Vowel formants from the wave equation

Antti Hannukainen, Teemu Lukkari, Jarmo Malinen and Pertti Palo

Institute of Mathematics, Helsinki University of Technology
P.O. Box 1100, FI-02015 TKK, Espoo, FINLAND
Antti.Hannukainen@tkk.fi, Teemu.Lukkari@tkk.fi, Jarmo.Malinen@tkk.fi, and Pertti.Palo@tkk.fi

Abstract: This article describes modal analysis of acoustic waves in the human vocal tract (VT) while the subject is pronouncing [øː]. The model used is the wave equation in three dimensions, together with physically relevant boundary conditions. The geometry is reconstructed from anatomical MRI data obtained by other researchers. The computations are carried out using the Finite Element Method. The model is validated by comparing the computed modes with measured data.

PACS numbers: 43.70.Bk, 43.20.Ks

1. Introduction

The purpose of this article is to study vowel production by the wave equation with boundary conditions as specified by Eq. (2) below. This model constitutes the input part of a *(scattering) conservative linear dynamical system* as presented by, e.g., Malinen *et al.* (2006); Malinen and Staffans (2006: 2007). A preliminary version of the present work was presented at the Phonetics Symposium 2006 (Hannukainen *et al.* 2006).

In the past, the vocal tract (VT) acoustics has been modelled in a number of different ways. Electrical transmission lines have been used for long time (see, e.g., Dunn 1950). The celebrated Kelly–Lochbaum model makes use of reflection coefficients obtained from a variable diameter tube (Kelly and Lochbaum 1962). Such reflection coefficients appear in, e.g., models from geophysics and in interpolation theory (see Foias and Frazho 1990). We remark that the Kelly–Lochbaum model is closely related to the horn model described by the Webster equation (see Chiba and Kajiyama 1958; Fant 1970). All these models have produced very accurate simulation results with a relatively light computational load, and they have applications, e.g., in mobile phones. More advanced two- and three-dimensional descendants of the Kelly–Lochbaum model are the transmission line networks that have been developed by El Masri *et al.* (1996: 1998); Mullen *et al.* (2006). For a recent review and further references, see Palo (2006).

Equation (2) in an anatomically realistic geometry has a more direct basis in physics than any of the approaches discussed in the previous paragraph. This is particularly useful in some applications, for example, in modelling the effects of anatomical abnormalities and maxillofacial surgery on speech (Dedouch *et al.* 2002a; Nishimoto *et al.* 2004; Švancara and Horáček 2006). As solving Eq. (2) analytically is possible only in a radically simplified geometry (see Sondhi 1986), we solve the problem numerically by *Finite Element Method* (FEM). This is the approach used by, e.g., Lu *et al.* (1993), Suzuki *et al.* (1993), Kawanishi *et al.* (1996), Niikawa *et al.* (2002), Dedouch *et al.* (2002b), Sasaki *et al.* (2003), and Švancara *et al.* (2004), too. Unfortunately, heavy computations are involved in this method.

We present a modal analysis of an anatomical configuration of [øː] as produced by a native Swedish speaker. We obtain resonance frequencies computationally, which correspond to formants. Unlike the scattering transfer function estimation used by Nishimoto *et al.* (2004) and Sasaki *et al.* (2003), our method does not necessarily require taking into account the radiation impedance at the mouth. Our approach is more closely related to Dedouch *et al.* (2002b) but instead of Neumann boundary condition on the glottis, we use a reflection-free boundary condition slightly above the glottis (see the last lines of Eq. (2) and Eq. (4)). Using reflection-free boundary conditions Eq. (3), our Eq. (2) can be coupled to a glottis model in a physically

realistic manner. Our results indicate that the computationally obtained formants identify the vowel [ø:] correctly in a larger set of measured data.

For numerical computations, a detailed geometric description of the VT is necessary. Nowadays, accurate anatomical data can be obtained using Magnetic Resonance Imaging (MRI). We are indebted to Dr. Olov Engwall (KTH) for kindly providing us with the required data.

