Proceedings of the 24th Nordic Seminar on Computational Mechanics
J. Freund and R. Kouhia (Eds.)

Interaction of vocal fold and vocal tract oscillations

Atte Aalto[1], Daniel Aalto[2], Jarmo Malinen[1], and Martti Vainio[3]

(1)Dept. of Mathematics and Systems Analysis, Aalto SCI, aaalto@cc.hut.fi, jmalinen@cc.hut.fi
(2)Institute of Mathematics, Åbo Akademi, daniel@iki.fi
(3)Institute of Behavioural Sciences (Phonetics), University of Helsinki, martti.vainio@helsinki.fi

Summary. We study the mechanical feedback coupling between the human vocal folds and vocal tract (VT) by simulating fundamental frequency glides over the lowest VT resonance. In the classical source–filter theory of speech production, the vocal folds produce a signal which is filtered by the resonator, vocal tract without any feedback. We have developed a computational model of the vocal folds and the VT that also includes a counter pressure from the VT to the vocal folds. This coupling gives rise to new computational observations (such as modal locking) that can be established experimentally.

Key words: Speech modelling, vocal folds model, flow induced vibrations, modal locking

Introduction

According to the classical source–filter theory of vowel production, the source (i.e., the *glottis*, meaning the aperture between vocal folds) operates independently of the filter (i.e., the vocal tract (VT)) whose resonances modulate the harmonic contents of the resulting sound, see [1, 2]. It is well known that this theory is adequate for modelling a wide range of phenomena in speech production. However, when the vocal folds' oscillation frequency f_0 (a.k.a. the *fundamental frequency*) is near the lowest VT resonance F_1 (i.e., the first *formant frequency*), the vocal folds' oscillations are affected by the acoustics of the VT as is observed in the computational and experimental works [3, 4, 5]. Such phenomena appear, e.g., in soprano singers phonation.

We use the computational model developed in [6, 7] to simulate f_0-glides on a steady vowel. The model includes an additional aerodynamic load from the resonator (filter) that is fed back to the vocal fold (source) equations of motion. In simulations we observe a strong and consistent modal locking between vocal folds oscillations and the acoustic vibrations in the VT. This phenomenon can also be detected in the preliminary experimental materials that we briefly introduce in the work.

Computational model

The model consists of three subsystems: vocal folds, glottal flow, and vocal tract. Since f_0–F_1 crossover typically occurs in females whereas the original model parameters corresponds to male physiology (see [7]), we have scaled the vocal fold masses by a factor of 0.253 and the stiffnesses by 0.836, see [8]. For details, model parameters, and numerical realisation, we refer to [6, 7].

Vocal folds

The vocal fold model in Fig. 1 consists of two wedge-shaped elements with two degrees of freedom. The distributed mass of these elements is reduced into three mass points and the

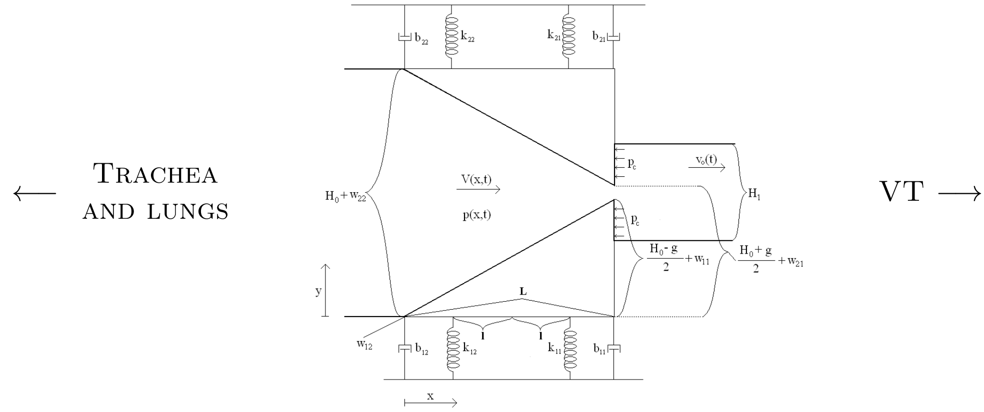

Figure 1. The geometry of the glottis model and the symbols used.

