

SELF-OSCILLATING MODELS OF THE TONGUE TIP FOR SIMULATING ALVEOLAR TRILLS

Benjamin Elie

Loria, Inria/CNRS/Université de Lorraine
email:benjamin.elie@loria.fr

Yves Laprie

Loria, Inria/CNRS/Université de Lorraine

Alveolar trills are sounds produced by the self-oscillation of the tongue tip in the alveolar region. Specific aeroacoustic conditions around the lingual constriction are required to produce such sounds, but have not been fully investigated due to the lack of physical modeling. This study discusses the possibility to model the tongue tip behavior with lumped mass-spring systems similar to those used for the vocal folds. It compares a single mass model to a two-mass model with mobile air flow separation point. The tongue tip model is introduced into a transmission line circuit analog model of the vocal tract to which a self-oscillating model of the vocal folds is connected, in order to accurately reproduce the aerodynamic conditions at the vicinity of the lingual constriction. Results show that, although the phase shift between masses with the two-mass model is very small, it differs with the single mass model by accounting for less acoustic interactions with the vocal tract and the vocal folds, and proves to be more robust for the simulation of alveolar trills than the single mass model. The study also proposes to model the incomplete closure of the vocal tract during contacts between the tongue tip and the hard palate in the oscillation process, by connecting an acoustic waveguide both downstream and upstream of the lingual constriction, producing a lateral air leakage. The model presented in the paper is then a useful tool to study the articulatory and physiological conditions required for the production of alveolar trills.

Keywords: Alveolar trills, Self-oscillations, Articulatory synthesis

Introduction

Alveolar trills are characterized by a periodic modulation of the amplitude of the acoustic pressure waveform radiated at the lips. This amplitude modulation is known to be due to the self-oscillations of a target articulator, namely the tongue tip in this case. They are driven by the aerodynamic forces applied at its vicinity, and this requires specific articulatory and aerodynamic conditions to produce alveolar trills. Curiously, only a few studies have attempted to define these conditions by studying the acoustic phenomena involved in trill production. Actually, except for studies on acoustic properties of the speech signal [1, 2], or on the linguopalatal patterns [3], the simplified model by McGowan [4] is, to the best of our knowledge, the sole attempt to study the production of trills using numeric simulations. Although he used a single degree of freedom system to represent the tongue-tip oscillations, he also discussed the possibility to include a lumped mass spring system derived from the classic two-mass model used for the vocal folds [5], since it presents the advantage to simulate phase shifts between the upstream and downstream parts of the vocal folds, as observed in real world experiments, and also to account for a mobile flow separation point [6, 7]. This paper investigates this question,

in Sec. 4, by comparing the mechanical behavior of the single-mass model, similar to the one introduced by Flanagan and Langraf [8], and a two-mass model with smooth contours and a mobile flow separation point [7]. They are detailed in Sec. 2.

The paper also proposes to account for the potential incomplete closure of the vocal tract during linguopalatal contacts by connecting a lateral acoustic waveguide connected both at the upstream and the downstream parts of the tongue. This lateral connection is made possible thanks to the recent Extended Single-Matrix Formulation (ESMF) of the vocal tract [9].

Acoustic model

The simulation framework to compute the acoustic propagation inside the vocal tract is derived from the *Transmission Line Circuit Analog* (TLCA) approach [10]. It considers plane waves propagating along a spatially sampled vocal tract, modeled as a set of connected acoustic tubes, or *tubelets*. Unlike the other widely used approach, the *Reflection Type Line Analog* (RTLCA) model [11, 12], it easily deals with time-varying lengths of the vocal tract, and also with uneven spatial sampling of the vocal tract. For further information, the reader may find a detailed review of existing techniques for speech synthesis in [13]. The framework that is used in this paper considers recent improvements of TLCA-based techniques [14, 9], such as the possibility to connect self-oscillating models of the vocal folds with a glottal chink. In this paper, the acoustic model is extended to include the potential oscillations of the tongue tip. Note that a very simplified vocal tract geometry (a cylinder with a narrow constriction representing the lingual constriction) is used, but the acoustic model also deals with realistic geometries. The choice of a simplified geometry is motivated by the focus on the mechanical behavior of the lumped mass-spring models of the tongue tip.