2. Acoustic model

Deriving the wave equation for sound pressure starts by assuming that the total pressure $p = p(\mathbf{r}, t)$ and the density $\rho = \rho(\mathbf{r}, t)$ can be expressed as

$$p(\mathbf{r}, t) = p_0 + p'(\mathbf{r}, t) \quad \text{and} \quad \rho(\mathbf{r}, t) = \rho_0 + \rho'(\mathbf{r}, t), \tag{1}$$

respectively, where p_0 and ρ_0 are independent of time t and space variable $\mathbf{r}$. For linearisation of the equations, it is assumed that $p' = p'(\mathbf{r}, t) \ll p_0$ and $\rho' = \rho'(\mathbf{r}, t) \ll \rho_0$ are small perturbations at point $\mathbf{r} = (x, y, z) \in \Omega$ at time t. Here $\Omega \subset \mathbb{R}^3$ denotes the interior of the VT with boundary $\partial\Omega = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$, where Γ_1 is the mouth opening, Γ_2 denotes the walls of the VT, and Γ_3 is a virtual boundary control surface a small distance above the glottis.

By $\mathbf{v} = \mathbf{v}(\mathbf{r}, t)$ denote the velocity field of the flow described by p and ρ. A velocity potential $\Phi = \Phi(\mathbf{r}, t)$ is any function that satisfies $\mathbf{v} = -\nabla\Phi$. With this notation, our acoustic model is given by

$$\begin{cases} \Phi_{tt} = c^2 \Delta\Phi \text{ on } \Omega, \\ \Phi = 0 \text{ on } \Gamma_1, \quad \frac{\partial\Phi}{\partial\nu} = 0 \text{ on } \Gamma_2, \text{ and} \quad \Phi_t + c\frac{\partial\Phi}{\partial\nu} = 2\sqrt{\frac{c}{\rho_0}}u \text{ on } \Gamma_3, \end{cases} \tag{2}$$

where $u = u(\mathbf{r}, t)$ is the incoming power (per unit area) at glottis input, c is the sound velocity in the VT, ν is the exterior unit normal on $\partial\Omega$, and $\frac{\partial\Phi}{\partial\nu} = \nu \cdot \nabla\Phi$. The problem is to compute the velocity potential $\Phi(\mathbf{r}, t)$ for a given glottal input function $u(\mathbf{r}, t)$.

To derive Eq. (2) from "first principles", one needs to assume that an isentropic thermodynamic equation of state for pressure $p = p(s, \rho)$ holds where s, ρ are the entropy and density, respectively. Then we define the sound speed c by linearising the equation of state $p' = p(s, \rho_0 + \rho') - p(s, \rho_0) \approx c^2\rho'$ where $p_0 = p(s, \rho_0)$ and $c^2 = \frac{\partial p}{\partial \rho}(s, \rho_0)$. In this approximation, the entropy s is kept constant since the associated thermodynamic process is assumed to be reversible. In the case of monatomic ideal gas, we have $p/\rho^\gamma = p_0/\rho_0^\gamma$ and $c^2 = \gamma p_0/\rho_0$ where $\gamma = 5/3$ is the adiabatic constant.

Now the wave equation $\Phi_{tt} = c^2\Delta\Phi$ can be derived by a linearisation argument involving the continuity equation, Euler equation and linearised equation of state $p' = c^2\rho'$. Having computed Φ, we obtain the perturbation pressure from $p' = \rho_0\Phi_t$. All this can be found, e.g., in Fetter and Walecka (1980: Chapter 9).

Equation (2) is sophisticated enough to capture many relevant properties of wave propagation in three-dimensional geometry (e.g., to detect cross modes). It can also be used as the theoretical starting point in deriving the Webster equation mentioned above. However, it does not take into account turbulence, shock formation, or losses due to viscosity, heat conduction, or boundary dissipation.

We also need to take into account the walls and both ends of the VT. The last three lines in Eq. (2) specify the required boundary conditions. We regard the mouth as an open end of an acoustic tube, and this is described by the Dirichlet condition $\Phi(\mathbf{r}, t) = 0$. More complicated models for the mouth opening or the surrounding acoustic space have been considered by Kawanishi *et al.* (1996) (an impedance model involving Bessel functions), Nishimoto *et al.* (2004) (an impedance model consisting of a small reflecting hemisphere), and Švancara *et al.* (2004) (an exterior model of two concentric spheres with an absorbing outer boundary).

