ASSESSMENT OF GLOTTAL INVERSE FILTERING BY USING AEROELASTIC MODELLING OF PHONATION AND FE MODELLING OF VOCAL TRACT

P. Alku[1], J. Horáček [2], M. Airas[1], A-M. Laukkanen[3]
[1]Lab. of Acoustics and Audio Signal Processing, Helsinki University of Technology, Finland
[2]Institute of Thermomechanics, Academy of Sciences of the Czech Republic, Prague, Czech Republic
[3]Dept. of Speech Communication and Voice Research, University of Tampere, Finland

Performance of glottal inverse filtering (IF) is evaluated in this paper by using speech material produced with computational modelling of voice production represented by an aeroelastic model of vocal folds and a Finite Element (FE) model of the vocal tract. An inverse filtering algorithm was used in order to estimate the glottal flow from the speech pressure signal generated by the model. Comparison between the estimated glottal flow and the original flow generated by computational modelling shows that the IF method is able to yield an accurate estimate for the glottal flow.

I. INTRODUCTION

Inverse filtering (IF) is a non-invasive method to estimate the source of voiced speech, the glottal volume velocity waveform. In this technique, a model for the vocal tract transfer function is first computed. The effect of the vocal tract is then cancelled from the produced speech waveform by filtering this through the inverse of the model. As an input to IF, it is possible to use either the oral flow recorded in the mouth with a flow mask (e.g. [1]) or the pressure waveform captured by a microphone in free field outside the mouth (e.g. [2]).

Performance of an inverse filtering method is practically impossible to assess with natural speech. This comes from the fact that it is not possible to analyse how closely the estimated glottal flow given by an inverse filtering algorithm corresponds to the true glottal flow because the latter can not be measured. It is, however, possible to assess inverse filtering by using synthetic speech that has been created using a known, artificial waveform of the glottal excitation. This kind of evaluation, however, is not truly objective, because speech synthesis and inverse filtering analysis are typically based on similar models of the human voice production apparatus (e.g. the source filter model [3]).

In the current study, we combine *physical modelling* of voice production in order to synthesize speech with a known, realistic glottal flow waveform. By using the pressure signals given by the physical models as an input to an inverse filtering method, it is then possible to analyze how closely the obtained estimate of the voice source matches the original glottal flow.

The paper first describes in section II the methodology used both in physical modelling (sections IIA and IIB) and in inverse filtering (section IIC). The results obtained for a sustained male vowel are described in section III and the paper is finished with short conclusions in section IV.

II. METHODOLOGY

A. Aeroelastic model of the vocal folds

Recently an aeroelastic model was developed by Horáček et al. [4, 5] that allows numerical simulation of self-oscillations of the vocal folds. The incompressible 1-D fluid flow theory is used in the model for expressing the unsteady aerodynamic forces and the Hertz model is used for the impact forces. The parameters of the model, i.e., the mass, stiffness and damping matrices are approximately related to the geometry, size and material density of real vocal folds as well as to a prescribed fundamental frequency (F0) and damping. In this contribution, the output of the numerical simulation, i.e., the intraglottal airflow rate is used to excite an FE model of human vocal tract representing the vowel /a/.

Symmetric oscillations are assumed and hence the vibration of only one vocal fold is modelled. Vocal fold oscillations are simulated by a vibrating element of length L with mass m and moment of inertia I with two-degrees-of-freedom supported by an elastic foundation in the wall of a channel conveying air (Fig. 1). The motion of an equivalent three mass system on two springs can be described by the following equation:

$$\overline{\mathbf{M}}\,\ddot{\mathbf{V}} + \overline{\mathbf{B}}\,\dot{\mathbf{V}} + \overline{\mathbf{K}}\,\mathbf{V} + \mathbf{F} = \mathbf{0}, \qquad (1)$$

where $\overline{\mathbf{M}}, \overline{\mathbf{B}}, \overline{\mathbf{K}}$ are the structural mass, damping and stiffness matrices, respectively, and ${}^{\mathrm{T}}\mathbf{V}=[V_1(t),V_2(t)]$ is the

