M. Laukkanen, and J. Švec. Finite element modelling of vocal tract changes after voice therapy. *Applied and Computational Mechanics*, 5(1), 2011.

[8] T. Vampola, J. Horáček, A.-M. Laukkanen, and J. Švec. Human vocal tract resonances and the corresponding mode shapes investigated by three-dimensional finite-element modelling based on CT measurement. *Logopedics Phoniatrics Vocology*, pages 1–10, 2013.

[9] A. Webster. Acoustic impedance, and the theory of horns and of the phonograph. *Proc. Natl. Acad. Sci. USA*, 5:275–282, 1919.