On the walls of the VT, we use the same Neumann condition $\frac{\partial \Phi}{\partial \nu}(\mathbf{r}, t) = 0$ as one would use at the closed end of a resonating tube. These two boundary conditions are discussed by Fetter and Walecka (1980: pp. 306-307).

At the glottis end, we use a scattering boundary condition that specifies the ingoing sound energy wave. For motivation, we define the ingoing wave $u(\mathbf{r}, t)$ and the outgoing wave $y(\mathbf{r}, t)$ for $\mathbf{r} \in \Gamma_3$ by

$$u = \sqrt{\frac{\rho_0}{4c}} \left(c \frac{\partial \Phi}{\partial \nu} + \Phi_t \right) \quad \text{and} \quad y = \sqrt{\frac{\rho_0}{4c}} \left(c \frac{\partial \Phi}{\partial \nu} - \Phi_t \right). \tag{3}$$

First of these equations coincides with the third boundary condition in (2). The net power absorbed by the interior domain Ω through the control/observation boundary at time t satisfies

$$\int_{\Gamma_3} |u(\mathbf{r}, t)|^2 \, d\omega(\mathbf{r}) - \int_{\Gamma_3} |y(\mathbf{r}, t)|^2 \, d\omega(\mathbf{r}) = \int_{\Gamma_3} (-\nu(\mathbf{r})) \cdot \mathbf{j_e}(\mathbf{r}, \mathbf{t}) \, d\omega(\mathbf{r})$$

where $\mathbf{j_e} = -\rho_0 \Phi_t \nabla \Phi = p' \mathbf{v}$ is the energy-flux vector as introduced in Fetter and Walecka (1980: pp. 307).

Instead of solving Eq. (2), we solve an easier — yet relevant — problem related to Eq. (2). More precisely, we determine the resonance frequencies corresponding to a particular vowel articulation position. By Malinen and Staffans (2006: Theorem 2.3), the resonances of Eq. (2) can be solved by finding the discrete, complex frequencies λ and the corresponding nonzero eigenfunctions $\Phi_\lambda(\mathbf{r})$ such that the equations

$$\begin{cases} \lambda^2 \Phi_\lambda = c^2 \Delta \Phi_\lambda \text{ on } \Omega, \\ \Phi_\lambda = 0 \text{ on } \Gamma_1, \quad \frac{\partial \Phi_\lambda}{\partial \nu} = 0 \text{ on } \Gamma_2, \text{ and} \quad \lambda \Phi_\lambda + c \frac{\partial \Phi_\lambda}{\partial \nu} = 0 \text{ on } \Gamma_3 \end{cases} \tag{4}$$

are satisfied. The time harmonic extension $\Phi(\mathbf{r}, t) = \Phi_\lambda(\mathbf{r}) e^{\lambda t}$ of Φ_λ satisfies clearly Eq. (2). Using the connection $p' = \rho_0 \Phi_t$, the corresponding perturbation pressure distribution is given by $p'(\mathbf{r}, t) = p_\lambda(\mathbf{r}) e^{\lambda t}$, where $p_\lambda(\mathbf{r}) := \rho_0 \lambda \Phi_\lambda(\mathbf{r})$. Thus Eq. (4) are satisfied with p_λ in place of Φ_λ.

3. Finite element modelling

The variational formulation of Eq. (4) (with p_λ in place of Φ_λ) is

$$\lambda^2 \int_\Omega p_\lambda \phi \, \mathrm{d}\Omega + \lambda c \int_{\Gamma_3} p_\lambda \phi \, \mathrm{d}\omega + c^2 \int_\Omega \nabla p_\lambda \cdot \nabla \phi \, \mathrm{d}\Omega = 0, \tag{5}$$

where ϕ is an arbitrary test function in Sobolev space $H^1_{\Gamma_1}(\Omega) = \{f \in H^1(\Omega) : f(\mathbf{r}) = 0 \text{ for } \mathbf{r} \in \Gamma_1\}$. The Finite Element Method (FEM) can be used to approximately solve Eq. (5); see, e.g., Johnson (1987) for an elementary treatment. We use piecewise linear shape functions and a tetrahedral mesh of $n = 64254$ elements which gives sufficiently accurate results. We obtain three $n \times n$ matrices, namely the stiffness matrix $\boldsymbol{K}$, the mass matrix $\boldsymbol{M}$, and $\boldsymbol{P}$ representing the glottis boundary condition in Eq. (4).