Models and analysis of vocal emissions for biomedical applications. 4th international workshop. October 29-31, 2005 – Firenze, Italy.
Edited by C. Manfredi. ISBN 88-8453-320-1 (online)

vector for rotation and translation of the vibrating element. The vector for nonlinear aerodynamic and collision forces can be expressed as

$$^{T}\mathbf{F} = [F_1(t), F_2(t)],$$
$$F_{1,2} = \rho \sum_{i,j=0}^{2} \sum_{k,l=0}^{2} {}^{1,2}K_{i,j,k,l} \left[V_1^{(i)}(t)\right]^k \left[V_2^{(j)}(t)\right]^l \quad (2)$$

where the superscripts of V_1 and V_2 denote the order of time derivatives and $K_{i,j,k,l}$ are constant coefficients. For the numerical simulations Eq. 1 was transformed into a system of four 1st order ordinary differential equations and 4th order Runge-Kutta method was used for the calculations.

The following parabolic function is used to approximate the geometry of the vocal folds:

$$a(x) = 1.858 - 159.86\, x^2 \quad (3)$$

The airflow velocity U_0 at the inlet (x=0) to the glottal region is simply related to the mean glottal volume velocity according to $Q = U_0 2H_0 h$ and to the static subglottal pressure according to:

$$P_{\text{sub}} = 1/2\, \rho\, U_0^{\,2} \left\{ H_0 / \left[H_0 - a(L) \right] \right\}^2 \quad (4)$$

During the vocal folds collision, the static subglottal pressure is constant that equals the pressure in the lungs (P_{lungs}). H_0 and h denote the height and width of the channel, respectively. Using tissue density ρ_h=1020kg/m^3, thickness L=6.8mm and length of the vocal fold h=10mm, eccentricity (e), total mass (m) and moment of inertia (I) were calculated. As the value of air density we used ρ=1.2kg/m^3. A tuning procedure was used to adjust the stiffness of the elastic foundation of the vibrating element and the damping coefficients in order to approximate the fundamental frequency F0 by setting the natural frequencies f_1=$F0$, f_2=$F0$+5Hz and 3dB half-power bandwidths $\Delta f_{1,2}$ of both resonances. The optimum distance between the two supporting springs was adjusted to l=0.344L, for which the real values of the stiffness coefficients c_1, c_2 can be calculated for the prescribed frequencies f_1, f_2. In the example studied in this paper, the following values were used for the input data: prephonatory glottal half-width g=0.2mm, F0≅100Hz, Δf_1=23Hz, Δf_2=29Hz, U_0=1.6m/s, Q=0.18l/s, P_{lungs}=380Pa and the Hertz coefficient for the vocal folds collisions k_H=730Nm$^{-2/3}$. The following main output data resulted from the simulation: open quotient OQ=0.72, skewing (speed) quotient QS=1.56, closing quotient CQ=0.28, fundamental frequency F0=1/T=100.77Hz calculated from the period T of the self-oscillations, maximum glottis opening GO=1.27mm, maximum impact stress IS=1328Pa and supraglottal pressure SPL=124dB.

