



**HAL**  
open science

# Simulating alveolar trills using a two-mass model of the tongue tip

Benjamin Elie, Yves Laprie

► **To cite this version:**

Benjamin Elie, Yves Laprie. Simulating alveolar trills using a two-mass model of the tongue tip. Journal of the Acoustical Society of America, In press, 142 (5). hal-01525882v2

**HAL Id: hal-01525882**

**<https://hal.science/hal-01525882v2>**

Submitted on 16 Aug 2017 (v2), last revised 8 Nov 2017 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Simulation of alveolar trills using lumped self-oscillating models of the tongue tip

Benjamin Elie<sup>1, a)</sup> and Yves Laprie<sup>1</sup>

*LORIA, INRIA/CNRS/Université de Lorraine, Nancy,  
France*

(Dated: August 16, 2017)

This paper investigates the possibility of reproducing the self-sustained oscillations of the tongue tip in alveolar trills. The interest is to study the articulatory and phonatory configurations that are required to produce alveolar trills. Using a realistic geometry of the vocal tract, derived from cineMRI data of a real speaker, the paper studies the mechanical behavior of a lumped two-mass model of the tongue tip. Then, the paper proposes a solution to simulate the incomplete occlusion of the vocal tract during linguopalatal contacts by adding a lateral acoustic waveguide. Finally, the simulation framework is used to study the impact of a set of parameters on the characteristic features of the produced alveolar trills. It shows that the production of trills is favored when the tongue tip position is slightly away of the alveolar zone, and when the glottis is fully adducted.

PACS numbers: 43.70.Bk, 43.72.Ja

---

<sup>a)</sup>benjamin.elie@inria.fr

## I. INTRODUCTION

During the production of trill consonants (bilabial, alveolar, or uvular), aerodynamic forces applied to a particular articulator (lips, tongue tip, and uvula, respectively) are known to be responsible for the oscillation of the target articulator. Its oscillation leads to a periodic amplitude modulation of the acoustic pressure waveform radiated at the lips. This periodic amplitude modulation is the main characteristic feature of trills.

This paper focuses on the mechanical and acoustic modeling of the production of alveolar trills, which are produced by the oscillation of the tongue tip in the alveolar region. This trill is the most common across languages<sup>1</sup>, in comparison with other places of articulation. However, this is still a hard-to-produce sound that is mastered late in the acquisition process<sup>2</sup>. This suggests that it requires specific articulatory and aerodynamic conditions, which have been reported to be the reason of assimilation in neighboring phonetic segments<sup>3</sup>. Curiously, only a few studies have investigated the acoustic and aerodynamic characteristics of the production of alveolar trills. Acoustic properties have been deeply detailed via analysis of the speech signal<sup>4</sup> and the intraoral pressure<sup>5</sup>. Linguopalatal contacts, along with the speech signal characteristics have also been investigated<sup>6</sup>.

Despite these works, there is still a lack of acoustic and/or mechanical models of the production of trills. To the best of our knowledge, the work by McGowan<sup>7</sup> is the sole attempt to study the production of trills using numeric simulations. He investigated the mechanisms involved in the self-oscillations of the tongue tip with a very simplified vocal tract. Simulations show that the oscillations of the tongue tip are sustained thanks to the wall compliance of the vocal tract, hence a behavior similar to a one-mass system. The author also briefly discusses the potential impact of classic two-mass lumped models, similar to those used for the vocal folds<sup>8,9</sup>. Indeed, two-mass models (TMM) present the advantage of simulating phase shifts between the upstream and downstream parts of the vocal folds, as observed in real world experiments<sup>8</sup>, and also to take a mobile flow separation point<sup>9,10</sup> into account. One might expect the same behavior for the tongue tip. The reader may find a detailed review of existing mass-spring systems for modeling the self-oscillation of the vocal folds in Birkholz *et al.*<sup>11</sup>, for instance.

In this work, the glottal source is connected to the classic transmission line paradigm<sup>12</sup>, considering recent improvements<sup>13</sup> in order to deal with a modified self-oscillating model of

the vocal folds to account for partial glottal abduction. This is one of the major contribution of the paper in comparison with the work of McGowan<sup>7</sup>, since it enables the influence of the acoustic pressure variations upstream of the lingual constriction due to the vocal folds motion to be investigated with real anatomic data. Area functions have been derived from real-time cineMRI using a sparse reconstruction technique<sup>14</sup>, detailed in Sec. II.

The other contribution is the possibility of the presented simulation framework to simulate the incomplete closure of the vocal tract during linguopalatal contacts. Indeed, results from Solé<sup>5</sup> suggest that there is no complete occlusion during the production of consonant trills, since the low frequency component of the mouth airflow is constantly strictly positive. Consequently, when dealing with acoustic simulations of the alveolar trills, one should guarantee that tongue tip oscillations with, or without contacts can be simulated, and also that linguopalatal contacts do not completely stop the main air path in the vocal tract. In Sec. IV C, the paper presents a solution to deal with this situation by adding 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<sup>13</sup>. The simulation framework including our self-oscillating tongue tip model is finally used to investigate the influence of various parameters on the production of alveolar trills. Results of the simulations are shown and discussed in Sec. V.

## II. ACOUSTIC AND ARTICULATORY CHARACTERISTICS OF ALVEOLAR TRILLS

### A. Acoustic characteristics

Alveolar trills [r] are sounds that are characterized by a periodic oscillation of the tongue tip. The vibration of the articulator is not directly controlled by the speaker, namely it is not due to an active muscular control, but is due to the aerodynamic forces applied at their vicinity<sup>7</sup>. Studies about the production of alveolar trills include the mastering process<sup>2</sup> by children, occurrence in world's languages<sup>1</sup>, coarticulatory effects<sup>3,15,16</sup>, acoustic characteristics<sup>4,5</sup>, and air-tissue interaction modeling<sup>7</sup>.

From these studies, it appears that alveolar trills are difficult to produce<sup>2</sup>, and may interfere with neighboring phonetic segments<sup>3</sup>. Aerodynamic measurements on speakers show that a

high intraoral pressure is required to produce trills, and even higher for voiceless alveolar trills<sup>5</sup>. The intraoral pressure show significant variability among subjects<sup>5</sup>. The oscillation frequency is in the range 28-33 Hz for voiceless trills and 26-29 Hz for voiced trills<sup>5</sup>. The open phase is 1.94 times longer than the closed phase for voiceless trills, and 1.34 times longer for voiced trills.

Electropalatographic data (EPG) show that linguopalatal contacts occur during the production of consonant trills<sup>3,6,16</sup>. However, contact patterns significantly vary across speakers, and even across repetitions of a single speaker. For a few speakers, contacts occur only at the edge of the tongue dorsum, and the alveolar region does not show any contact between the tongue tip and the hard palate. Hence, there is no evidence of systematic contacts between the tongue and the hard palate, or the teeth, during the oscillation cycle. It is not unlikely that the tongue tip freely vibrates slightly away from the alveolar zone. Additionally, in all studies<sup>4,5</sup>, the presented acoustic waveforms during the so-called closed phase is not null, as well as the airflow measured at the mouth. This suggests that trills do not involve complete occlusion. It is also clear in Fig. 3 of the paper of Solé<sup>5</sup> that the mouth airflow is systematically strictly positive during the trill production.

Consequently, when dealing with numeric simulations, one should consider the possibility of taking into account the incomplete closure of the vocal tract during linguopalatal contacts.

Fig. 1 shows two examples of acoustic speech signal of alveolar trills uttered in the pseudoword /ara/. The time envelope of the acoustic pressure waveform shows periodic cycles during the production of the alveolar trill [r]. It oscillates between a maximum, at a level slightly less than the adjacent vowels, and a minimum, which is not null.

## **B. Articulatory characteristics**

Observing tongue tip movements during alveolar trills is very difficult. Indeed, since the oscillation has a small amplitude and is relatively fast, it requires high spatiotemporal resolutions. Thus, high-speed Ultrasound imaging and Electromagnetic articulography (EMA)<sup>17</sup> have been used to observe the time evolution of the tongue tip position during alveolar trills. These studies provide data from alveolar trill realizations of German<sup>17</sup>, Czech<sup>18</sup>, Spanish, and Russian speakers<sup>19</sup>. All of them show oscillation amplitudes of a few millimeters. EMA measurements also show that the tongue exhibits a flatter shape in the coronal plane during

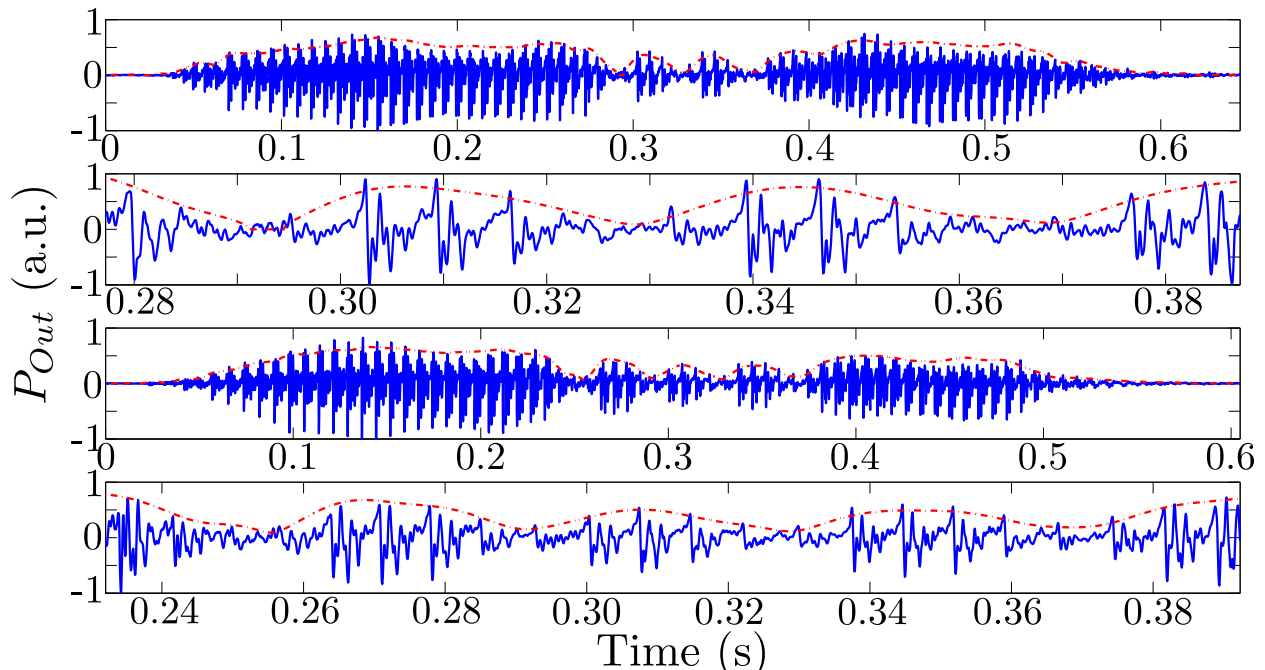


Figure 1. Audio speech signal of two utterances of the pseudoword /ara/, with a zoomed-in version of the alveolar trill [r]. The time envelope is plotted as a dashed line

trills in comparison with fricatives<sup>18</sup>.