When treating Eq. (5) we proceed to solve the following linear algebra problem: find all complex numbers λ and corresponding nonzero vectors $\mathbf{x}(\lambda)$ such that

$$\lambda^2 \boldsymbol{K} \mathbf{x}(\lambda) + \lambda c \boldsymbol{P} \mathbf{x}(\lambda) + c^2 \boldsymbol{M} \mathbf{x}(\lambda) = 0 \quad \Leftrightarrow \quad \boldsymbol{A} \mathbf{y}(\lambda) = \lambda \boldsymbol{B} \mathbf{y}(\lambda). \tag{6}$$

where $\boldsymbol{A} = \begin{bmatrix} -c\boldsymbol{P} & -c^2\boldsymbol{M} \\ \boldsymbol{I} & 0 \end{bmatrix}$, $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{K} & 0 \\ 0 & \boldsymbol{I} \end{bmatrix}$, and $\mathbf{y}(\lambda) = \begin{bmatrix} \lambda \mathbf{x}(\lambda) \\ \mathbf{x}(\lambda) \end{bmatrix}$ (Saad 1992). The numbers λ are good approximations of the λ's appearing in Eq. (4), provided that the number n of elements is high enough. The lowest formants F1, F2, ..., correspond to the numbers λ in the order of increasing imaginary part.

4. Data

Figure 1 in in Hannukainen *et al.* (2006) shows a sliced representation of the VT geometry that we have used as the basis of our analysis. There are 29 slices, each consisting of 51 points, and they define the VT from glottis to mouth. For faster computation, the slices were down-sampled by taking into account only every fourth point.

The raw MRI data was collected from a native male speaker of Swedish while he pronounced a prolonged vowel [øː] in supine position. Engwall and Badin (1999) describe the MR imaging procedure and image post-processing. Corresponding formant measurement data is also available in the same article. The formants were estimated from speech recorded on a different occasion but with the same subject in a similar supine condition.

5. Results and conclusions

The latter form of Eq. (6) was solved in MATLAB environment, and the formants F1 to F4 that we obtained are shown in Table 1. These computed formants are roughly $3\frac{1}{2}$ semitones too high compared to the measured values, and we will discuss the physical background of this discrepancy below. The bottom row in Table 1 shows the computed formants multiplied by 0.817, which corresponds to a difference of $3\frac{1}{2}$ semitones.

TABLE 1. Computed, measured, and scaled formants for [øː] in kHz

	F1	F2	F3	F4
Computed	0.68	1.35	2.71	3.79
Measured	0.50	1.06	2.48	3.24
Scaled	0.56	1.11	2.22	3.10

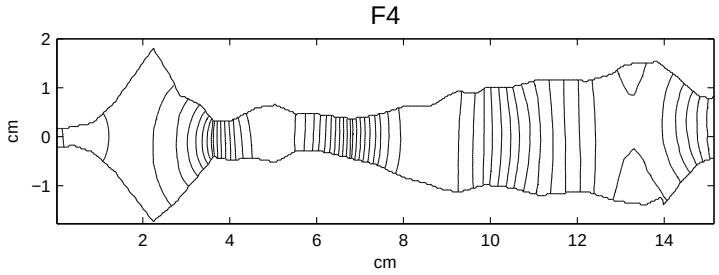

FIG. 1. Isobars corresponding to F4 along a mid-line cut. The mouth is on the right.