Acoustic propagation

The vocal tract is seen as a concatenation of acoustic tubelets having cross-sectional areas that approximate the vocal tract area function. It has been shown [9], following TLCA-based techniques [10, 14], that the wave propagation is driven by the following system of equations

$$\mathbf{f} = \mathbf{Z}\mathbf{u}_Z + \mathbf{Q}\mathbf{u}_Q, \quad (1)$$

where $\mathbf{f} \in \mathbb{R}^{(N+1)}$ is a vector containing pressure forces, $\mathbf{Z} \in \mathbb{R}^{(N+1) \times (N+1)}$ is a tridiagonal matrix containing impedance and loss terms associated to each tubelets, where \mathbf{Q} is a square matrix the same size as \mathbf{Z} having only one non-zero element, that is $Q_{(1,1)} = R_b$, $\mathbf{u} \in \mathbb{R}^{N+1}$ is the vector containing the volume velocities inside each tubelet, and $\mathbf{u}_Q \in \mathbb{R}^{(N+1)} = [U_1^2, U_2^2, \dots, U_N^2]^T$ is the vector containing the square power of the volume velocities. The term R_b is the Bernoulli resistance, as defined in [9].

Self-oscillation models of the tongue tip

Self-oscillating models of the tongue tip are represented in Fig. 1. The single-mass model (SMM) may be sufficient to simulate the behavior of the tongue tip during the production of alveolar trills. Indeed, the oscillation frequency of such trills is no more than 35 Hz [2], which is much below the first formant frequency of the human vocal tract, which is one necessary condition for the SMM to oscillate [15, 16]. However, the SMM has been reported to create unrealistic acoustic interactions between the vocal folds and the vocal tract [17], which could be also be true between the vocal tongue tip and the rest of the vocal system. This will be discussed in the next section.

The two-mass model (TMM) has the advantage to simulate higher modes than the sole transverse mode simulated by the SMM. It thus allows phase shift between the upstream and the downstream masses, as observed in the motion of the vocal folds in natural speech. For tongue tip oscillations, this has been previously discussed by McGowan [4], but there is, so far, no evidence of such behavior

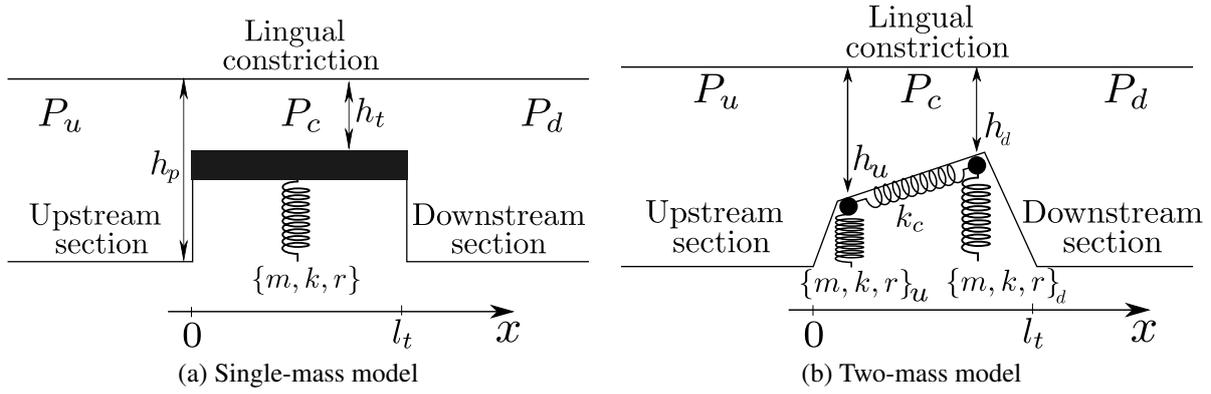


Figure 1: a) single-mass model for the tongue tip, and b) the two-mass model. P_d , P_u , and P_c denote the pressure downstream, upstream and inside the lingual constriction, respectively. The mass, stiffness and the damping of the tongue are denoted by the triplet $\{m, k, r\}$. The height of the lingual constriction is denoted by h_t . Indices u and d denote the upstream and downstream parts respectively.