The simulation framework presented in this study requires the whole area function of the vocal tract. For that purpose, the vocal tract geometry has been observed via real-time cineMRI techniques. It consists in a reconstruction of the midsagittal view of the vocal tract at a reasonably high frame rate, using sparse reconstruction methods<sup>14</sup>. The frame rate is 29 frames per second, with a spatial resolution of  $1 \times 1 \text{ mm}^2$ . MRI experiments were performed on a 3T Signa HDxt MR system (General Electric Healthcare, Milwaukee, WI). Speech MRI data were obtained from a healthy volunteer with written informed consent and approval of local ethics committee. He is a 35 years old male and a French native speaker. Although French does not have alveolar trills, our speaker's ability of uttering alveolar trills comes from a second language acquisition (Spanish) at an intermediate fluency level. The data were collected with a 16-channel neurovascular coil array. The protocol consisted in a sagittal slice through the middle of the vocal tract acquired with a custom modified Spoiled Fast Gradient Echo (Fast SPGR, TR 3.5ms, TE 1.1ms, line BW 83.33 kHz, flip angle 30 degrees, half echo in frequency direction, matrix  $256 \times 256$ , 512 temporal frames). For each temporal frame the modification consisted in acquiring only a randomized fixed size sample

(10 lines) of the 256 phase lines of the  $k$ -space. The alveolar trill is pronounced in the intervocalic context /ara/.

We considered that acquiring MR images at 29 frames per second, namely similar to the trill frequency, is sufficient to get a good idea of the global position of the tongue tip during oscillations along with the rest of the vocal tract since we acquire several images per /r/ segment. For the study, the speaker was asked to slightly lengthen the duration of the /r/ segment, so that between 5 and 10 images per phonetic segment are available.

Fig. 2 shows an example of the midsagittal view of the vocal tract during the production of an alveolar trill in the pseudoword /ara/. All the contours which are sufficiently invariant because they correspond to a non-deformable structure are tracked by correlation with a reference image chosen in the first images of the sequence. This is the case of the incisor roots for locating the mandible, and the upper part of the head and the vertebra used for locating the palate and pharyngeal walls. Deformable contours presenting a small number of deformation modes, like lips and larynx are tracked by means of a semi automatic method inspired from Fontecave and Berthommier<sup>20</sup>. The main principle is to delineate the target contour in key images randomly selected in the cineMRI, and then to index images where the contour has to be tracked with respect to the key images by using a Discrete Cosine Transform (DCT) distance. The image used for indexing images is limited to the region where the contour to be tracked is located. The resulting contour is obtained by averaging contours of the closest images. This semi-automatic tracking requires supervision from the user who can add key images to correct errors corresponding to an image too far from the existing key images. Finally, the tongue contour which presents a high variability, due to its movements and the dynamic MRI acquisition technique as well, has been delineated by hand. From the contour extraction, area functions are obtained by, first, dividing the vocal tract shape in tubelets perpendicular to the centerline, and then applying  $\alpha$   $\beta$  transformations to recover the area<sup>21</sup>. This transformation is a power function that gives the cross-sectional area  $a(x)$  as a function of the midsagittal distance  $d(x)$ :

$$a(x) = \alpha(x)d(x)^{\beta(x)}, \quad (1)$$

where  $\alpha$  and  $\beta$  are ad hoc coefficients. Their values depend on the region of the vocal tract, and  $x$  is the distance to the glottis along the centerline. Several studies have provided different values for  $\alpha$  and  $\beta$ . We kept those given by Soquet *et al.*<sup>21</sup> for male speakers. The

determination of the centerline plays a critical role in synthesis since it influences the length of the vocal tract. A specific algorithm was designed purposely<sup>22</sup>.

As previously explained, the real-time cineMRI reconstruction enables several area functions to be recovered for a single [r] realization. Since the main aim of the paper is to investigate the possibility to simulate self-oscillations of the tongue tip in alveolar trills with lumped mass-spring models integrated into realistic shapes of the vocal tract, using area functions of alveolar trills on different contexts is not in the scope of the paper. Consequently, for the sake of brevity, we chose to use only one area function for the rest of the paper: it corresponds to the median area function extracted from the acquired sequence of images at the trill instants. The simultaneous audio recordings of the utterances allows each reconstructed MR image to be labeled with its corresponding produced phoneme. The MRI noise is removed from the recorded audio signal thanks to a source separation technique<sup>23</sup>. The images corresponding to the trill instants are then the ones that correspond to the /r/ segment. The choice of a median area function is mainly for the purpose of compensating for the potential errors due to the manual delineation, as well as the variations in the alveolar constriction region due to the tongue tip movement during the trill production. The resulting area function is shown in Fig. 2. In the simulations, it corresponds to that of the vocal tract when the tongue is at its rest position. Oscillations of the tongue tip will then modify the cross-sectional area of the lingual constriction, namely the area of the tubelet having the smallest area (at 15.9 cm in Fig. 2 (c)). Note that the back of the tongue creates a narrowing at the middle of the vocal tract. This has been previously reported as a common features of rhotic sounds, including alveolar and uvular trills<sup>24</sup>.

### III. ACOUSTIC MODEL

The simulation framework used to compute the acoustic propagation inside the vocal tract is derived from the *transmission line circuit analog model* (TLCA) approach<sup>12</sup>. 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* (RTLTA) model<sup>25,26</sup>, 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 Ref.<sup>27</sup>.



## Production of alveolar trills

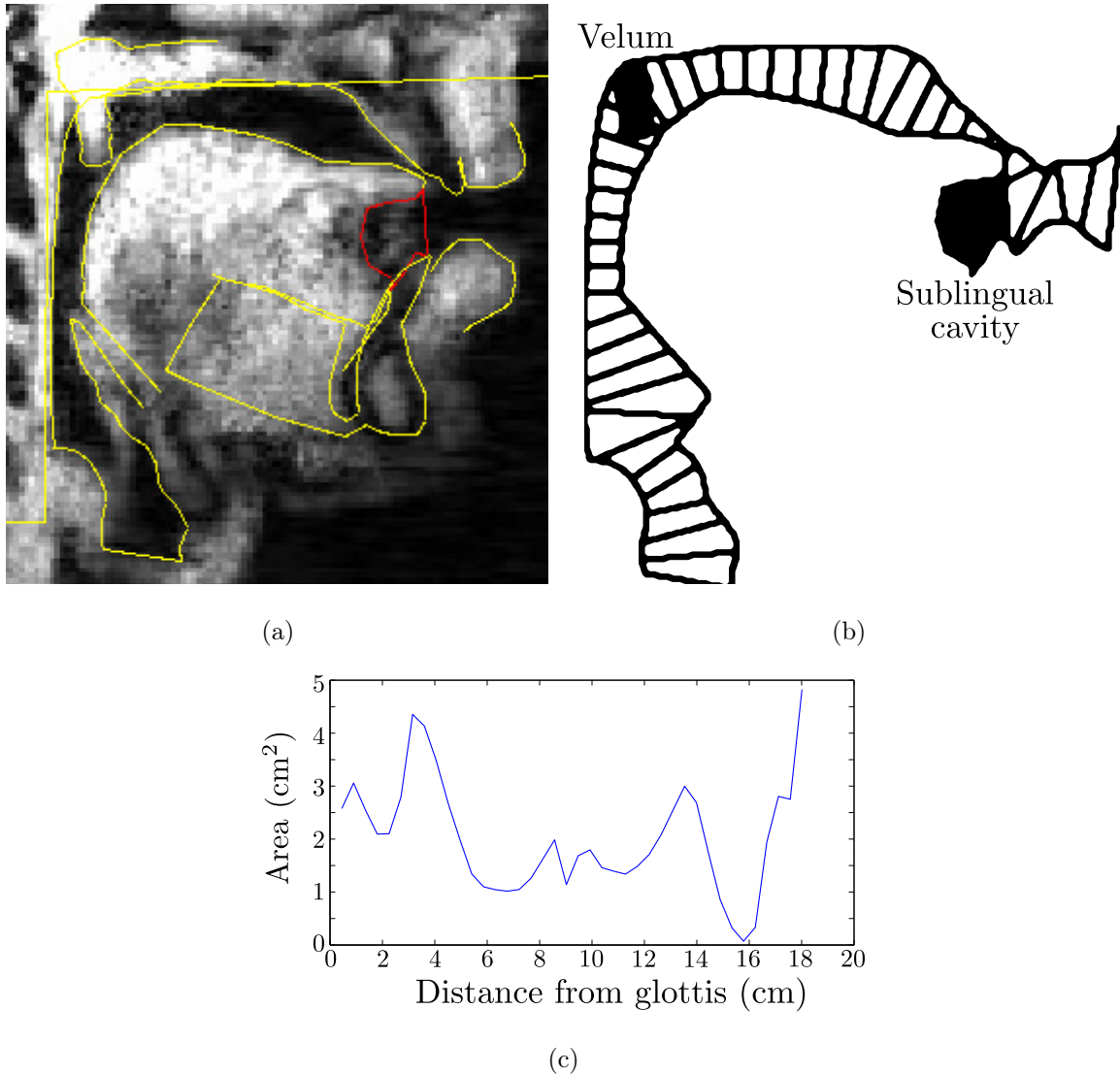


Figure 2. (a) MR image of a midsagittal view of the vocal tract during the production of an alveolar trill in the sequence /ara/. Contours of the articulators are shown in the MR image, as well as (b) the spatial sampling of the vocal tract into acoustic tubelets. (c) resulting median area function obtained from real-time cineMRI data.

The framework that is used in this paper, called ESMF<sup>13</sup>, considers recent improvements of TLCA-based techniques, such as the possibility to connect self-oscillating models of the vocal folds with a membranous glottal chink (a gap in the membranous portion of the glottis due to its partial abduction). In this paper, the acoustic model is extended to include the potential oscillations of the tongue tip.