We also obtained the resonance modes p_λ — see Eq. (4) — corresponding to the formants F1-F4. They are computed as linear combinations of the element basis functions, using the components of $\mathbf{x}(\lambda)$ as weights. We note that the perturbation pressures p_λ are not given here in any physically relevant scale but they have been normalised so that the maximum deviation from the static pressure p_0 is either 1 or -1. Figure 1 shows isobars for the fourth mode. Figure 2 shows the pressure distributions of the modes. Figures 1 and 2 are plotted along a cross-sagittal mid-line cut (see Fig. 1, Hannukainen *et al.* 2006). We remark that Fig. 1 supports the hypothesis that a weak cross-mode resonance related to F4 should appear in the oral cavity.

The vowels from Engwall and Badin (1999: Table 4), together with the scaled and computed $[\text{øː}]_{s,c}$ from Table 1, are plotted in the (F2, F1)-plane in Fig. 3. Clearly, $[\text{øː}]_{s,c}$ is closre to measured [øː] than to any other measured vowel, *except* possibly [ɑː]. To further clarify the

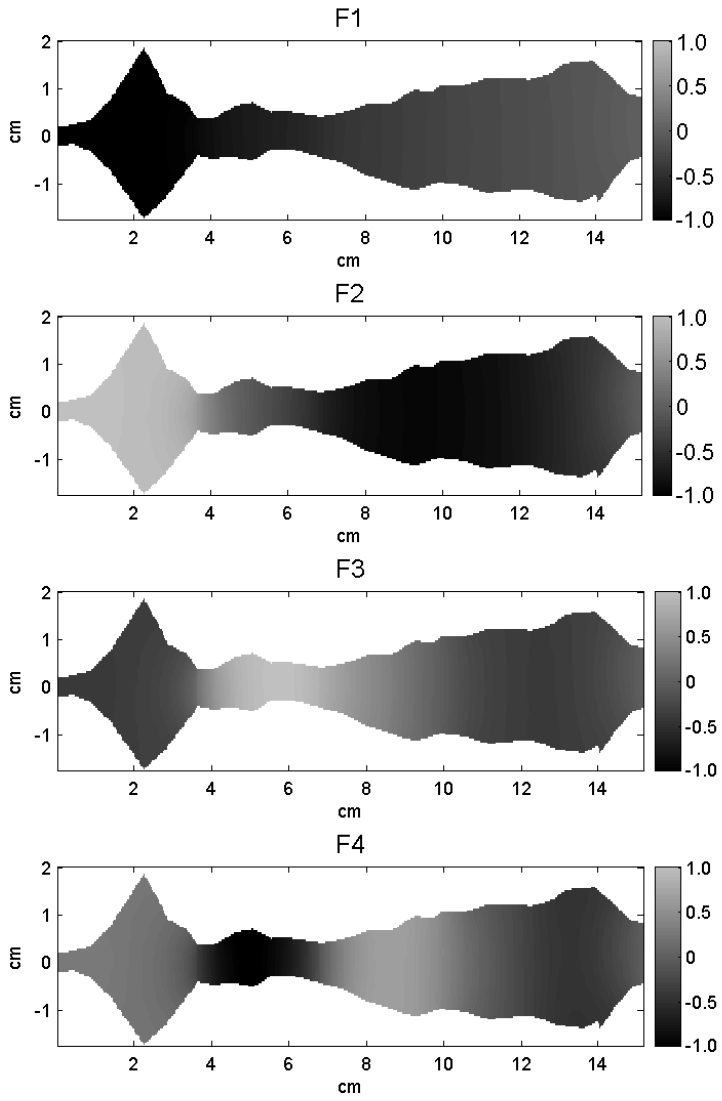

FIG. 2. Pressure distributions for F1-F4 along a mid-line cut. The mouth is on the right.

situation, let us consider the formants F1 to F4 for $[\text{ø:}]_{s,c}$, [ø:], and [ɑ:] as vectors: $[\text{ø:}]_{s,c} = (0.56, 1.11, 2.22, 3.10)$, $[\text{ø:}] = (0.5, 1.06, 2.48, 3.24)$, and $[\text{ɑ:}] = (0.56, 0.94, 2.74, 3.24)$. Then the Euclidean distance between $[\text{ø:}]_{s,c}$ and [ø:] is 0.31, but the distance between $[\text{ø:}]_{s,c}$ and [ɑ:] is significantly larger, equalling 0.57. This difference is explained by F3, since the fourth formants are almost the same. We conclude that the *first two* formants classify the scaled, computed vowel $[\text{ø:}]_{s,c}$ almost correctly. Moreover, if we look at *all four* available formants, even the remaining ambiguity disappears.