of the tongue tip during trill productions. The TMM used for the tongue tip is then similar to the one used for the vocal folds. It includes smooth contours [6, 7], and the computed pressure forces account for viscous losses and unsteady flow effects [18, 19].

The equation of motion are, for the SMM

$$m\ddot{h}_t(t) + r\dot{h}_t(t) + k\Delta h_t(t) = F(t), \quad (2)$$

and for the TMM

$$\begin{cases} m_u\ddot{h}_u(t) + r_u\dot{h}_u(t) + (k_u + k_c)\Delta h_u(t) - k_c\Delta h_d(t) = F_u(t) \\ m_d\ddot{h}_d(t) + r_d\dot{h}_d(t) + (k_d + k_c)\Delta h_d(t) - k_c\Delta h_u(t) = F_d(t) \end{cases}, \quad (3)$$

where m , r , k are respectively the mass, the damping, and the stiffness of the tongue tip, and F is the pressure forces that are applied to the system. The elongation is $\Delta h = h - h_0$, where h_0 denotes the rest position of the tongue. Indices u and d denote the upstream and the downstream parts, respectively.

Pressure forces are defined as

$$F = \int_0^{l_t} P(x)dx, \quad (4)$$

where l_t is the length of the lingual constriction. The pressure distribution along the lingual constriction is given by

$$P(t) = P_u(t) + Be(t) + Po(t) + In(t) \quad (5)$$

where $Be(t)$, $Po(t)$, and $In(t)$ are respectively the steady term of the Bernoulli equation, the Poiseuille corrective term and the unsteady term of the Bernoulli equation.

In the case of the SMM, they are defined as:

$$\begin{aligned} Be(t) &= -\frac{\rho_a U_t^2(t)}{2w_t^2} \left[\frac{1}{h_t^2(t)} - \frac{1}{h_p^2(t)} \right], \\ Po(t) &= -\frac{12\mu_a l_t U_t(t)}{w_t h_t^3(t)}, \\ In(t) &= -\frac{\rho_a l_t}{w_t} \frac{\partial}{\partial t} \left[\frac{U_t(t)}{h_t(t)} \right], \end{aligned} \quad (6)$$

where w_t and l_t are respectively the width and the length of the lingual constriction, ρ_a and μ_a are respectively the mass density and the shear viscosity of the air. The height h_p is the height of the section just downstream of the lingual constriction.

In the case of the TMM, the potential phase shift between both masses modifies the expressions of $Be(x, t)$, $Po(x, t)$, and $In(x, t)$ in Eq. (6), which depend on the position along the x -coordinate:

$$\begin{aligned} Be(x, t) &= -\frac{\rho U_t^2(t)}{2w_t^2} \left[\frac{1}{h_t^2(x, t)} - \frac{1}{h_t^2(x_0, t)} \right], \\ Po(x, t) &= -\frac{12\mu U_t(t)}{w_t} \int_{x_0}^x \frac{dx}{h_t^3(x, t)}, \\ In(x, t) &= -\frac{\rho}{w_t} \frac{\partial}{\partial t} \left[U_t(t) \int_{x_0}^x \frac{dx}{h_t(x, t)} \right]. \end{aligned} \quad (7)$$

Eq. (7) is valid for x located upstream of the separation point. Indeed, in this paper, similarly to the vocal folds model, the TMM includes a mobile separation point x_s , where x_s is such that $h_s = h(x_s) = 1.2h_u$ if $1.2h_u < h_d$, and $x_s = x_d$ otherwise. The value of 1.2 is a ad hoc criterion [7]. The pressure $P(x, t) = P_d(t)$ if $x > x_s$.

