Improved fold closure in mass-spring low-dimensional glottal models

C. Drioli[1] and F. Avanzini[2]
[2] Dept. of Computer Science, University of Verona, Italy
[1] Dept. of Information Engineering, University of Padova, Italy

Abstract: **This work presents a low-dimensional physical model of the glottis in which a 2-D fold displacement representation allows to represent both the vertical and longitudinal displacements of the folds. We use a one-mass mechanical model, coupled to aerodynamic driving forces, and we use a delay line representation to account for the propagation of the displacement on the body-cover. The waveform is characterized by means of a set of acoustic parameters (open quotient, speed quotient, return quotient, fundamental frequency F_0, etc.) that are used in the literature as typical voice source quantification parameters. The paper provides comparisons between values of these parameters computed for the proposed model and for analytical models (LF) of the flow.**

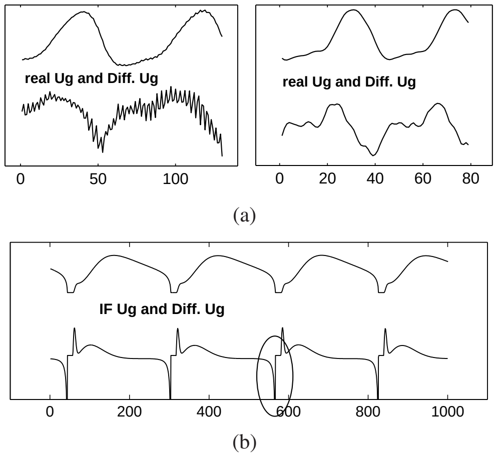
Keywords: **Voice source, Low-dimensional models, Voice source parameters, Voice quality**

I. INTRODUCTION

Low-complexity physical models based on the one- and two-mass paradigm have demonstrated to possess desirable properties: they are computationally efficient and stable, they offer physically justified control for basic glottal flow cues, and they can reproduce modal and non-modal phonation modalities for generating a wide range of phonatory styles and voice qualities [1], [2], [3], [4].

An open issue concerning simplified physical models of the vocal apparatus is that not always they allow to reproduce all the possible configurations and patterns of oscillation which can be observed in actual glottal flow waveforms. In particular, mass-spring models such as the classic Ishizaka-Flanagan (IF) model [5] are often characterized by unrealistic behavior in the closing phase due to very crude folds collision representations, and the abrupt closure often negatively affects the perceptual result of the synthesis. Smooth closing patterns are usually observable in inverse-filtered glottal waveforms, see Fig. 1, and are considered in many non-physical models (e.g., the well known Liljencrant-Fant (LF) analytical representation [6]).

This work focuses on a low-dimensional physical model of the glottis. We use a one-mass mechanical model, coupled to aerodynamic driving forces. We introduce a 2-D fold displacement representation, in order to be able to represent both the vertical and the longitudinal displacements of the fold through delay lines taking into account the propagation of the displacement on the body-cover.



(a)

(b)

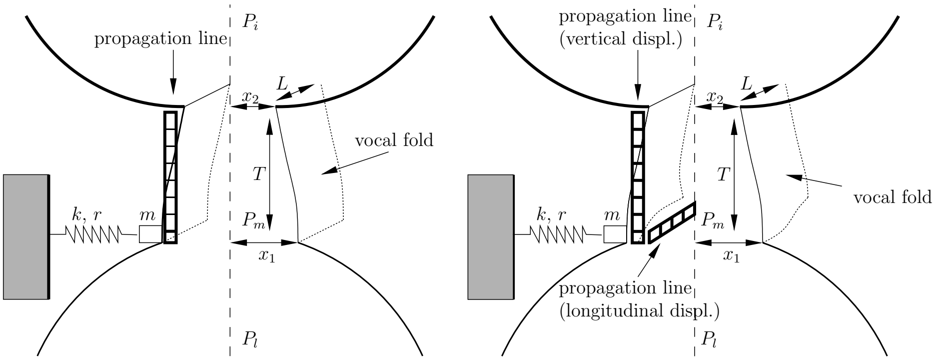
Fig. 1. Panel a): real glottal flow waveforms obtained by inverse filtering; panel b): glottal flow synthesis from an implementation of the IF model. The typical abrupt closure shape is highlighted.

The paper is organized as follows. Section II gives an overview of the voice production model under investigation and presents the details of the refinements proposed. In Section III, some experimental results are presented and some properties of the new model are discussed by comparing it with the LF analytical model. In Section IV the conclusions are given.

II. METHOD

The glottis model adopted here is a low-dimensional body-cover model in which the lower edge of the folds is represented by a single mass-spring system k, r, m and the propagation of the displacement is represented by a delay line of length T [1], see Fig. 2(a). The structure is a one-mass model with a propagation line aimed at simulating the propagation of the motion along the thickness of the fold, in agreement with the body-cover model proposed by [7]. A second-order resonant filter represents the oscillating fold, a simplified and an impact model reproduces the impact distortions on the fold displacement and adds an offset x_0 (the rest position of the folds).