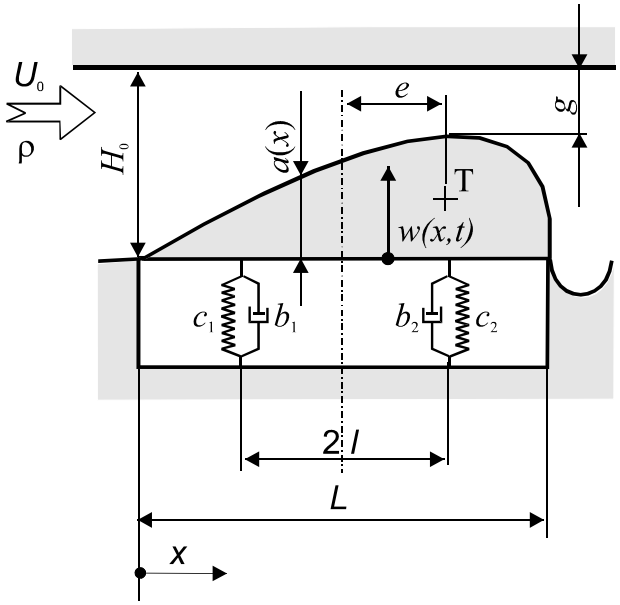

Figure 1. Two-degrees of freedom model of the vocal fold.

B. FE model of the vocal tract

FE modelling was used in numerical simulation of the vocal tract filtering by using the Czech vowel /a/ produced by a male speaker. The model was designed based on MRI data described in [6]. The vocal tract geometry was obtained from a native Czech speaker during phonation. The MRI of the vocal tract for the mid-sagittal cross-section and the designed FE model are shown in Fig. 2. The vocal tract was modelled by the ANSYS FE code using acoustics finite elements FLUID 30 with speed of sound c_0=343m/s and ρ=1.2kg/m^3.

The acoustic pressure p is described by the equation:

$$\nabla^2 p = \frac{1}{c_0^2} \frac{\partial^2 p}{\partial t^2} \quad (5)$$

and in FE formulation it can be written in the matrix form in the global co-ordinate system as

$$\mathbf{M}\, \ddot{\mathbf{P}} + \mathbf{B}\, \dot{\mathbf{P}} + \mathbf{K}\, \mathbf{P} = \mathbf{f}(t) \quad (6)$$

where **M**, **B**, **K** are the mass, acoustic boundary damping and stiffness matrices, respectively; **P** and **f** are the vectors of nodal acoustic pressures and excitation forces, respectively. The transient analysis with the Newmark integration method was used for numerical simulation of

the acoustic signal near the lips whereas the excitation was applied at the position of the vocal folds. The effect of outgoing acoustic energy was modelled by an absorption boundary condition at the lips, where a boundary admittance was prescribed in correspondence to the 3dB half-power bandwith known for formant (acoustic resonant) frequencies. The excitation signal was the intraglottal airflow volume velocity $Q(t)$ resulting from the aeroelastic model of the vocal folds.

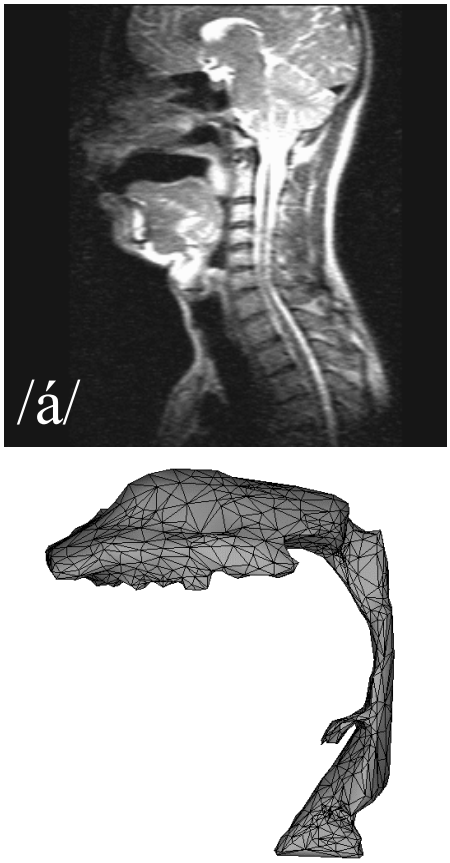

Figure 2. MRI of the male subject during phonation (upper) and FE model of the vocal tract for the Czech vowel /a/ (lower).