The collision model uses an artificial stiffening of the tongue tip, namely the stiffness parameters k_u and k_d are multiplied by a factor 4 if h_u , or h_d , are less than or equal to 0, respectively, and the damping is multiplied by a factor 2.2. These factors are also ad hoc criteria, chosen following the original TMM by Ishizaka and Flanagan [5].

Data

Area function and tongue-tip parameters

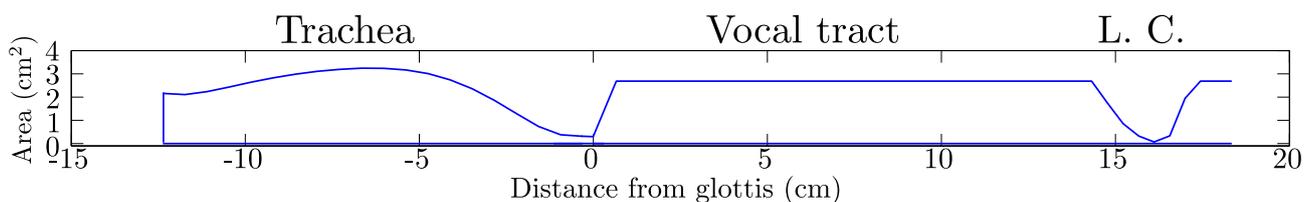


Figure 2: Area function of the vocal system, from the lung (left) to the lips (right). The lingual constriction is denoted by L.C.

The vocal tract area function used for this study is shown in Fig. 2. Note that we use a subglottal system represented by a waveguide with an area function borrowed from [12]. The stiffness values have been chosen according to the mass values so that the resonance frequency of the uncoupled mechanical system is 25 Hz, which is in the range of typical values observed in real speakers [2]. The chosen values for the model are summarized in Tab. 1. They have been derived empirically so that it creates oscillations with linguopalatal contacts and remains physiologically realistic.

Comparison between lumped mechanical models

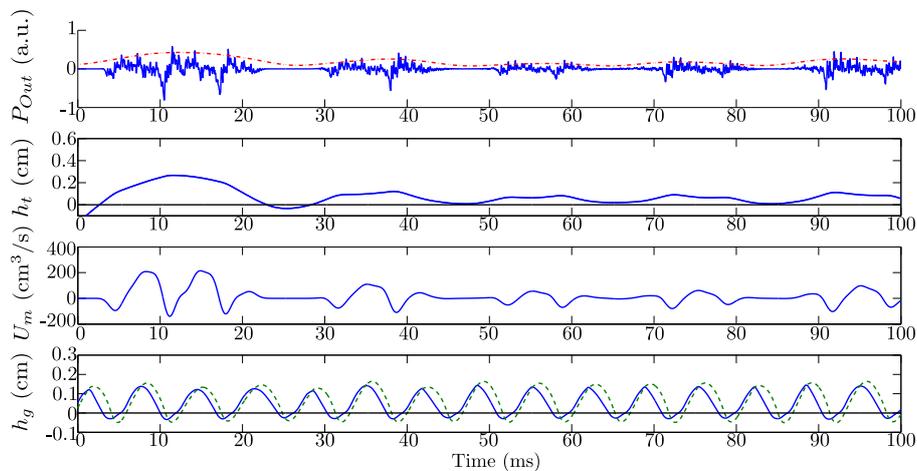
Using the area function shown in Fig. 2, sustained alveolar trills are simulated, first with the SMM, and then with the TMM. The parameters of the TMM are detailed in Tab. 1.

Results of the simulations are displayed in Fig. 3. It shows the oscillations of the tongue tip position around its rest position. When linguopalatal contacts occur, this completely stops the air flow inside the vocal tract, hence the zero values of the air flow at the mouth U_m and the acoustic pressure P_{Out} . In both cases, the time envelope of the acoustic pressure is then a periodic function, which is at its minimum when the tongue tip is close to the hard palate (or even is in contact with), and at its maximum when the tongue tip is far from the hard palate.