elastic support is approximated by two springs. The equations of motion for the vocal folds are

$$\begin{cases} \mathbf{M}_1\ddot{\mathbf{W}}_1(t) + \mathbf{B}_1\dot{\mathbf{W}}_1(t) + P\mathbf{K}_1\mathbf{W}_1(t) = -\mathbf{F}(t), \\ \mathbf{M}_2\ddot{\mathbf{W}}_2(t) + \mathbf{B}_2\dot{\mathbf{W}}_2(t) + P\mathbf{K}_2\mathbf{W}_2(t) = \mathbf{F}(t) \end{cases} \tag{1}$$

where $\mathbf{W}_j = (w_{j1}, w_{j2})^T$ are the displacements of the right and left endpoints of the j^{th} fold, $j = 1, 2$. The glottal opening at the narrowest point is denoted by ΔW_1. At the other end (towards the trachea), the opening is ΔW_2. These are given by (1) through $\left[\begin{smallmatrix}\Delta W_1 \\ \Delta W_2\end{smallmatrix}\right] = \mathbf{W}_2 - \mathbf{W}_1 + \left[\begin{smallmatrix} g \\ H_0\end{smallmatrix}\right]$. The parameter g is the glottal opening in neutral position. The mass, damping, and stiffness matrices are denoted by $\mathbf{M}_j$, $\mathbf{B}_j$, and $\mathbf{K}_j$. Control parameter P is used for simulating f_0-glides.

During the glottal open phase (when $\Delta W_1(t) > 0$), the load terms of (1) are given by $\mathbf{F} = (F_{A,1}, F_{A,2})^T$ that are given below in Eq. (3). During the glottal closed phase (when $\Delta W_1(t) < 0$), there are no aerodynamic forces apart from the acoustic counter pressure from the VT. Instead, there is a nonlinear spring force for the elastic collision of the vocal folds, given by the Hertz impact model (see, e.g., [9]), $\mathbf{F} = \begin{bmatrix} k_H|\Delta W_1|^{3/2} - \frac{H_0 - H_1/2}{2L}\frac{H_1}{2} h \cdot p_c \\ \frac{H_0 - H_1/2}{2L}\frac{H_1}{2} h \cdot p_c \end{bmatrix}$.

Glottal flow

An incompressible 1D flow through the glottal opening with velocity v_o is described by

$$\dot{v}_o(t) = \frac{1}{C_{iner} h H_1}\left(p_{sub} - \frac{C_g}{\Delta W_1(t)^3} v_o(t)\right) \tag{2}$$

motivated by the Hagen–Poiseuille law. The parameter p_{sub} is the subglottal pressure and h is the width of the rectangular flow channel. The parameter C_{iner} regulates the flow inertia, and C_g regulates the viscous pressure loss in the glottis.

In the glottis, the flow velocity $V(x, t)$ is assumed to satisfy the mass conservation law $H(x, t)V(x, t) = H_1 v_o(t)$ for static incompressible flow where $H(x, t) = \Delta W_2(t) + \frac{x}{L}(\Delta W_1(t) - \Delta W_2(t))$, $x \in [0, L]$ is the height of the flow channel inside the glottis. The pressure $p(x, t)$ in the glottis is given by the Bernoulli law $p(x, t) + \frac{1}{2}\rho V(x, t)^2 = p_{sub}$ for static flow.

This pressure and the VT counter pressure p_c are reduced to a force pair $(F_{A,1}, F_{A,2})^T$ where $F_{A,1}$ affects at the narrow end of the glottis ($x = L$) and $F_{A,2}$ at the wide end ($x = 0$). This reduction is carried out by using the total force and moment balance equations $F_{A,1} + F_{A,2} = h\int_0^L (p(x, t) - p_{sub})\, dx$ and $L \cdot F_{A,1} = h\int_0^L x(p(x, t) - p_{sub})\, dx - p_c \cdot h\frac{H_1}{2}\frac{H_0 - H_1/2}{2}$ that yield

$$\begin{cases} F_{A,1} &= \frac{1}{2}\rho v_o^2 hL\left(-\frac{H_1^2}{\Delta W_1(\Delta W_2 - \Delta W_1)} + \frac{H_1^2}{(\Delta W_1 - \Delta W_2)^2}\ln\left(\frac{\Delta W_2}{\Delta W_1}\right)\right) - \frac{H_1(H_0 - H_1/2)}{4L} h p_c, \\ F_{A,2} &= \frac{1}{2}\rho v_o^2 hL\left(\frac{H_1^2}{\Delta W_2(\Delta W_2 - \Delta W_1)} - \frac{H_1^2}{(\Delta W_1 - \Delta W_2)^2}\ln\left(\frac{\Delta W_2}{\Delta W_1}\right)\right) + \frac{H_1(H_0 - H_1/2)}{4L} h p_c. \end{cases} \tag{3}$$