### A. Acoustic propagation

The vocal tract is seen as a waveguide network, in which each waveguide represents a side cavity<sup>13,28</sup>. In our study, side cavities include the sublingual cavity (*cf.* Fig. 2) and the lateral channels around the sides of the tongue tip. It has been shown, following the recent ESMF paradigm<sup>13</sup>, that the wave propagation inside such networks, connected to a self-oscillating model of the vocal folds, is driven by the following system of equations

$$\mathbf{f} = \mathbf{Z}\mathbf{u}_Z + \mathbf{Q}\mathbf{u}_Q, \quad (2)$$

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 tubelet,  $\mathbf{Q}$  is a square matrix the same size as  $\mathbf{Z}$  having only one non-zero element, that is  $Q_{(1,1)} = R_b$ ,  $\mathbf{u}_Z \in \mathbb{R}^{N+1}$  is the vector containing the volume velocities inside each tubelet, and  $\mathbf{u}_Q \in \mathbb{R}^{(N+1)} = [U_1^2, U_2^2, \dots, U_{N+1}^2]^T$  is the vector containing the square power of the volume velocities. The term  $R_b$  is the Bernoulli resistance, as defined in Ref.<sup>13</sup>. Eq. 2 is computed at each time step, and the volume velocities  $\mathbf{u}$  are the solution of the equation. The acoustic pressure  $P_{Out}$  is computed as the first time derivative of the volume velocities at the lip termination, i.e.  $U_{N+1}$ , where  $N$  is the number of tubelets that models the vocal tract.

The acoustic propagation model supports the connection of self-oscillating models of the vocal folds with a membranous glottal chink<sup>13</sup>. In such case, the membranous portion of the glottis is composed of two distinct parts along the length of the vocal folds: an adducted part, modeled with a classic two-mass self-oscillating system<sup>9</sup>, and an abducted portion, i.e. the so-called *membranous glottal chink*, modeled as an acoustic branch. This allows incomplete closure of the glottis during oscillation cycles of the vocal folds due to the partial abduction of the glottis to be modeled, and has been found to be important for simulating breathy voice<sup>29</sup>, and voiced fricatives<sup>30</sup>. In this paper, the degree of abduction, denoted by  $D_{ab}$ , is defined as the ratio, in percent, between the area of the membranous chink and the area of the fully abducted glottis. Consequently,  $D_{ab} = 0\%$  when there is no membranous chink, and  $D_{ab} = 100\%$  when the glottis is fully abducted.

## B. Self-oscillating model of the tongue tip

In order to reproduce self-oscillations of the tongue tip, our model considers the lower wall of the tubelet corresponding to the lingual constriction to be a lumped mass-spring system, namely the TMM, derived from the classic two-mass mechanical system for the vocal folds by Ishizaka and Flanagan<sup>8</sup>. In this paper, the self-oscillation is used to update the cross-sectional area of the tubelet that corresponds to the lingual constriction, namely the tubelet that presents the minimal cross-section area. Consequently, only the extremity of the tongue tip is assumed to oscillate, without modifying the cross-section areas of the neighboring tubelets. In the following presentation of the model, the upstream and the downstream sections, as in Fig. 3, represent the tubelets located just downstream and upstream of the tubelet with the minimal cross-section area.

For simulating the self-oscillation of the vocal folds, the choice of a two-mass system instead of a single-mass one is motivated by the possibility to reproduce higher modes of oscillation, and especially phase shifts between the upstream and the downstream parts of the vocal folds. Indeed, at least two modes of oscillation have been extensively observed and shown to be important to consider in the simulations<sup>31</sup>. For tongue tip oscillations, this has been previously discussed by McGowan<sup>7</sup>, but there is, so far, no evidence of such a phase shift between the upstream and the downstream parts of the tongue tip during trill productions.

The TMM used for the tongue tip is then similar to the one used for vocal folds. It includes smooth contours that generates a glottal area function without discontinuities, a mobile flow separation point<sup>9,10</sup>, and the computed pressure forces consider viscous losses<sup>9</sup>.

The geometry of the tongue tip is then represented in Fig. 3. In this paper, indices  $u$  and  $d$  denote the upstream and downstream parts respectively. The position along the  $x$ -axis inside the lingual constriction is denoted by  $x_i$ , with  $x = 0, 1, 2, 3$ , corresponding to the constriction input, the position of the upstream mass, the position of the downstream mass, and the constriction output, respectively. The total length of the lingual constriction is then  $l_t = x_3 - x_0$ , and its width is denoted  $w_t$ . The equation of the tongue motion is

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

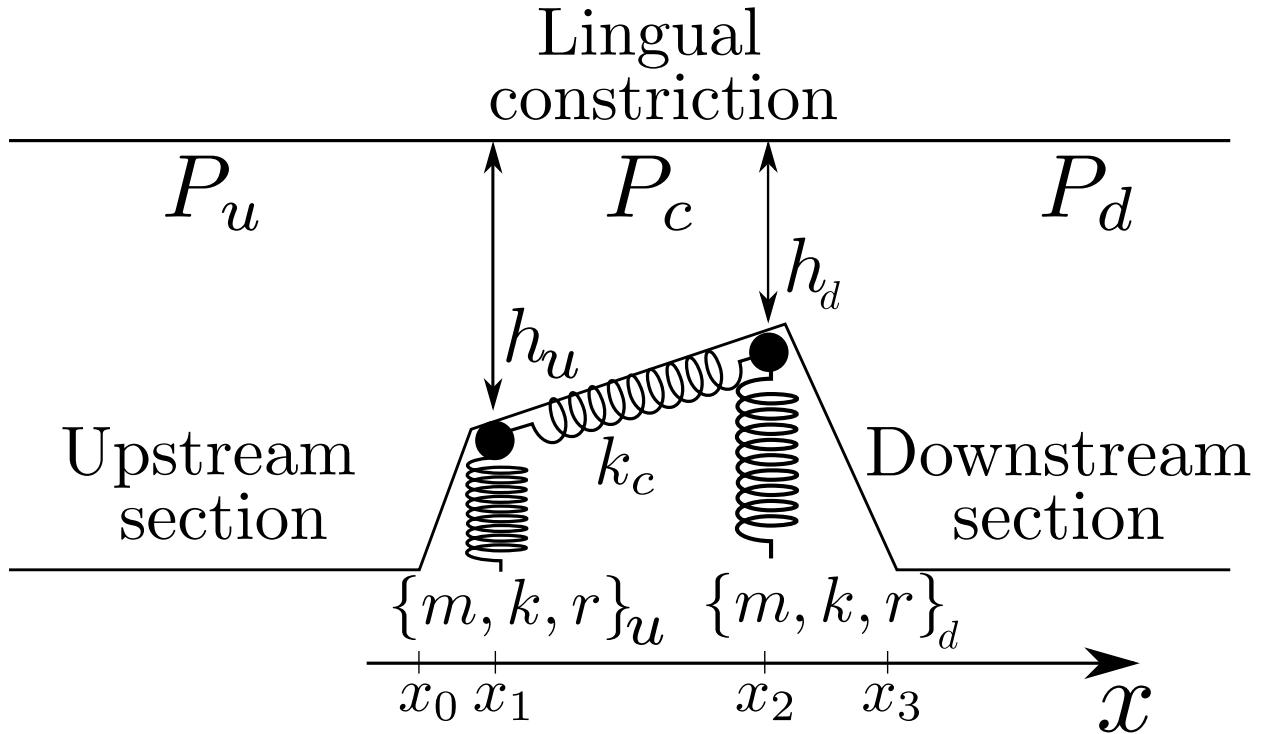


Figure 3. Two-mass model for the tongue tip.  $P_d$ ,  $P_u$ , and  $P_c$  denote the pressure downstream, upstream and inside the lingual constriction, respectively. The mass, stiffness and the damping of the tongue are denoted by the triplet  $\{m, k, r\}$ . They are coupled by a the spring  $k_c$ . The height of the lingual constriction is denoted by  $h$ .

where  $F_u$  and  $F_d$  are the pressure forces applied at the upstream and at the downstream masses, respectively. They are defined as

$$F_u(t) = w_t \int_{x_0}^{x_1} \frac{x - x_0}{x_1 - x_0} P(x, t) dx + w_t \int_{x_1}^{x_2} \frac{x - x_2}{x_1 - x_2} P(x, t) dx$$

$$F_d(t) = w_t \int_{x_1}^{x_2} \frac{x - x_1}{x_2 - x_1} P(x, t) dx + w_t \int_{x_2}^{x_3} \frac{x - x_3}{x_2 - x_3} P(x, t) dx,$$

To compute the pressure distribution  $P_c(x, t)$  inside the lingual constriction, the airflow in the oral channel is considered as incompressible. Additionally, the viscous effects, which cannot be neglected when the lingual constriction is almost closed, are taken into account via a Poiseuille corrective term. Our model considers the viscous effects as predominant over the inertial effects so that the latter are not taken into account, as suggested by Lous *et al*<sup>9</sup> and Deverge *et al.*<sup>32</sup>. Considering these assumptions, the pressure distribution  $P_c(x, t)$  is given by

$$P_c(x, t) = P_u(x, t) + Be(x, t) + Po(x, t) \quad (4)$$

where  $Be(x, t)$  and  $Po(x, t)$  are respectively the steady term of the Bernoulli equation and the Poiseuille corrective term. They are defined as:

$$\begin{aligned} Be(x, t) &= -\frac{\rho U_t^2(t)}{2w_t^2} \left[ \frac{1}{h_t^2(x, t)} - \frac{1}{h_t^2(x_0, t)} \right], \\ Po(x, t) &= -\frac{12\mu U_t(t)}{w_t} \int_{x_0}^x \frac{dx}{h_t^3(x, t)}, \end{aligned} \quad (5)$$

where  $U_t(t)$  is the instantaneous volume velocity of the airflow inside the lingual constriction.

Eq. (5) is valid for  $x$  located upstream of the separation point. Indeed, in this paper, similarly to the vocal folds model, the two-mass model of the tongue tip includes a mobile separation point  $x_s$ . For the vocal folds model, the separation point has been shown to have a significant impact on the oscillating behavior<sup>33</sup>. However, predicting its location accurately is a very hard task as it requires numerical simulations of the full three-dimensional flow<sup>33,34</sup>. For lumped models as used in this paper, simplified methods to predict the location of the separation point have been proposed<sup>9,10</sup>. They usually consider the separation point to be located at a point downstream of the minimal glottal constriction, where the cross-section area is that of the minimal glottal constriction multiplied by an ad hoc criterion, usually taken between 1.15, as in Haas *et al.*<sup>35</sup> and 1.2, as in Lous *et al.*<sup>9</sup>. This is a very strong assumption, which may lead to inaccurate predictions of the separation point location, as it has been shown by numerical studies<sup>36</sup>. Yet, the paper considers the quasi-static criterion as there is still a lack of accurate prediction methods suitable to lumped models.