The areas at entry and exit of the glottis can be respec-



(a) (b)

Fig. 2. Low-dimensional body-cover model of the vocal folds. Panel (a): the vertical displacement of the fold modeled through a single propagation line; panel (b): the vertical and longitudinal displacements of the fold are modeled through a two propagation lines. In both panels, from bottom to top, P_l is the lung pressure, P_m is the driving pressure acting on the vocal folds, m, k, and r represent respectively the mass, stiffness, and damping of the fold, T represents its thickness, x_1 and x_2 are the fold displacements at entrance and exit of the glottis, and P_i is the pressure at entrance of the vocal tract.

tively defined as

$$\begin{aligned} a_1(t) &= 2L(x_{01} + x_1(t)) && (1) \\ a_2(t) &= 2L(x_{02} + x_1(t) - \tau \dot{x}_1(t)) \\ &= 2L(x_{02} + x_2(t)), && (2) \end{aligned}$$

where L is the length of the folds, x_{01} and x_{02} are the rest positions of the fold at entrance and exit to the glottis, and $\tau = T/c_f$ (c_f being the wave velocity on the fold surface) is the time taken by the wave to propagate from the entrance to the upper end of the glottis. The glottal area is finally modeled as the minimum cross-sectional area between the areas at lower and the upper vocal fold edge, i.e., $a = \min\{a_1, a_2\}$. A detailed description of the aerodynamics of the model can be found in the referenced papers [2].

We develop here an extension to this one-delayed mass model by allowing the layer to propagate along two directions: 1. the vertical axis, and 2. the horizontal axis. The scheme is loosely inspired to the 16-mass model introduced in [8], in which an array of two-mass-spring systems is organized longitudinally in order to represent horizontal differences along the length of the cord, and the longitudinal propagation on the body-cover of the fold. The proposed model is shown in Fig. 2(b). The areas at entry and exit of the glottis should be now computed taking into account that the displacement may be not constant along the longitudinal axis:

$$\begin{aligned} a_1(t) &= 2\int_0^L (x_{01} + x_1(l,t))dl && (3) \\ a_2(t) &= 2\int_0^L (x_{02} + x_2(l,t))dl, && (4) \end{aligned}$$

where $x_{1,2}(l,t) = x_{1,2}(0,t) - \tau_l \dot{x}_{1,2}(l,t)$, $\tau_l = L/c_f$.

III. RESULTS

Let us characterize the glottal waveform by means of a set of voice source parameters, allowing us to better evaluate how the new longitudinal displacement parameter affects the shape of the glottal flow pulse. Figure 3 shows the time instants usually defined for a glottal cycle, referred to an LF model.

Figure 4 shows three simulations performed with the proposed low-dimensional body-cover model. The parameter τ_l controlling the displacement delay on the longitudinal axis is gradually increased from left to right. It can be noticed that changes in this parameter mainly affect the closing phase of the glottal cycle. More precisely, the *return time*, i.e. the time interval between the minimum of the flow derivative at time instant t_e and the closing instant t_c (see Fig. 3), scales with τ_l.

Typical voice source quantification parameters extracted from the flow and the differentiated flow are *direct* ones, such as P (the glottal cycle period), $F_0 = 1/P$ (the fundamental frequency of oscillation), t_o (the opening instant), t_p (the maximum flow amplitude instant), t_e (the negative peak instant), t_c (the closing instant), and *derived* ones, such as the speed quotient SQ, the open quotient OQ, the opening quotient $OingQ$, the closing quotient $CingQ$, the return quotient RQ. For our discussion we focus on the following ones, which are among the most used in the literature [9]: return quotient $RQ = (t_c - t_e)/P$, open quotient $OQ = (t_e - t_o)/P$, and speed quotient $SQ = (t_p - t_o)/(t_c - t_p)$. The return quotient is directly related to the return phase duration, the open quotient is directly related to the duration of the open glottis interval that precedes the closure instant, and speed quotient is a measure of the ratio of the open phase to the closing and return phases. Most of these cues have been recognized to be particularly relevant for the study of the perceptual