As we pointed out earlier, the computed formants F1 to F4 differ from the corresponding measured formants by $3\frac{1}{2}$ semitones. Having said that, the *ratios* between the computed formants and the measured formants match each other very well. There is a simple physical explanation why such a discrepancy is to be expected. In Eq. (2), we use the Dirichlet bound-

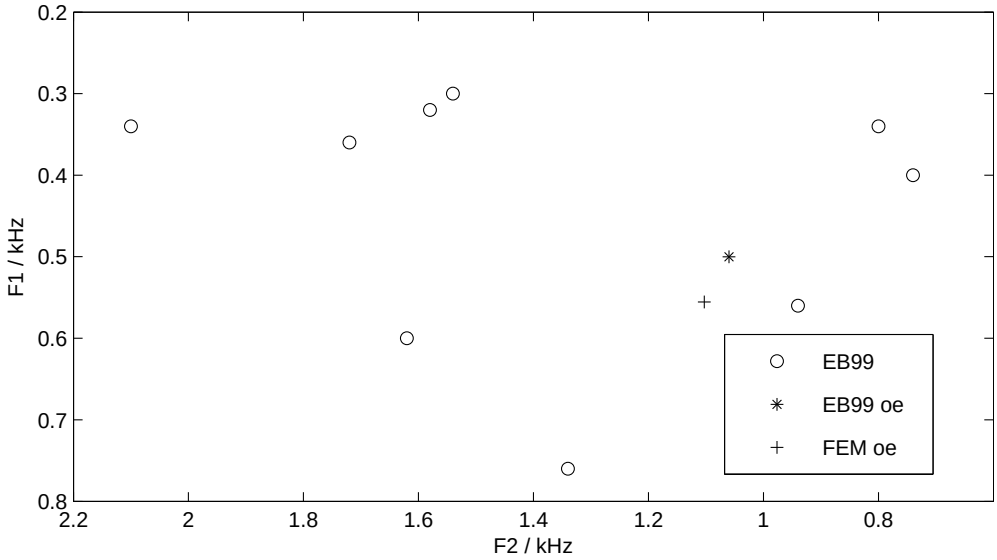

FIG. 3. Vowels in the (F2,F1)-plane. *FEM oe* (+) is the scaled, computed [øː], *EB99 oe* (*) is the measured [øː] and *EB99* (o) are other measured vowels. (EB99 denotes Engwall and Badin (1999).)

ary condition on the lip opening. This results in a vibrational node at the opening. In reality, such a node would appear further away outside the mouth since we are surely able to hear the sound outside of a speakers VT. In that sense, the real life VT is effectively longer than the one described by Eq. (2), resulting in lower formants. To get rid of this artefact, we should also model the surrounding acoustic space.

Surrounding acoustic space has been modelled by a lumped impedance for a transmission line (Laine 1982), by using a "small space" model with impedance termination on the outer shell (Nishimoto *et al.* 2004), and by using a "large space" model with an absorbing outer boundary (Švancara *et al.* 2004). The first two of these approaches include a tuning parameter to be determined experimentally so that the measured and computed formants coincide. We remark that impedance termination for the wave equation is inherently more difficult than for the transmission line, since the termination must be of boundary control instead just of point control type.

Acknowledgements

We would like to thank Olov Engwall from KTH, Stockholm, for providing the articulation geometry for this study. We would also like to thank an anonymous reviewer for valuable comments.

Antti Hannukainen has been supported by the Academy of Finland.

References

Chiba, T. and Kajiyama, M. (**1958**). The Vowel, Its Nature and Structure, Phonetic Society of Japan.

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

Dedouch, K., Horáček, J., Vampola, T., Švec, J., Kršek, P., and Havlík, R. (**2002**b). "Acoustic modal analysis of male vocal tract for Czech vowels," in "Proceedings Interaction and Feedbacks '2002," 13 – 19.