Thus, considering the ad hoc criterion proposed by Lous *et al.*<sup>9</sup> of 1.2, the separation point  $x_s$  is defined such that  $h_s = h(x_s) = 1.2h_u$  if  $1.2h_u < h_d$ , and  $x_s = x_d$  otherwise. Finally,  $P_c(x, t) = P_d(t)$  if  $x > x_s$ .

The collision model uses artificial stiffening of the tongue tip, namely the stiffness parameters  $k_u$  and  $k_d$  are multiplied by a factor  $\kappa$  if  $h_u$ , or  $h_d$ , are equal or less than 0, respectively, and the damping is multiplied by a factor  $\eta$ . These factors are also ad hoc criteria, and their estimation is discussed in Sec. IV B.

#### IV. SIMULATIONS OF ALVEOLAR TRILLS

Using the area function derived from cineMRI data of a real speaker, sustained alveolar trills are simulated with the TMM. First, an alveolar trill is simulated without a lateral waveguide around the lingual constriction, such that linguopalatal contacts create a complete

occlusion of the vocal tract. Then, a solution consisting in adding a lateral branch around the lingual constriction is presented to simulate the incomplete closure of the vocal tract during linguopalatal contacts.

### A. Parameters of the tongue tip model

Table I. 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 <sup>-1</sup>	$0.02\sqrt{k.m/2}$
Coupling spring $k_c$	N/m	$k/2$
Stiffness factor when collision $\kappa$		1.5
Damping factor when collision $\eta$		1.2

The parameters chosen for the TMM are summarized in Tab. I. The width of the lingual constriction  $w_t$  and the height at rest  $h_0$  are derived from the MRI data. Since lumped mass-spring systems models consider the shape of the constriction as rectangular,  $w_t$  is estimated such that the constriction area  $a_t = h_0 w_t = \alpha_t h_0^\beta$  matches the cross-sectional area given by the power transformation of Eq. (1), hence

$$w_t = \alpha_t h_0^{\beta_t - 1}, \quad (6)$$

where  $\alpha_t$  and  $\beta_t$  are the transformation coefficients from Soquet *et al.*<sup>21</sup> in the alveolar region, namely  $\alpha_t = 1.92$  and  $\beta_t = 1.2$  when  $w_t$  and  $h_0$  are expressed in cm. 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<sup>5</sup>. Assuming the mass density of the tongue  $\rho_t$  is slightly more than water<sup>37</sup>, i.e.  $\rho_t \simeq 1.04 \times 10^6 \text{ g.m}^{-3}$ , the distance between the base and the extremity of the tongue tip  $h_{tt}$ , measured in MR images, is around 1.1 cm, and the tongue tip is a parallelepiped, each mass is

$$m = \frac{1}{2} \rho_t \times l_t \times h_{tt} \times w_t \simeq 0.26\text{g}.$$

The total mass of the system is then  $0.26 \times 2 = 0.52\text{g}$ . The ad hoc criteria  $\kappa$  and  $\eta$  of the collision model are estimated empirically. Their values impact the closed phase duration, and consequently, the open quotient of the tongue tip oscillation, namely the ratio of the duration of the open phase in comparison with the trill period. They have been estimated as the values of  $\{\kappa, \eta\}$  such that the framework simulates trills with an open quotient similar to those experimentally estimated on real speakers, namely 57.3 %, as reported by Solé<sup>5</sup>. This open quotient has been reached for  $\kappa = 1.5$  and  $\eta = 1.2$ . The collision model uses these values for the rest of the paper.

## B. Simulation of an alveolar trill with the lingual constriction modeled as a single waveguide

In this section, the lingual constriction is seen as a single waveguide. Results of the simulations are displayed in Fig. 4. It shows the oscillation of the lingual constriction height around its rest position. In this example, periodic contacts between the tongue tip and the hard palate occur. 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 is in contact with), and at its maximum when the tongue tip is far from the hard palate. The displacement of the upstream and downstream masses of the tongue tip model shows no significant phase shift, which suggests a behavior similar to a single-mass system. Simulations also reveal that the tongue tip oscillations may disturb the motion of the vocal folds. Indeed, the amplitude of the oscillation of the vocal folds increases when the tongue tip is close to, or in contact with, the hard palate. This is visible in the bottom plot in Fig. 4, representing the glottal opening  $h_g$ , where the maximal values of  $h_g$  slightly increase during two glottal pulses at  $t = 8\text{ms}$ ,  $t = 48\text{ms}$ , and  $t = 88 \text{ms}$ . For each of these time instants, the tongue tip is in contact

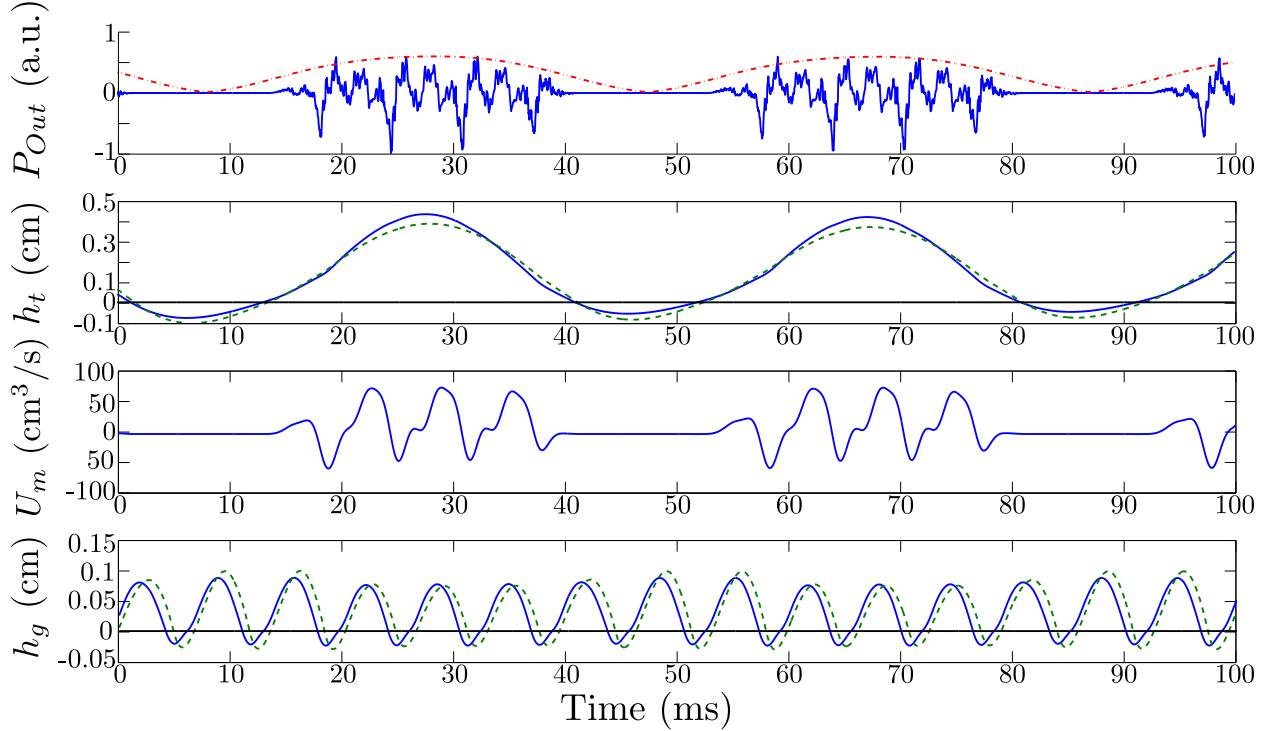


Figure 4. Simulation of an alveolar trill with linguopalatal contacts using the two-mass model. From top to bottom are plotted the acoustic pressure radiated at the lips ( $P_{Out}$ ) and its time envelope (dot-dashed line), the time evolution of the lingual constriction height  $h_t$  at the upstream (solid line) and downstream (dashed line) sections, the low frequency filtered airflow at the mouth  $U_m$ . The bottom figure plots the displacement  $h_g$  of the upstream (solid line) and downstream (dashed line) parts of the vocal folds. Positive values of  $h_t$  and  $h_g$  indicate open lingual and the glottal constriction, respectively, while null and negative values indicate closed constrictions.

with the hard palate. Basically, the amplitude of the glottal opening waveform follows the lingual constriction closure with a small time delay of a few milliseconds. More interestingly, the downstream part of the vocal folds is more impacted, as it exhibits a larger oscillation amplitude than the upstream part in the closing phase of the lingual constriction. Since the closing phase of alveolar trills is known to exhibit a rise of the supraglottal pressure<sup>5</sup>, this is certainly the main reason of this modification of the glottal opening waveform. The upstream part of the vocal folds is the most impacted since it is directly connected to supraglottal part of the vocal system.

However, the acoustic and the airflow waveforms  $P_{Out}$  and  $U_m$  show discrepancies with real alveolar trills. Indeed, the simulated waveform exhibits null values during the closed



phase, while this is never observed in real trills. This is due to the fact that the vocal tract is seen here as a single waveguide, which leads to the complete occlusion of the vocal tract during linguopalatal contacts. The next section presents a method to fix this issue by adding lateral channels that allows incomplete closures to be simulated.

### C. Simulations of the incomplete closure of the vocal tract at the linguopalatal contact instants

As discussed in Sec. II, in real trills linguopalatal contacts do not lead to the complete occlusion of the oral channel. Indeed, both the air flow at the mouth and the acoustic pressure waveform are not null. In order to simulate the incomplete closure of the vocal tract during linguopalatal contacts, we suggest to add a lateral acoustic branch to simulate the air path going along the side of the tongue. The ESMF paradigm that is used in this paper enables such situations<sup>13</sup> to be taken into account in the acoustic simulations. Due to the lack of data (cineMRI films provide only midsagittal views of the vocal tract), the area function of the lateral branch is estimated according to these assumptions: i) connections at the upstream and at the downstream parts are located respectively at 14.0 cm and 16.2 cm from the glottis, ii) the cross-section area of the lateral branch is taken at a certain ratio  $r_l$  of the vocal tract area function located between the connections, iii) the length of the lateral branch is the same as the length of the main air path. The presented simulation uses the TMM of the tongue tip with the same mechanical parameters as in Tab. I, integrated into a vocal tract modeled with a lateral branch that is connected to the main path with an area ratio  $r_l$  of 50%.