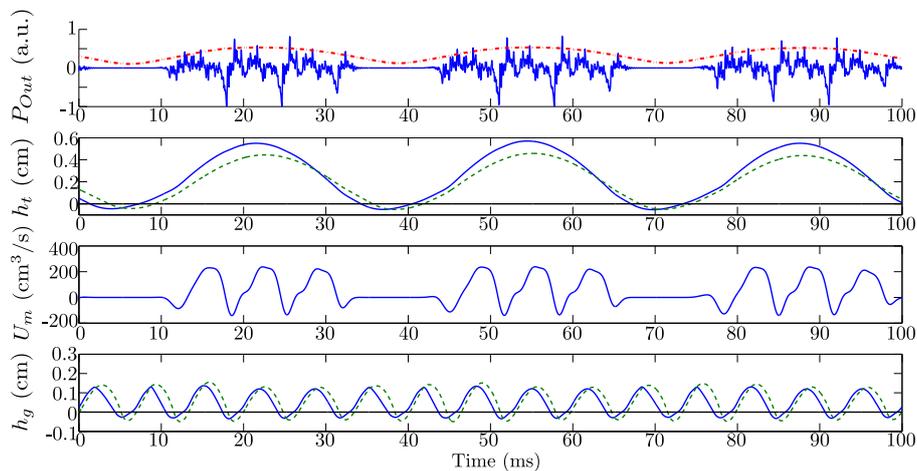
Significant differences are clearly visible between the results provided by the models. Firstly, the tongue oscillation frequency f_t of the SMM is much larger than the one of the TMM (45 Hz versus 29

Table 1: Parameters of the two-mass model of the tongue tip.

Parameter	Unit	Value
Subglottal pressure P_{sub}	Pa	1000
Rest position of the tongue tip h_0	mm	0.7
Length of the lingual constriction l_t	mm	4.4
Width of the lingual constriction w_t	mm	10.5
Mass $m = m_1 = m_2$	g	0.26
Stiffness $k = k_1 = k_2$	N/m	228
Nominal damping coefficient r_i	kg.rad.s ⁻¹	$0.02\sqrt{k.m/2}$
Coupling spring k_c	N/m	$k/2$



(a)



(b)

Figure 3: Simulation of alveolar trills with linguopalatal contacts by using a) a single-mass model, and b) a two-mass model. For each case, from top to bottom are plotted the acoustic pressure radiated at the lips P_{Out} , as well as the time envelope (dot-dashed line), the time evolution of the lingual constriction height h_t (the dashed line in b) represents the downstream part of the tongue tip in the two-mass model), the low frequency filtered airflow at the mouth U_m , and the displacement of the upstream (solid line) and downstream (dashed line) parts of the vocal folds h_g .

Hz). Secondly, the tongue tip displacement waveform computed with the SMM is much less smooth

than the waveform given by the TMM. Indeed, high-order harmonics seem to significantly contribute to the displacement waveform. The main acoustic consequence is a more erratic time envelope of the acoustic waveform. One possible explanation of the presence of these high-order harmonics could be an excessively high coupling between the tongue tip and the vocal folds. Indeed, since the oscillations of the tongue tip are driven by the acoustic pressure upstream of the lingual constriction, it is very likely that the oscillations of the vocal folds disturb the oscillations of the tongue tip.

Simulations of incomplete closure of the vocal tract at the tongue palate contact instants

Previous works about alveolar trills [2, 4] report airflow at the mouth which is not null during the closed phase of the trill. The same observation applies for the acoustic pressure waveforms as well. A possible explanation is that there is no complete occlusion of the vocal tract during linguopalatal contacts, due to air leakage around both sides of the tongue tip. One idea to be able to simulate this phenomenon is to add a lateral acoustic branch to simulate the air path going through obstacle created by the tongue-palate contact. The ESMF paradigm that is used in this paper allows such modeling [9]. With this simplified model, the area function and the length of the lateral branch are the same as those of the lingual constriction branch when the tongue is at rest, according to an aspect ratio r_l . The upstream and the downstream connections are taken at the location where the area function starts to decrease, i.e. at 14.7 cm and 17 cm from the glottis, respectively. Presented simulations use the TMM with the same mechanical parameters as in Tab. 1 in a vocal tract modeled with a lateral branch connected to the main path with an aspect ratio r_l of 50% (the lateral branch has the same cross-sectional area as the lingual constriction at rest).