Vocal tract

The VT is modelled by Webster's lossless horn equation $\Psi_{tt}(s,t) - \frac{c^2}{A(s)}\frac{\partial}{\partial s}\left(A(s)\frac{\partial \Psi(s,t)}{\partial s}\right) = 0$ where $\Psi(s,t)$ is a velocity potential, c is the sound velocity, $s \in [0, L_{VT}]$ is the distance from the glottis measured along the VT centreline, and L_{VT} is the length of the VT. The area function $A(\cdot)$ is the cross-sectional area of the VT. The sound pressure is given by $p = \rho\Psi_t$.

The resonator is controlled by the glottal flow velocity from (2) through the boundary condition $\Psi_s(0,t) = -v_o(t)$. The boundary condition at lips is $\Psi_t(L_{VT},t) + \theta c \Psi_s(L_{VT},t) = 0$ representing a frequency-independent acoustic resistance. The resonator exerts a counter pressure $p_c(t) = \rho\Psi_t(0,t)$ to (1) through (3), forming a feedback loop between the vocal folds and the VT.

Simulation and experimental results

We denote by $\tilde{f}_0$ the nominal frequency of the vocal fold oscillations if the feedback from the VT was removed. The actual, observed vocal fold oscillation frequency is denoted by f_0.

Frequency glides are simulated by quadratically increasing the parameter P in (1) so that $\tilde{f}_0$ increases linearly from 350 Hz to 810 Hz during a 2 s time interval. The spectrogram of the pressure at lips is shown in Fig. 2a with auxiliary lines showing the glide of $\tilde{f}_0$ and $F_1 = 647$ Hz. It is observed that f_0 coincides first with $\tilde{f}_0$, but then it suddenly jumps upwards to F_1 when it reaches about 470 Hz. The wave form of the glottal pulse near the transition is a superposition of two signals with frequencies $\tilde{f}_0$ and F_1. When $\tilde{f}_0$ exceeds F_1, then f_0 and $\tilde{f}_0$ coincide again. In the downward glide, a similar behaviour occurs as presented in Fig. 2b.

In the experiments, female subjects were asked to follow a target glide: a triangle wave sweep whose frequency grows from 170 Hz to 340 Hz linearly in logarithmic scale. In Fig. 3, a subject produces a rising glide of vowel /i/ for which 250 Hz $< F_1 <$ 300 Hz. Here the formant value has been measured from non-periodic phonation but it could also be obtained from the anatomy-based resonance analysis using MRI and FEM. Subjects use auditory, tactile, and proprioceptive feedback in controlling their phonation, and this is not modelled at all. A subject can start the corrective pitch regulation (e.g., near the modal locking) as early as 80 ms after a deviation from the given auditive target glide is perceived. Such corrective control actions may cause chaotic patterns such as the subphonation episode in Fig. 3.

Conclusions

We have introduced a simple model for simulating human vowel production. This model was used for studying the feedback effect from the vocal tract to the vocal folds in time domain complementing [4]. Qualitatively, the computational results are well in line with experiments: (1) in normal speech the vocal folds are not affected by the acoustics of the VT but (2) when

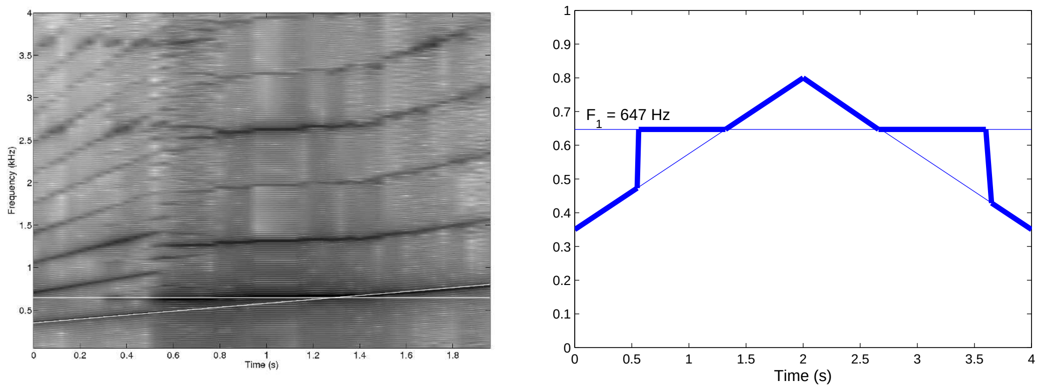

Figure 2. (a): Simulated f_0-glide upwards 350–810 Hz. (b): A sketch of the modal locking in an f_0-glide over F_1 first upwards and then downwards. The thick (thin) line shows f_0 (resp. $\tilde{f}_0$) during the glide.