The simulation with a ratio  $r_l = 50\%$ , displayed in Fig. 5, shows that this vocal tract configuration enables linguopalatal contact with incomplete 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. These features are in agreement with what is commonly observed in real trills (see Fig. 1, for instance).

Note that the inclusion of the lateral waveguide does not significantly modify the global behavior of the tongue tip oscillations, except for a slight decrease of the trill amplitude. There is still no phase shift between the upstream and the downstream parts of the tongue tip, still suggesting a behavior similar to a single mass system. Note also that the rise of the

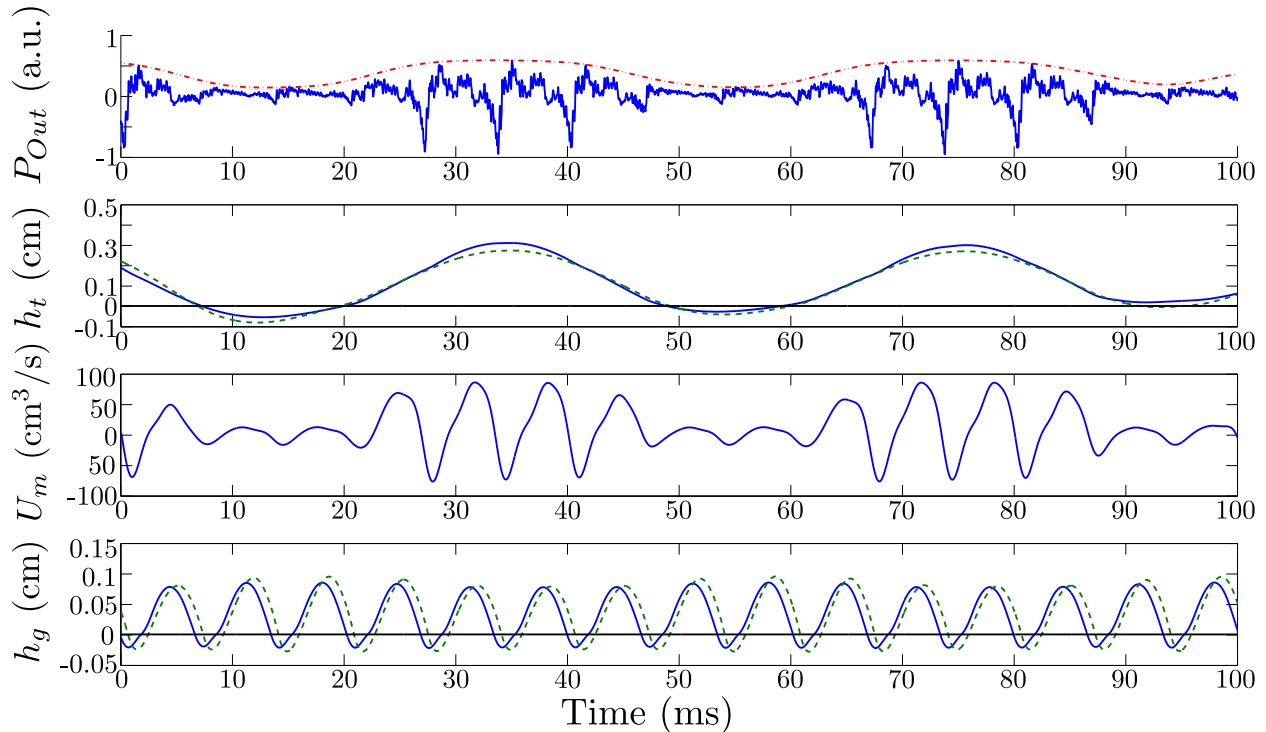


Figure 5. Results of the simulation of a sustained alveolar trill with a lateral configuration ( $r_l = 50\%$ ). From top to bottom: acoustic pressure radiated at the lips  $P_{Out}$  (solid line) and its time envelope (dash-dotted line), lingual constriction height  $h_t$  at the upstream (solid line) and at the downstream (dashed line) parts, and air flow  $U_m$  at the mouth. The bottom plot represents the displacement  $h_g$  of the vocal folds at the upstream (solid line) and at the downstream (dashed line) sections. Positive values of  $h_t$  and  $h_g$  indicate open lingual and the glottal constriction, respectively, while null and negative values indicate closed constrictions.

vocal folds opening during closed phase is still visible at  $t = 10$  ms,  $t = 50$ ms, and  $t = 90$ ms.

## V. CONDITIONS OF PRODUCTION OF ALVEOLAR TRILLS

This section investigates the influence of various configurations of the vocal tract and of the glottis on the simulation of the tongue tip oscillation. Studied parameters include the subglottal pressure, the mass of the tongue tip, the height of the lingual constriction at rest, the glottal abduction degree, and the lateralization of the air flow.

### A. Trill characteristic features

The analyzed trill characteristic features are the trill frequency  $f_t$ , the trill amplitude, denoted  $\hat{h}_t$ , and the trill contact ratio, denoted  $C_r$ . For each simulation of sustained trills, a group of 10 successive periods are considered. Fig. 6 illustrates the identification of the different features. For each oscillating cycle  $n$ ,  $A_n$  is the difference between the maximal and the next minimal lingual constriction height,  $T_n$  is the time difference between two successive maxima, and  $c_n$  is a binary number, equal to 0 if there is no linguopalatal contact during cycle  $n$ , and 1 if there is a contact.

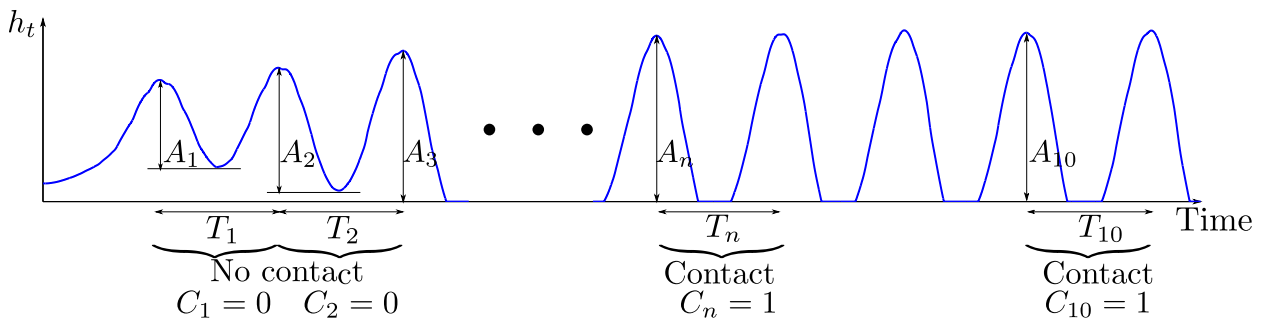


Figure 6. Example of a typical lingual constriction aperture waveform and identifying features  $A_n$ ,  $T_n$ , and  $c_n$ , for the computation of the trill characteristics.

The trill frequency is the inverse of the mean period value, namely

$$f_t = \frac{10}{\sum_{n=1}^{10} T_n}.$$

The trill amplitude  $\hat{h}_t$  is the mean value of the difference between the maximal and the next minimal lingual constriction height in the group of 10 successive oscillatory cycles, hence

$$\hat{h}_t = \frac{1}{10} \sum_{n=1}^{10} A_n.$$

The trill contact ratio  $C_r$  is the percentage of oscillatory cycles where there is a linguopalatal contact among 10 successive periods:

$$C_r(\%) = 100 \times \frac{1}{10} \sum_{n=1}^{10} c_n.$$

## B. Effect of the mass and the subglottal pressure

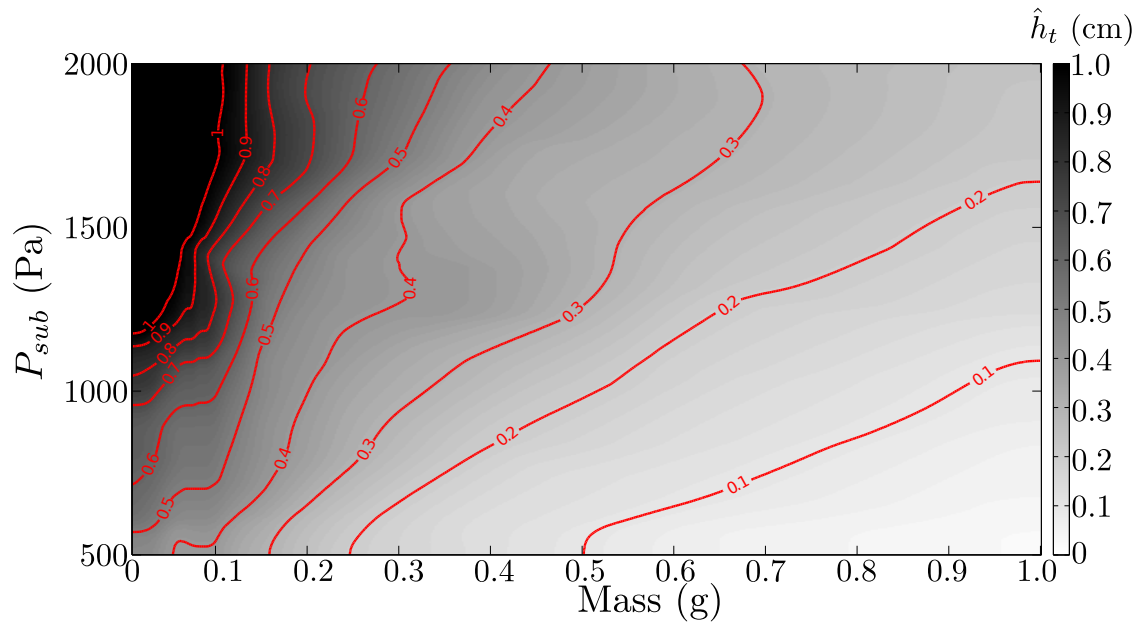
For these simulations, the glottis is entirely adducted, i.e.  $D_{ab} = 0\%$ , and there is no lateralization, i.e.  $r_l = 0\%$ . The other parameters, namely the rest position  $h_0$ , the length and the width of the lingual constriction, and the contact model parameters  $\kappa$  and  $\eta$ , are set to the values shown in Tab. I. The masses  $m_1$  and  $m_2$  vary from 10 mg to 1 g, and the subglottal pressure from 500 to 2000 Pa. For each simulation,  $m_1 = m_2$ .