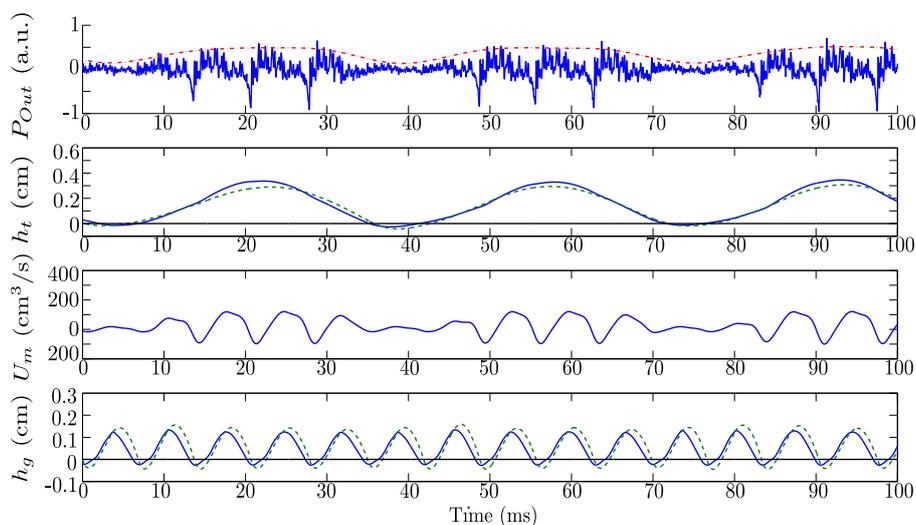


Figure 4: Results of the simulation of a sustained alveolar trill with a lateral configuration, using the two-mass model, and $r_l = 50\%$. From top to bottom are plotted the acoustic pressure radiated at the lips P_{Out} , as well as the time envelope (dot-dashed line), the time evolution of the lingual constriction height h_t at the upstream (solid line) and at the downstream (dashed line) parts of the tongue tip, the low frequency filtered airflow at the mouth U_m , and the displacement of the upstream (solid line) and downstream (dashed line) parts of the vocal folds h_g .

The simulation with a ratio $r_l = 50\%$, displayed in Fig. 4, shows that this vocal tract configuration enables linguopalatal contact without complete closure of the vocal tract to be simulated. When the tongue tip and the hard palate are in contact ($h_t \leq 0$), the air flow U_m is not null, nor is the acoustic signal pressure. This can be compared with Fig. 3 b), which shows the results of a simulation with similar parameters, but without the lateral waveguide.

Conclusion

This paper has described the mechanical behavior of lumped mass-spring systems when used to simulate the self-oscillations of the tongue tip during alveolar trills. One important point is the response of the models to the oscillations of the pressure upstream of the lingual constriction due to the vibrations of the vocal folds. Simulations have shown that the two-mass model is less sensitive to the vocal folds, and that the mass-spring system oscillates almost independently. Connecting a lateral branch to the lingual constriction allows incomplete closure of the vocal tract to be simulated during linguopalatal contacts. Thus, the commonly observed features of non-zero airflow and acoustic pressure during the closed phase can be reproduced. Using a lumped two-mass system to simulate the tongue tip oscillations in alveolar trills may then be very useful to investigate the conditions required for their production. We are currently using this simulation framework with realistic geometries of the vocal tract, derived from cineMRI [20], to investigate the impact of various parameters on the tongue tip behavior.

Acknowledgements

This study is supported by the ANR (*Agence Nationale de la Recherche*) ArtSpeech project.