C. Inverse filtering

The inverse filtering method used is based on our previous experiments in developing automatic methods to estimate the glottal flow from the speech pressure waveforms with the Iterative Adaptive Inverse Filtering (IAIF) method [7]. The current method, the flow diagram of which is shown in Fig. 3, is a slightly modified version from our previous ones. Parametric spectral models that are used in various blocks of the flow diagram are computed with the Discrete All-pole Modeling (DAP) method [8] instead of the conventional linear predictive analysis. This makes it possible to obtain estimates of the formant frequencies that are less biased by the harmonic structure of the speech spectrum. The detailed description of the IAIF-method can be found in [9].

The IAIF method has limitations. It is based on straightforward linear modelling of speech production without taking into account, for example, the interaction between the glottal source and the vocal tract. Moreover, the digital model of the vocal tract is a pure all-pole filter, which is not accurate for nasals. Despite these inherent limitations, the proposed technique provides a promising method to estimate the glottal flow especially given the fact that the method can be implemented (if desired) in a completely automatic manner with a reasonable computational cost.

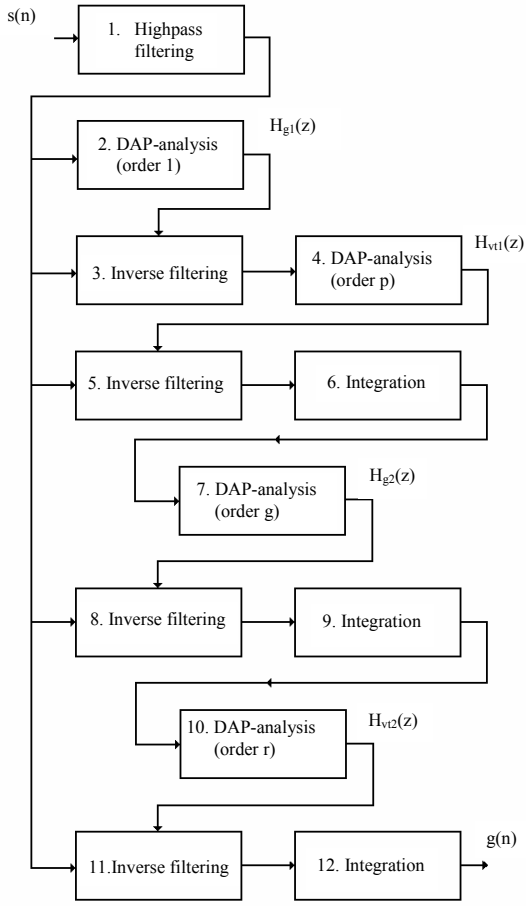

Figure 3. Block diagram of the IAIF method

III. RESULTS

The vowel sound produced by the physical modelling was inverse filtered with the IAIF method by using the following parameters (see Fig. 3): p = r = 12, g = 4. The sampling frequency was 10 kHz. The length of the analysis window was 50 ms. The lip radiation effect

(blocks no 6, 9 and 12 in Fig 3) was cancelled by a first order all-pole filter with its pole at z = 0.96.

The glottal flow estimate computed by the IAIF method is shown together with the original flow generated by physical modelling in Fig. 4. Both of the two time-domain waveforms were parameterised using the Normalized Amplitude Quotient (NAQ) [10]. The value of the NAQ parameter equalled 0.2085 and 0.2038 for the original and estimated flow, respectively. Hence, in terms of the NAQ parameter, the difference between the estimated glottal flow and the original one was approximately 2 %.