Fig. 7 shows the trill contact ratio and the trill amplitude as a function of the values of each mass modeling the tongue tip and the subglottal pressure. Not surprisingly, the trill amplitude is the largest for small values of the tongue mass and for large values of the subglottal pressure  $P_{sub}$ . For mass values smaller than 0.1 g, the amplitude of the tongue tip oscillation may be larger than 1 cm, which seems unrealistically high, since it is out of the range of reported data on real speakers<sup>17,18,38</sup>. This leads to area variations of approximately 1.4 cm<sup>2</sup>, according to Eq. (1), and is much larger than data provided by McGowan<sup>7</sup>, who reported maximal constriction area between 0.27 and 0.67 cm<sup>2</sup>. These areas correspond to maximal constriction heights of 0.44 cm and 0.21 cm, respectively. In the range of subglottal pressure values usually observed in natural speech<sup>39</sup>, i.e. under 1000 Pa, the trill amplitude values that lies in this range (0.21-0.44 cm) are reached for mass values between 0.2 g and 0.4 g. In this range, linguopalatal contacts occur almost systematically ( $C_r > 80\%$ ). Note that the contact ratio  $C_r$  follows the same evolution as the trill amplitude. For the rest of the paper, both masses modeling the tongue will be set to 0.26 g, as in Sec. IV, since it roughly corresponds to anatomical data, and additionally lies in the range of mass values that simulate realistic oscillations of the tongue tip.

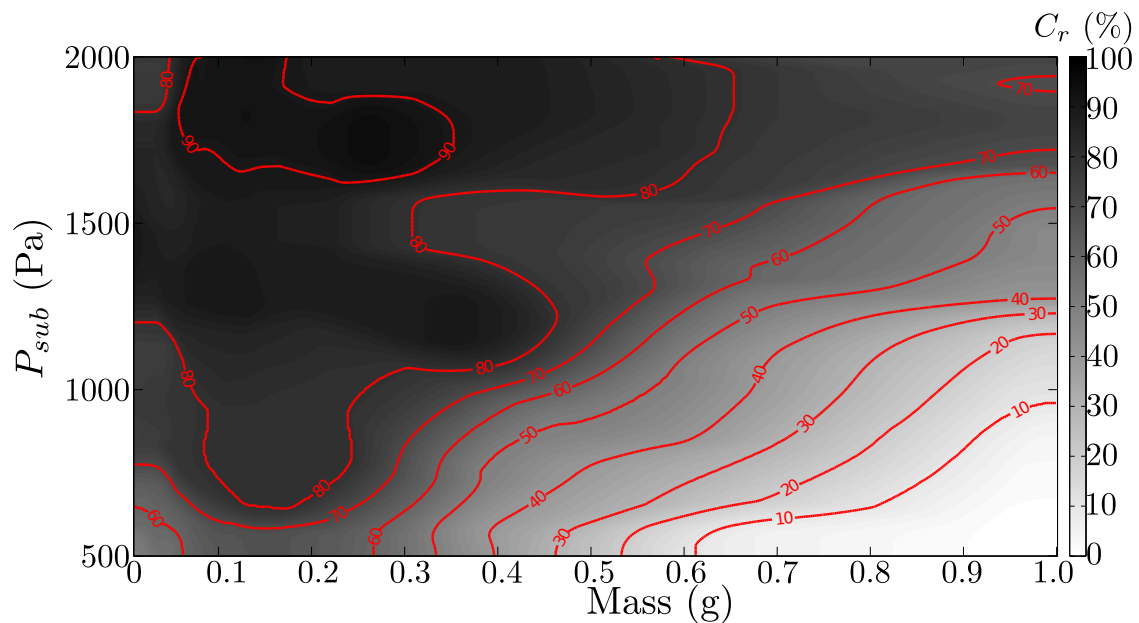
## C. Effect of the tongue initial position

For these simulations, the glottis is entirely adducted, i.e.  $D_{ab} = 0\%$ , and there is no lateralization, i.e.  $r_l = 0\%$ . The length and the width of the lingual constriction are set to the values shown in Tab. I. The masses  $m_1$  and  $m_2$  are set to 0.26 g. The rest position varies from 0 to 0.4 cm and the subglottal pressure from 500 to 2000 Pa.

The tongue initial position  $h_0$  is set to different values to study its impact on the self-oscillating motion of the tongue tip. Fig. 8 shows the trill frequency  $f_t$ , the trill amplitude



(a)



(b)

Figure 7. (a) Trill amplitude and (b) trill contact ratio, as a function of the mass  $m_1 = m_2$  and the subglottal pressure  $P_{sub}$ . Solid lines represent the contour lines.

$\hat{h}_t$ , and the trill contact ratio  $C_r$  as a function of the subglottal pressure  $P_{sub}$  and the initial tongue position  $h_0$ . It clearly shows that the initial tongue position has a significant influence on the production of alveolar trills. For instance, for a given  $P_{sub}$ , the trill amplitude as a

function of  $h_0$  exhibits a local maximum, i.e. when  $h_0$  is between 0.5 and 1 mm. This implies that the tongue should not be in contact with the hard palate at the beginning of the trill. Results also show that when  $h_0$  is larger than approximately 1.25 mm, linguopalatal contacts occur ( $C_r \geq 10\%$ ) only when  $P_{sub}$  is larger than a certain threshold, between 1400 and 1500 Pa, for  $h_0 \leq 0.25$ cm, and larger than 1500 Pa for  $h_0 \geq 0.25$ cm. This threshold value increases as  $h_0$  increases: it is larger than 2000 Pa for  $h_0 \geq 0.3$ cm. These observations show that the alveolar trills are favored when the tongue tip is very close to the hard palate, i.e. less than 1.25 mm, but not in contact. Consequently, producing alveolar trills requires a very fine adjustment of the tongue tip position in order to generate the self-oscillation; which confirms the fact that alveolar trills are difficult to produce. It is also worth noting that larger lingual constriction heights at rest tend to increase the trill frequency. The trill frequency is the most sensitive to the rest position when  $h_0$  is smaller than the local maximum, i.e. when  $h_0 < 0.5$  mm.

#### D. Effect of the incomplete closure of the vocal tract

For these simulations, the glottis is entirely adducted, i.e.  $D_{ab} = 0\%$ , the rest position, the length and the width of the lingual constriction are set to the values shown in Tab. I. The masses  $m_1$  and  $m_2$  are set to 0.26 g. The lateral area ratio  $r_l$  varies from 0 to 50%, and the subglottal pressure from 500 to 2000 Pa.

The incomplete closure of the vocal tract is studied via the lateralization ratio  $r_l$ . Fig. 9 shows that this has no significant effect on the trill characteristics: the trill frequency is barely lowered by 1 Hz when there is an incomplete closure of the vocal tract, and, for a given  $P_{sub}$ , the trill amplitude and the trill contact ratio exhibit a slight tendency to decrease as the incomplete closure increases (larger  $r_l$ ). However, although the incomplete closure of the vocal tract during linguopalatal contacts has a very small impact on the mechanical behavior of the tongue tip, it has been shown in Sec. IV C that the acoustic impact is not negligible. Consequently, when dealing with acoustic synthesis of alveolar trills using physical models, one should consider the lateralization of the airflow.

### E. Effect of the glottal abduction

For these simulations, there is no lateralization, i.e.  $r_l = 0\%$ , the rest position, the length and the width of the lingual constriction are set to the values shown in Tab. I. The masses  $m_1$  and  $m_2$  are set to 0.26 g. The degree of glottal abduction  $D_{ab}$  varies from 0 to 100%, and the subglottal pressure from 500 to 2000 Pa.

Fig. 10 shows that the presence of the partial abduction has a significant impact on the trill characteristics. Indeed, it has for main effect, for a given  $P_{sub}$ , to increase the trill frequency, the trill amplitude, and also the trill contact ratio. Indeed, when the degree of abduction  $D_{ab}$  is greater than approximately 30%, the trill amplitude is smaller than 2 mm, and the trill contact ratio shows that there is almost no linguopalatal contacts. There is a transition zone, when  $D_{ab}$  is between 20 and 40%, where the trill contact ratio is very sensitive to small perturbations of the glottal abduction. Before this transition, for  $D_{ab} < 25\%$ , the trill contact ratio is 100%, and it vanishes for  $D_{ab} > 40\%$ , except when  $P_{sub}$  is larger than 1700 pa, approximately.

Note that the increase of the trill frequency for voiceless trills is in agreement with the experimental observations by Solé<sup>5</sup>, who reported higher trill frequencies for voiceless trills (mean 29.3 Hz) than for voiced trills (mean 28.1 Hz). Another explanation of an higher trill frequency of voiceless trills could be the fact that they require higher subglottal pressures. Indeed, for a given abduction degree, increasing the subglottal pressure increases the trill frequency, as shown in Fig. 10 (a).

## VI. CONCLUSION AND FURTHER WORKS

This paper has discussed the possibility to model the self-oscillation of the tongue tip in alveolar trills via classic lumped mass-spring systems. Based on a model similar to one used for vocal folds, the two-mass model<sup>9</sup>, it has been shown that a periodic oscillation of the tongue tip could be reproduced to simulate the production of alveolar trills. In comparison with previous works<sup>7</sup>, the simulation framework presented in this paper can support physiologically realistic area functions of the vocal tract, and also the connection with a self-oscillating model of the vocal folds with a membranous glottal leakage to simulate voiced and voiceless alveolar trills. It can also account for the incomplete closure of the

vocal tract during linguopalatal contacts due to the lateral leakage of the airflow around the tongue.

The two-mass model of the tongue tip oscillates at frequencies similar to those observed in natural speech<sup>5,7</sup>, i.e. in the range 25-30 Hz. In our simulations, both masses do not show significant phase shifts, which suggests that, unlike the vocal folds, its behavior is very similar to the single-mass model.

The simulation framework presented in this paper also deals with the commonly observed case of the incomplete closure of the vocal tract during linguopalatal contacts. It is taken into account thanks to the connection of lateral branches around the tongue. Although our simulations show that the connection of lateral branches does not significantly modify the oscillation characteristics of the tongue tip, it has a strong acoustic impact: the intraoral airflow and the acoustic pressure waveforms are not completely null in the closed phase of the trill, as it has been observed in natural speech in all of the previous studies<sup>5,7</sup>. Consequently, the incomplete closure of the vocal tract should be taken into account when dealing with the acoustic synthesis of alveolar trills using physical models.

Additionally, our simulations have revealed the fact that there is an initial position of the tongue tip that favors the production of alveolar trills. Indeed, for a given subglottal pressure, the trill amplitude as a function of the rest position of the tongue tip admits a local maximum, located between 0.5 and 1 mm off the hard palate. When this tongue-palate distance at rest is greater than 1 mm, the trill amplitude significantly decreases, making the production of alveolar trills more difficult. This might be one of the reason why consonant trills are hard-to-produce sounds in human spoken languages.