References

1. Mendoza, E., Gloria Carballo. Acoustic characteristics of trill productions by groups of spanish children, *clinical linguistics & phonetics*, **14** (8), 587–601, (2000).
2. Solé, M.-J. Aerodynamic characteristics of trills and phonological patterning, *J. of Phon.*, **30**, 655–688, (2002).
3. Recasens, D. and Pallarès, M. D. A study of /t/ and /t̪/ in the light of the "DAC" coarticulation model, *Journal of Phonetics*, **27**(2), 143–169, (1999).
4. McGowan, R. S. Todayngue-tip trills and vocal-tract wall compliance, *J. Acoust. Soc. Am.*, **91**(5), 2903–2910, (1992).
5. Ishizaka, K. and Flanagan, J. L. Synthesis of voiced sounds from a two-mass model of the vocal cords, *Bell Syst. Tech. J.*, **51**(6), 1233–1268, (1972).
6. Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J. and Auregan, Y. Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model, *J. Acoust. Soc. Am.*, **96**(6), 3416–3431, (1994).
7. Lous, N. J. C., Hofmans, G. C. J., Veldhuis, R. N. J. and Hirschberg, A. A symetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prothesis design, *Acta Acustica*, **84**, 1135–1150, (1998).
8. Flanagan, J. and Landgraf, L. Self-oscillating source for vocal-tract synthesizers, *IEEE Transactions on Audio and Electroacoustics*, **16** (1), 57–64, (1968).
9. Elie, B. and Laprie, Y. Extension of the single-matrix formulation of the vocal tract: Consideration of bilateral channels and connection of self-oscillating models of the vocal folds with a glottal chink, *Speech Communication*, **82**, 85–96, (2016).
10. Maeda, S. A digital simulation method of the vocal-tract system, *Speech Communication*, **1**, 199–229, (1982).

11. Kelly, J. L. and Lochbaum, C. C. Speech synthesis, *Proceedings of the Fourth International Congress on Acoustics*, pp. 1–4, (1962).
12. Story, B. H. Phrase-level speech simulation with an airway modulation model of speech production, *Computer Speech & Language*, **27(4)**, 989–1010, (2013).
13. Kröger, B. J. and Birkholz, P., (2009), Articulatory synthesis of speech and singing: State of the art and suggestions for future research. *Multimodal Signals: Cognitive and Algorithmic Issues*, pp. 306–319, Springer.
14. Mokhtari, P., Takemoto, H. and Kitamura, T. Single-matrix formulation of a time domain acoustic model of the vocal tract with side branches, *Speech Communication*, **50(3)**, 179 – 190, (2008).
15. Titze, I. R. The physics of small-amplitude oscillation of the vocal folds, *J. Acoust. Soc. Am.*, **83(4)**, 1536–1552, (1988).
16. Zhang, Z., Neubauer, J. and Berry, D. A. Aerodynamically and acoustically driven modes of vibration in a physical model of the vocal folds, *The Journal of the Acoustical Society of America*, **120** (5), 2841–2849, (2006).
17. Cataldo, E., Leta, F. R., Lucero, J. and Nicolato, L. Synthesis of voiced sounds using low-dimensional models of the vocal cords and time-varying subglottal pressure, *Mechanics Research Communications*, **33(2)**, 250–260, (2006).
18. Vilain, C., Pelorson, X., Fraysse, C., Deverge, M., Hirschberg, A. and Willems, J. Experimental validation of a quasi-steady theory for the flow through the glottis, *J. of Sound and Vibration*, **276(3-5)**, 475 – 490, (2004).
19. Bailly, L., Pelorson, X., Henrich, N. and Ruty, N. Influence of a constriction in the near field of the vocal folds: Physical modeling and experimental validation, *J. Acoust. Soc. Am.*, **124(5)**, 3296–3308, (2008).
20. Elie, B., Laprie, Y., Vuissoz, P.-A. and Odille, F. High spatiotemporal cineMRI films using compressed sensing for acquiring articulatory data, *Eusipco, Budapest*, August, pp. 1353–1357, (2016).