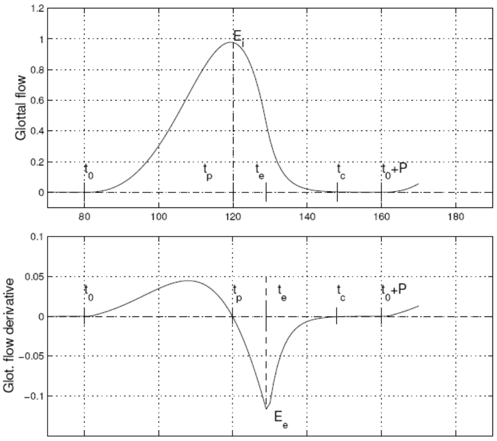
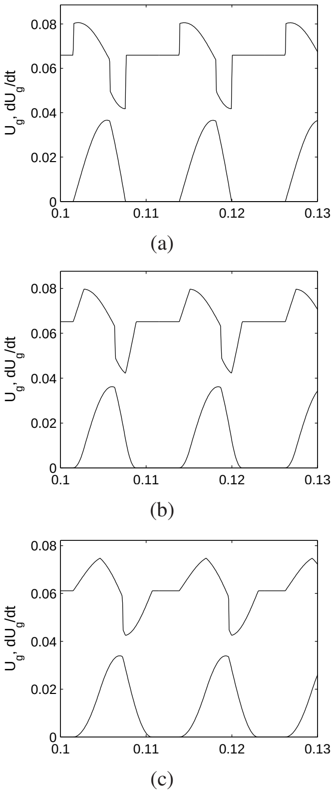


Fig. 3. Glottal flow parameters referred to the LF model: time of glottal opening t_o; time and value t_p,E_i of flow maximum; time and value t_e,E_e of flow derivative minimum; time of glottal closure t_c; glottal period P.

influence of the voice source characteristics, and for comparing different voice qualities (e.g., [10], [11]).

Analytical models, such as the LF, are widely appreciated due to their effectiveness in controlling the voice source parameters. One of the advantages is the possibility of controlling each source parameter by acting on a well identified analytical parameter. The return phase is typically easier to control in analytical models than in physical ones, where the simplifications in the representation of the folds collision usually results in an abrupt closure, corresponding to $RQ = 0$. We focus on this aspect and compare the improved low-dimensional physical model with the LF class. To this aim, a set of 9 glottal flow waveforms was generated both by an LF model and by the proposed model. In both cases the parameter related to the return phase was increased for each next run. The result of the simulations and of the computation of the voice source parameters is shown in Fig. 5.

The two sets of waveforms are characterized by same period length, and approximately same OQ values. Due to the differences in the shape of the pulse of the two models, it was more difficult to obtain similar values for the SQ parameter. The first thing that can be observed by comparing panels a) and b) of Fig. 5 is that in both models the parameter used to control the return phase does not produce appreciable pitch variations. If this property is an obvious one for analytical models, in which the period length is analytically imposed, the same behavior is not necessarily granted for a physical model, in which each component in the dynamic loop may potentially affect the stability and the frequency of oscillation. It has been observed, for instance, that changing the length T of the delay line representing the thickness of the fold, may affect only the duration of the closed phase in some circumstances, or both the duration of the closed phase and of the overall period (i.e., it may affect the pitch). In all the experiments conducted on the proposed model, no appreciable pitch variations were observed in response to variations of the parameter τ_l representing the length L of the fold.



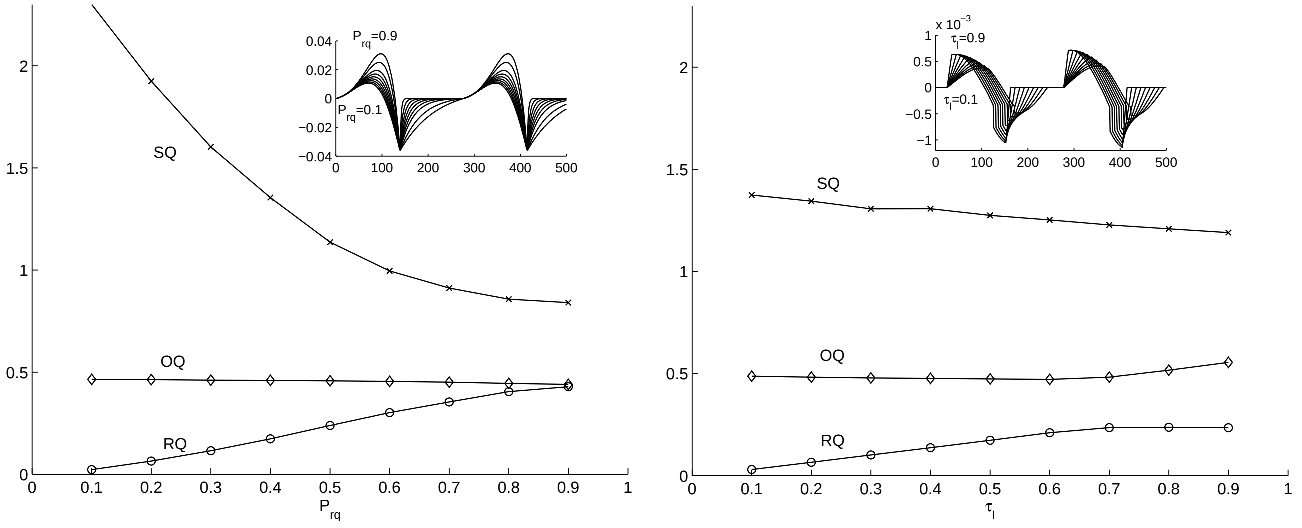
(a)

(b)

(c)

Fig. 4. Result from numerical simulations performed with the proposed low-dimensional body-cover model. Values of τ_l increase from panel (a) to panel (c).