Finally, it has been shown that the amplitude of the tongue oscillation is significantly reduced in voiceless configurations, i.e. when the vocal folds are partially, or completely abducted. This is especially true when the degree of glottal abduction exceeds a critical value, between 30 and 40% of the total glottal abduction, depending on the value of the subglottal pressure. This confirms the fact that voiceless trills are more difficult to pronounce than voiced trills, and this may also explain why voiceless trills are rare among human languages.

This first attempt to model the tongue tip oscillations with classic lumped mass-spring systems opens new ways of investigations of the production of trills. Indeed, one could imagine similar models applied to other trills, such as the lips for the bilabial trills, or



at the uvula for the uvular trills. The possibility to support realistic geometries of the vocal tract derived from cineMRI acquisitions of real speakers is a useful tool to study the impact of various articulatory configurations on the production of trills, and consequently to investigate the reasons of the difficulty encountered by numerous speakers to produce trills. Additional data, considering phonological contexts, are planned to be acquired in the near future to complete this investigation. Finally, the paper raises several questions about the dynamics of the tongue tip oscillation and about the aerodynamic behavior of the flow in the lingual constriction during alveolar trills that remain unsolved, mainly due to the lack of full numerical simulations, and/or fine experimental observations. Thus, this paper may constitute a base for such investigations in order to eventually address these important questions for the full understanding of the alveolar trills dynamics.

## ACKNOWLEDGEMENTS

The work belongs to the project ArtSpeech, with the financial support of the French ANR (*Agence Nationale de la Recherche*). The authors are grateful to the laboratory IADI (*Imagerie Adaptative Diagnostique et Interventionnelle*) and the CIC-IT (*Centre d'Intervention Clinique - Innovations Technologiques*) of Nancy for the MRI acquisitions.

## REFERENCES

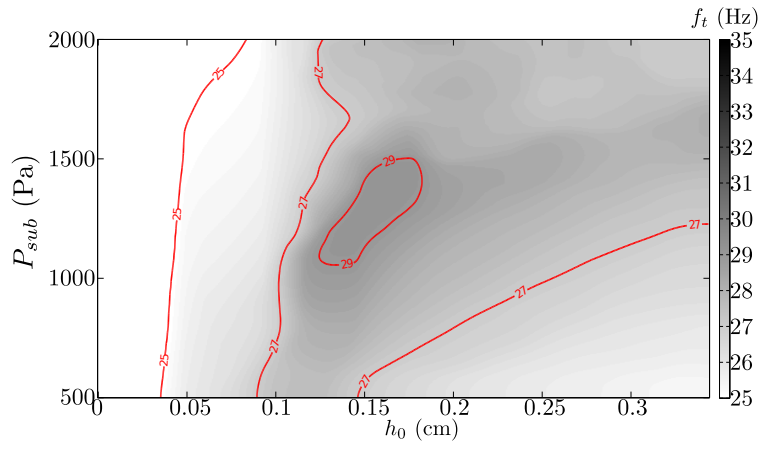
- <sup>1</sup>I. Maddieson, S. F. Disner, *Patterns of sounds*, Cambridge university press, 1984.
- <sup>2</sup>B. C. Jimenez, Acquisition of spanish consonants in children aged 3–5 years, 7 months, *Language, Speech, and Hearing Services in Schools* 18(4) (1987) 357–363.
- <sup>3</sup>M.-J. Solé, Production requirements of apical trills and assimilatory behavior, in: *Proceedings of the XIVth International Congress of Phonetic Sciences*, Vol. 1, 1999, pp. 487–490.
- <sup>4</sup>E. Mendoza, G. Carballo, Acoustic characteristics of trill productions by groups of spanish children, *clinical linguistics & phonetics* 14 (8) (2000) 587–601.
- <sup>5</sup>M.-J. Solé, Aerodynamic characteristics of trills and phonological patterning, *J. of Phon.* 30 (2002) 655–688.
- <sup>6</sup>D. Recasens, M. D. Pallarès, A study of /r/ and /r̄/ in the light of the "DAC" coarticulation model, *Journal of Phonetics* 27(2) (1999) 143–169.

- <sup>7</sup>R. S. McGowan, Tongue-tip trills and vocal-tract wall compliance, *J. Acoust. Soc. Am.* 91(5) (1992) 2903–2910.
- <sup>8</sup>K. Ishizaka, J. L. Flanagan, Synthesis of voiced sounds from a two-mass model of the vocal cords, *Bell Syst. Tech. J.* 51(6) (1972) 1233–1268.
- <sup>9</sup>N. J. C. Lous, G. C. J. Hofmans, R. N. J. Veldhuis, A. Hirschberg, A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design, *Acta Acustica* 84 (1998) 1135–1150.
- <sup>10</sup>X. Pelorson, A. Hirschberg, R. R. van Hassel, A. P. J. Wijnands, Y. Auregan, 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) (1994) 3416–3431.
- <sup>11</sup>P. Birkholz, B. Kröger, P. Birkholz, A survey of self-oscillating lumped-element models of the vocal folds, *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung* (2011) 47–58.
- <sup>12</sup>S. Maeda, A digital simulation method of the vocal-tract system, *Speech Communication* 1 (1982) 199–229.
- <sup>13</sup>B. Elie, Y. Laprie, 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 (2016) 85–96.
- <sup>14</sup>B. Elie, Y. Laprie, P.-A. Vuissoz, F. Odille, High spatiotemporal cineMRI films using compressed sensing for acquiring articulatory data, in: *Eusipco, Budapest, 2016*, pp. 1353–1357.
- <sup>15</sup>A. M. Lewis, Coarticulatory effects on spanish trill production, in: *Proceedings of the 2003 Texas Linguistics Society Conference, Vol. 116, 2004*, p. 127.
- <sup>16</sup>D. Recasens, Coarticulation in catalan dark [l] and the alveolar trill: general implications for sound change, *Language and speech* 56 (1) (2013) 45–68.
- <sup>17</sup>P. W. Schönle, K. Gräbe, P. Wenig, J. Höhne, J. Schrader, B. Conrad, Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract, *Brain and Language* 31 (1) (1987) 26–35.
- <sup>18</sup>P. Howson, A. Kochetov, P. van Lieshout, Examination of the grooving patterns of the czech trill-fricative, *Journal of Phonetics* 49 (2015) 117–129.
- <sup>19</sup>M. Proctor, Towards a gestural characterization of liquids: Evidence from spanish and russian, *Laboratory Phonology* 2 (2) (2011) 451–485.

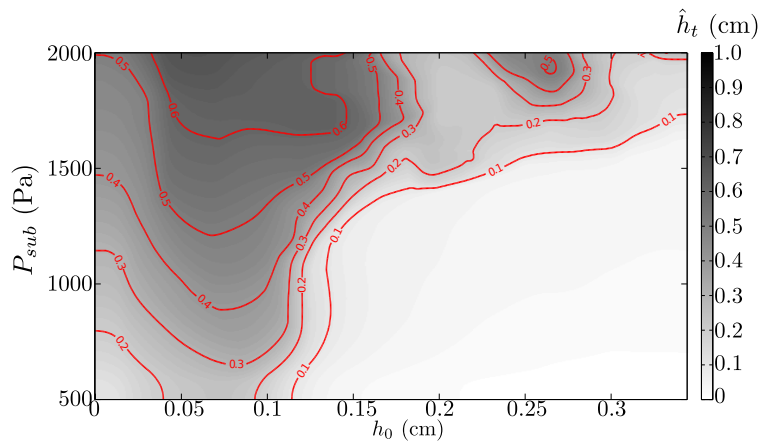
- <sup>20</sup>J. F. Jallon, F. Berthommier, A semi-automatic method for extracting vocal-tract movements from x-ray films, *Speech Communication* 51 (2) (2009) 97–115.
- <sup>21</sup>A. Soquet, V. Lecuit, T. Metens, D. Demolin, Mid-sagittal cut to area function transformations: Direct measurements of mid-sagittal distance and area with MRI, *Speech Communication* 36(3) (2002) 169–180.
- <sup>22</sup>Y. Laprie, M. Loosvelt, S. Maeda, E. Sock, F. Hirsch, Articulatory copy synthesis from cine X-ray films, in: *Interspeech 2013 (14th Annual Conference of the International Speech Communication Association)*, Lyon, France, 2013, pp. 1–5.
- <sup>23</sup>A. Ozerov, E. Vincent, F. Bimbot, A general flexible framework for the handling of prior information in audio source separation, *IEEE Transactions on Audio, Speech, and Language Processing* 20(4) (2012) 1118–1133.
- <sup>24</sup>S. E. Boyce, S. M. Hamilton, A. Rivera-Campos, Acquiring rhoticity across languages: An ultrasound study of differentiating tongue movements., *Clinical Linguistics & Phonetics* 30 (3–5) (2016) 174–201.
- <sup>25</sup>J. L. Kelly, C. C. Lochbaum, Speech synthesis, in: *Proceedings of the Fourth International Congress on Acoustics*, 1962, pp. 1–4.
- <sup>26</sup>B. H. Story, Phrase-level speech simulation with an airway modulation model of speech production, *Computer Speech & Language* 27(4) (2013) 989–1010.
- <sup>27</sup>B. J. Kröger, P. Birkholz, Articulatory synthesis of speech and singing: State of the art and suggestions for future research, in: *Multimodal Signals: Cognitive and Algorithmic Issues*, Springer, 2009, pp. 306–319.
- <sup>28</sup>P. Mokhtari, H. Takemoto, T. Kitamura, Single-matrix formulation of a time domain acoustic model of the vocal tract with side branches, *Speech Communication* 50(3) (2008) 179 – 190.
- <sup>29</sup>M. Zañartu, G. E. Galindo, B. D. Erath, S. D. Peterson, G. R. Wodicka, R. E. Hillman, Modeling the effects of a posterior glottal opening on vocal fold dynamics with implications for vocal hyperfunction, *J. Acoust. Soc. Am.* 136(6) (2014) 3262–3271.
- <sup>30</sup>B. Elie, Y. Laprie, A glottal chink model for the synthesis of voiced fricatives, in: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 5240–5244.
- <sup>31</sup>I. R. Titze, The physics of small-amplitude oscillation of the vocal folds, *J. Acoust. Soc. Am.* 83(4) (1988) 1536–1552.