Dunn, H. K. (**1950**). "The calculation of vowel resonances, and an electrical vocal tract," J. Acoust. Soc. Am. **22**, 740 – 753.

El Masri, S., Pelorson, X., Saguet, P., and Badin, P. (**1996**). "Vocal tract acoustics using the transmission line matrix (TLM) method," in "Proceedings of the 4th International Conference on Spoken Language Processing," 953 – 956.

El Masri, S., Pelorson, X., Saguet, P., and Badin, P. (**1998**). "Development of the transmission line matrix method in acoustics. applications to higher modes in the vocal tract and other complex ducts," Int. J. of Numerical Modelling **11**, 133 – 151.

Engwall, O. and Badin, P. (**1999**). "Collecting and analysing two- and three-dimensional MRI data for swedish," TMH-QPSR , 11–38.

Fant, G. (**1970**). Acoustic Theory of Speech Production, Mouton, The Hague.

Fetter, A. and Walecka, J. (**1980**). Theoretical Mechanics of Particles and Continua, McGraw–Hill, New York.

Foias, C. and Frazho, A. E. (**1990**). The Commutant Lifting Approach to Interpolation Problems, vol. 44 of *Operator Theory: Advances and applications*, Birkhäuser Verlag, Basel.

Hannukainen, A., Lukkari, T., Malinen, J., and Palo, P. (**2006**). "Formants and vowel sounds by finite element method," in "The Phonetics Symposium 2006," Helsinki, Finland, 24 – 33.

Johnson, C. (**1987**). Numerical Solution of Partial Differential Equations by the Finite Element Method, Cambridge University Press.

Kawanishi, Y., Obuchi, M., Dang, J., Nakai, T., and Suzuki, H. (**1996**). "Consideration of the acoustic characteristics of the pyriform fossa in transmission line model, 3D-FEM model and realistic model," Tech. Rep. IEICE **EA96-12**, 1 – 8.

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

Laine, U. K. (**1982**). "Modelling of lip radiation impedance in z-domain," in "IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP -82)," vol. 3.

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.

Malinen, J. and Staffans, O. J. (**2006**). "Conservative boundary control systems," J. Diff. Eq. **231**, 290 – 312.

Malinen, J. and Staffans, O. J. (**2007**). "Impedance passive and conservative boundary control systems," Compl. Anal. Oper. Theory **2**.

Malinen, J., Staffans, O. J., and Weiss, G. (**2006**). "When is a linear system conservative?" Quart. Appl. Math. **64**, 31 – 91.

Mullen, J., Howard, D., and Murphy, D. (**2006**). "Waveguide physical modeling of vocal tract acoustics: Flexible formant bandwith control from increased model dimensionality," IEEE Transactions on Audio, Speech and Language Processing **14**, 964 – 971.

Niikawa, T., Ando, T., and Matsumura, M. (**2002**). "Frequency dependence of vocal-tract length," in "Proceedings of the 7th International Conference on Spoken Language Processing," 1525 – 1528.

Nishimoto, H., Akagi, M., Kitamura, T., and Suzuki, N. (**2004**). "Estimation of transfer function of vocal tract extracted from MRI data by FEM," in "The 18th International Congress on Acoustics," vol. II, Kyoto, Japan, 1473 –1476.

Palo, P. (**2006**). A Review of Articulatory Speech Synthesis, Master's thesis, TKK, Helsinki.

Saad, Y. (**1992**). Numerical Methods for Large Eigenvalue Problems, Manchester University Press, Manchester.

Sasaki, K., Miki, N., and Miyanaga, Y. (**2003**). "FEM analysis based on 3-D time-varying vocal tract shape," in "EUROSPEECH-2003," 2357 – 2360.

Sondhi, M. M. (**1986**). "Resonances of a bent vocal tract," J. Acoust. Soc. Am. **79**, 1113–1116.

Suzuki, H., Nakai, T., Takahashi, N., and Ishida, A. (**1993**). "Simulation of vocal tract with three-dimensional finite element method," Tech. Rep. IEICE **EA93-8**, 17 – 24.

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," .