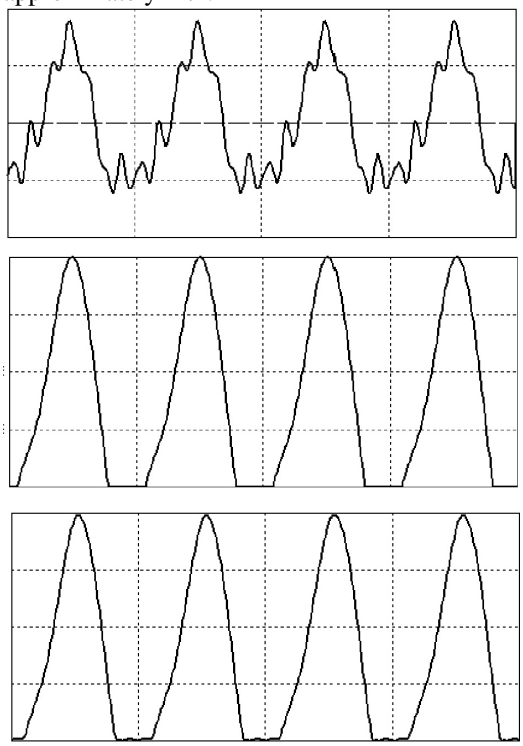

Figure 4. Speech pressure signal (top) and glottal flow (middle) generated by physical modelling. Estimated glottal flow (bottom) given by inverse filtering. All signals are in the time-domain, length of panel 40 ms.

IV. CONCLUSIONS

Evaluation of inverse filtering methods is problematic because direct measurements of the glottal flow are difficult, if not impossible. In addition, using synthetic speech as test material does not make a fully objective evaluation possible, because voice synthesis and inverse filtering are typically based on the same voice production models.

The present study aimed to avoid these fundamental limitations by using a vowel produced with physical modelling in evaluation of inverse filtering. The results were encouraging in showing that the difference between the original flow generated by physical modelling and estimated one was small.

The experiments of the present study were based on single vowel sound. In order to better understand the limitations of inverse filtering, the characteristics of the test material should be expanded. In particular, the range of F0 values used in the evaluation should be expanded to cover the pitch range of female speech.

ACKNOWLEDGEMENTS

This study was supported by the Academy of Finland (projects 200859 and 205962) and by the Grant Agency of the Academy of Sciences of the Czech Republic, project No IAA20766401 *Mathematical modelling of human vocal folds oscillations*.

REFERENCES

[1] M. Rothenberg, "A new inverse-filtering technique for deriving the glottal airflow waveform during voicing," *J. Acoust. Soc. Amer.*, vol. 53, pp. 1632-1645, 1973.

[2] D.Y Wong, J.D. Markel, and A.H. Gray, Jr., "Least squares glottal inverse filtering from the acoustic speech waveform," *IEEE Trans. on Acoust., Speech, and Signal Proc.*, vol. 27, pp. 350-355, 1979.

[3] G. Fant, *The Acoustics Theory of Speech Production*, the Hague: Mouton, 1960.

[4] J. Horáček, P. Šidlof, and J.G. Švec, "Numerical modelling of leakage-flow-induced vibrations of human vocal folds with Hertz impact forces," In: 3rd International Workshop MAVEBA 2003, pp. 143-146.

[5] J. Horáček, P. Šidlof, and J.G. Švec, "Numerical simulation of self-oscillations of human vocal folds with Hertz model of impact forces," In: Langre E, Axisa F, eds. Flow-Induced Vibration. Ecole Polytechnique, Paris, pp. 143-148.

[6] K. Dedouch, J. Horáček, J.G. Švec, P. Kršek, R. Havlík, and J. Vokřál, "Acoustic analysis of a male vocal tract for Czech vowels," In: Proc. Phoniatric Days of Eva Sedláčková, 11-13 Sept. 2003, Brno, pp 60-63.

[7] P. Alku, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," *Speech Comm.*, vol. 11, pp. 109-118, 1992.

[8] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling", *IEEE Trans. Signal Proc.*, vol. 39, pp. 411-423, 1991.

[9] P. Alku, B. Story, and M. Airas, "Evaluation of an inverse filtering technique using physical modeling of voice production," in CD Proc. of Int. Conf. on Spoken Lang. Proc. 2004.

[10] P. Alku, T. Bäckström, and E. Vilkman, "Normalized amplitude quotient for parameterization of the glottal flow," *J. Acoust. Soc. Amer.,* vol. 112, pp. 701-710, 2002.