- <sup>32</sup>M. Deverge, X. Pelorson, C. Vilain, P.-Y. Lagrée, F. Chentouf, J. Willems, A. Hirschberg, Influence of collision on the flow through in-vitro rigid models of the vocal folds, *J. Acoust. Soc. Am.* 114(6) (2003) 3354–3362.
- <sup>33</sup>Z. Zhang, Influence of flow separation location on phonation onset, *J. Acoust. Soc. Am.* 124 (3) (2008) 1689–1694.
- <sup>34</sup>W. Zhao, C. Zhang, S. H. Frankel, L. Mongeau, Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms, *J. Acoust. Soc. Am.* 112 (5) (2002) 2134–2146.
- <sup>35</sup>J. Haas, P. Luizard, X. Pelorson, J. C. Lucero, Study of the effect of a moderate asymmetry on a replica of the vocal folds, *Acta Acustica united with Acustica* 102 (2) (2016) 230–239.
- <sup>36</sup>D. Sciamarella, P. Le Quéré, Solving for unsteady airflow in a glottal model with immersed moving boundaries, *European Journal of Mechanics-B/Fluids* 27 (1) (2008) 42–53.
- <sup>37</sup>F. A. Duck, *Physical properties of tissues: a comprehensive reference book*, Academic press, 2013.
- <sup>38</sup>Y.-C. Kim, J. Kim, M. Proctor, A. Toutios, K. Nayak, S. Lee, S. S. Narayanan, Toward automatic vocal tract area function estimation from accelerated three-dimensional magnetic resonance imaging, in: *Speech Production in Automatic Speech Recognition*, Citeseer, 2013, pp. 1–4.
- <sup>39</sup>A. Giovanni, D. Demolin, C. Heim, J.-M. Triglia, Estimated subglottic pressure in normal and dysphonic subjects, *Annals of Otology, Rhinology & Laryngology* 109(5) (2000) 500–504.

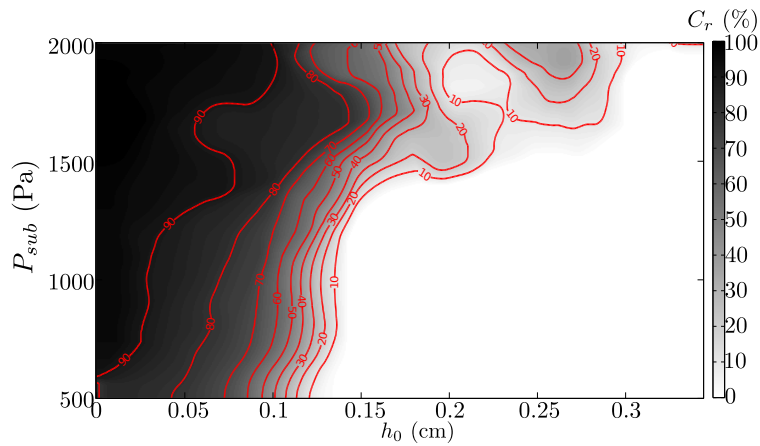
Production of alveolar trills



(a)



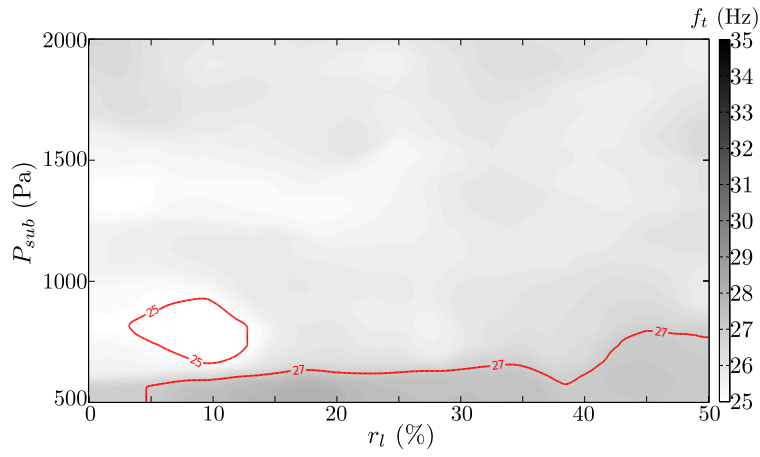
(b)



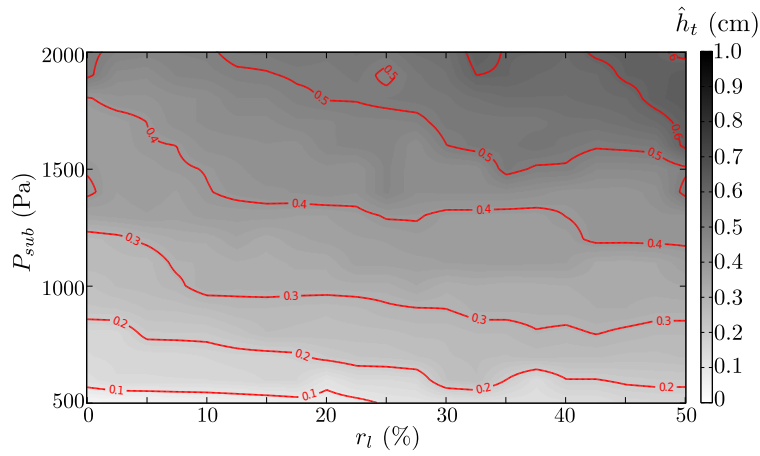
(c)

Figure 8. (a) Trill frequency  $f_t$ , (b) trill amplitude  $\hat{h}_t$ , and (c) trill contact ratio  $C_r$ , as a function of the subglottal pressure  $P_{sub}$  and the rest position of the tongue  $h_0$ . Solid lines represent the contour lines.

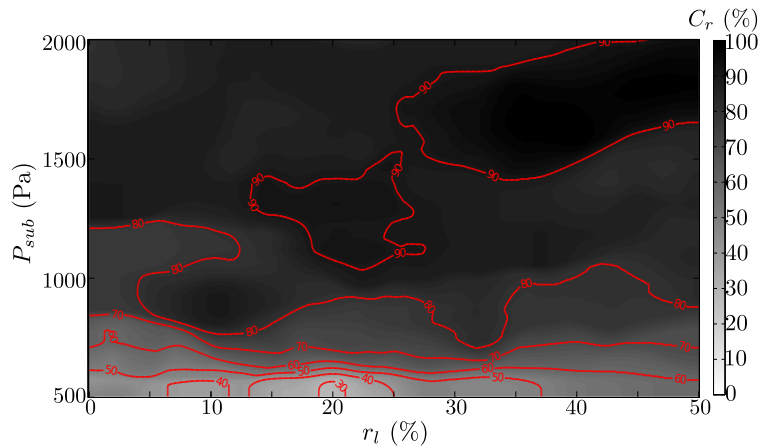
Production of alveolar trills



(a)



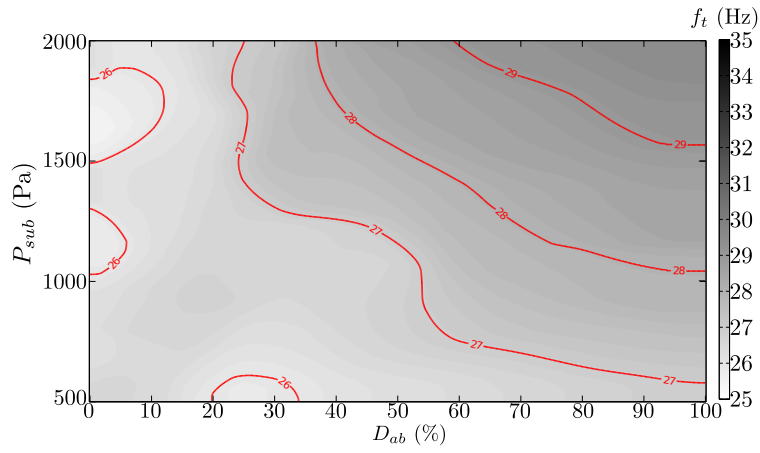
(b)



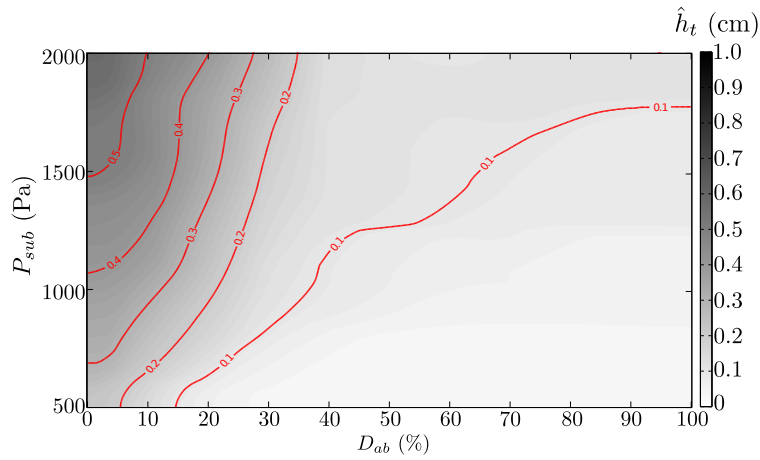
(c)

Figure 9. (a) Trill frequency  $f_t$ , (b) trill amplitude  $\hat{h}_t$ , and (c) trill contact ratio  $C_r$ , as a function of the subglottal pressure  $P_{sub}$  and the lateralization ratio  $r_l$ . Solid lines represent the contour lines.

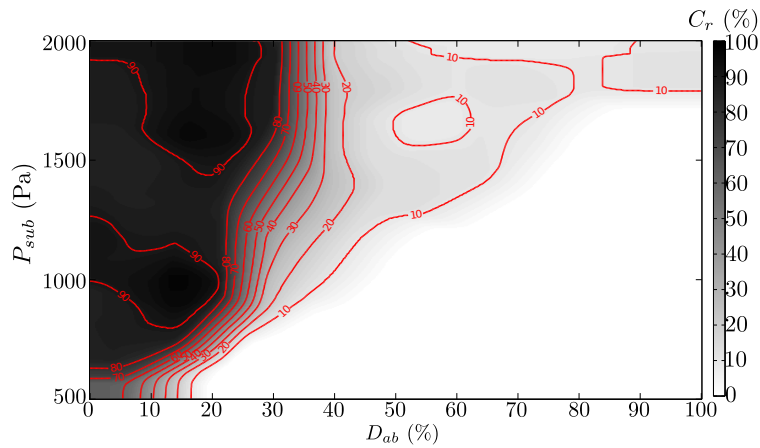
Production of alveolar trills



(a)



(b)



(c)

Figure 10. (a) Trill frequency  $f_t$ , (b) trill amplitude  $\hat{h}_t$ , and (c) trill contact ratio  $C_r$ , as a function of the subglottal pressure  $P_{sub}$  and the degree of glottal abduction  $D_{ab}$ . Solid lines represent the contour lines.