The comparison of panels a) and b) of the same figure leads to other considerations. It can be seen that, as the control parameters P_{rq} and τ_l are raised, the behavior of the three source parameters considered here, RQ, OQ, and SQ, is qualitatively the same: RQ increases as expected in both models, even if with this configuration of the low-dimensional glottal model it was not possible to reach the same values around 0.5 obtained with the LF one (instability of the oscillation was observed if the parameter τ_l was further increased). Such high return quotient values are however rarely observed in natural glottal flow recordings. The OQ parameter is approximately constant as expected (note from the definition that the return phase does not contribute to the open quotient OQ), except in the right-most part of the plot, where the curve rises slightly. Finally, both curves representing the SQ parameter show a decreasing trend, although the range spanned by the plot related to the LF model is appreciably larger than the range



(a) (b)

Fig. 5. Glottal flow parameters computed on a set of 9 waveforms obtained by running the LF model (panel a)) and the proposed low-dimensional body-cover model (panel b)) while increasing the parameter responsible for the duration of the return phase (parameters on the x axes are normalized).

spanned by the plot related to the low-dimensional model. Apparently, there is no straightforward explanation for this inequality, apart from the evident differences in the pulse shape.

In conclusion, some interesting properties of this class of physical models emerged from the comparison with analytical models, and further investigation will be conducted in this direction.

IV. CONCLUSIONS

We proposed an extension to the mechanical component of a low-dimensional vocal fold model previously introduced, and we discuss the affectiveness of the new scheme in terms of control of glottal flow cues providing comparisons with the LF analytical model. The additional degree of freedom introduced with this new scheme allows to control some relevant features of the glottal flow waveform, such as the return quotient, that are not directly accessible with similar models previously proposed in the literature. Future research on this class of models is foreseen with respect to a number of issues, including: 1. the perceptual assessment of the synthesis to gain understanding on the perceptual relevance of the new parameters in terms of naturalness of the synthesis and of voice quality controllability, 2. the refinement of the low-dimensional model to adapt its glottal pulse shape to the characteristics of the LF model, thus allowing improved comparisons between the two classes of models, and 3. the design of automatic parametric adaptation algorithms to fit the model to real glottal waveforms.

REFERENCES

[1] F. Avanzini, P. Alku, and M. Karjalainen, "One-delayed-mass model for efficient synthesis of glottal flow," *Proc. of Eurospeech Conf*, pp. 51–54, September 2001.

[2] C. Drioli, "A flow waveform-matched low-dimensional glottal model based on physical knowledge," *J. Acoust. Soc. Am.*, vol. 117, no. 5, pp. 3184–3195, 2005.

[3] F. Avanzini, S. Maratea, and C. Drioli, "Physiological control of low-dimensional glottal models with applications to voice source parameter matching," *Acta Acustica united with Acustica*, vol. 92, no. Suppl.1, pp. 731–740, Sep. 2006.

[4] C. Drioli and F. Avanzini, "Non-modal voice synthesis by low-dimensional physical models," *in Proc. 3rd Int. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)*, 2003.

[5] K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," *The Bell Syst. Tech. J.*, vol. 51, no. 6, pp. 1233–1268, July-August 1972.

[6] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," *STL-QPS*, pp. 1–13, 1985.

[7] I. R. Titze, "The physics of small-amplitude oscillations of the vocal folds," *J. Acoust. Soc. Am.*, vol. 83, no. 4, pp. 1536–1552, April 1988.

[8] I. R. Titze and W. J. Strong, "Normal modes in vocal cord tissues," *The Journal of the Acoustical Society of America*, vol. 57, no. 3, pp. 736–744, 1975.

[9] P. Alku and E. Vilkman, "A comparison of glottal voice quantification parameters in breathy, normal and pressed phonation of female and male speakers," *Folia Phoniatr. Logop.*, vol. 48, no. 5, pp. 240–254, Sep. 1996.

[10] A. E. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels," *J. Acoust. Soc. Am.*, vol. 49, no. 2, pp. 583–590, February 1971.

[11] D. Childers and C. Ahn, "Modeling the glottal volume-velocity waveform for three voice types," *J. Acoust. Soc. Am.*, vol. 97, no. 1, pp. 505–519, January 1995.