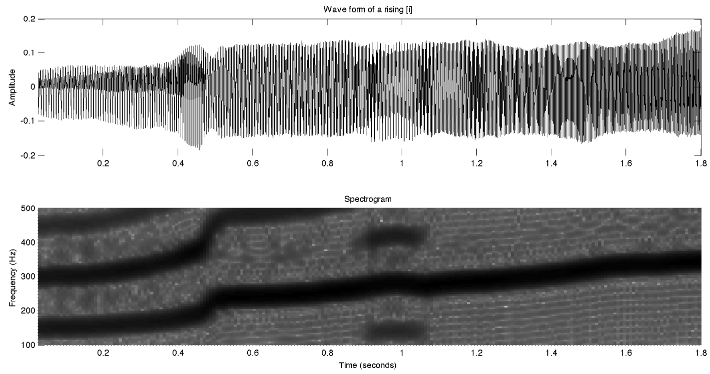

Figure 3. The wave form and the spectrogram (showing f_0) of a rising one octave glide with static vowel [i]. At 0.45–0.5 s a dip in amplitude coincides with a fast rise in f_0. At 0.9–1.06 s, subphonation occurs.

$|f_0 - F_1| < 100$ Hz, three kinds of instabilities are reported in [3]: fundamental frequency jumps, subharmonics, and chaotic behaviour. In simulations we have observed frequency jumps (of magnitude $\approx$170 Hz) at expected frequencies, and we propose that they are related to modal locking between the vocal fold and VT oscillations.

The experimental part of the work is in progress but some preliminary results are already available. It should be noted that the experimental results are complex: not all detected frequency jumps are due to the proposed coupling mechanism. Such jumps might occur because of register shifts, i.e., abrupt changes of the vibrating length of vocal folds. Subglottal formants may play a role at higher frequencies, and the muscle control may not be symmetric in falling and rising glides. Moreover, professional singers use at least following compensation mechanisms to avoid the coupling: change in F_1 by moving the tongue and pharynx configuration yet maintaining the vowel identifiable; reduction in the subglottal pressure inducing weaker glottal flow and hence weaker forces and weaker coupling; and a change to a more breathy phonation type.

References

[1] Chiba, T. and Kajiyama, M. *The Vowel, its Nature and Structure*. Tokyo-Kaiseikan Publishing Company Ltd., 1941.

[2] Fant, G. *The Acoustic Theory of Speech Production*. Moulton, The Hague, 1960.

[3] Titze, I., Riede, T., and Popolo, P. Nonlinear source-filter coupling in phonation: Vocal exercises, *Journal of the Acoustical Society of America*, 123(4):1902–1915, 2008.

[4] Titze, I. Nonlinear source-filter coupling in phonation: Theory, *Journal of the Acoustical Society of America*, 123(5):2733–2749, 2008.

[5] Hatzikirou H., Fitch W. T., and Herzel H. Voice Instabilities due to Source-Tract Interactions, *Acta Acustica united with Acustica*, 92:468–475, 2006.

[6] Aalto, A. *A low-order glottis model with nonturbulent flow and mechanically coupled acoustic load*, Master's thesis, TKK, Helsinki, 2009. `http://math.tkk.fi/research/sysnum/`

[7] Aalto, A., Alku, P., and Malinen, J. A LF-pulse from a Simple Glottal Flow Model, *MAVEBA2009 Proceedings*, Florence, Italy, 2009.

[8] Titze, I. Physiologic and acoustic differences between male and female voices, *Journal of the Acoustical Society of America*, 85(4):1699–1707, 1989.

[9] Horáček, J., Šidlof, P., and Švec, J. Numerical simulation of self-oscillations of human vocal folds with Hertz model of impact forces, *Journal of Fluids and Structures* 20:853–869, 2005.