# Acoustic impact of the gradual glottal abduction degree on the production of fricatives: A numerical study

Benjamin Elie[1, a)] and Yves Laprie[1]

*LORIA, INRIA/CNRS/Université de Lorraine, Vandoeuvre-les-Nancy,
France*

(Dated: August 16, 2017)

The paper presents a numerical study about the acoustic impact of the gradual glottal opening on the production of fricatives. Sustained fricatives are simulated by using classic lumped circuit element methods to compute the propagation of the acoustic wave along the vocal tract. A recent glottis model is connected to the wave solver to simulate a partial abduction of the vocal folds during their self-oscillating cycles. Area functions of fricatives at the three places of articulation of French have been extracted from static MRI acquisitions. Simulations highlight the existence of three distinct regimes, named $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$, depending on the degree of abduction of the glottis. They are characterized by the frication noise level: $\mathcal{A}$ exhibits a low frication noise level, $\mathcal{B}$, which is a transitional unstable regime, is a mixed noise/voice signal, and $\mathcal{C}$ contains only frication noise. They have significant impacts on the first spectral moments. Simulations show that their boundaries depend on articulatory and glottal configurations. The transition regime $\mathcal{B}$ is shown to be unstable: it requires very specific configurations in comparison with other regimes, and acoustic features are very sensitive to small perturbations of the glottal configuration abduction in this regime.

PACS numbers: 43.70.Bk
Keywords: Speech production; Phonetics; Fricative; Glottal gap

[a)] Electronic mail: benjamin.elie@loria.fr

# I. INTRODUCTION

Fricatives are a class of consonants that are produced by creating a supraglottal constriction in the vocal tract so that a turbulent airflow is generated, usually downstream of the constriction. This results in the production of the so-called frication noise, which is the main characteristic of the fricative consonants. Voiceless fricatives are produced by adjusting the glottal opening area such that it is significantly greater than the area of the supraglottal constriction[1]. In that case, the glottis is completely abducted, and the generated sound contains only frication noise. In voiced fricatives, the glottis is adjusted so that both the frication noise and the voiced source contributions are mixed in the produced speech. The aeroacoustic conditions required to produce voiced fricatives are then very specific.

Studies about fricatives have focused on the spectral characteristics of the produced sound[2], the frication noise source[3], the geometry of the supraglottal constriction[4], and the vocal tract configuration downstream of the supraglottal constriction[5], especially the potential obstacles (teeth, lips, etc) encountered by the airflow[6]. These studies have contributed to a better understanding of the specific aeroacoustic conditions that are required to produce fricatives. It has been shown that the frication noise is generated when the airflow becomes turbulent, namely when the Reynolds number is sufficiently high[1]. High Reynolds numbers occur when the cross-sectional area of the supraglottal constriction is small and/or when the low-frequency component of the acoustic volume velocity through the constriction is large. The latter condition implies that the glottal opening area should be sufficiently large. For voiced fricatives, in addition to these conditions for the generation of frication noise, other conditions at the vicinity of the glottis are required to produce the voicing: the vocal folds should not be completely abducted, and the transglottal pressure drop must be large enough to guarantee self-oscillations[7]. This implies subtle adjustments of the geometry of both the supraglottal constriction and the configuration at the glottis.

Most studies about the perceptual distinction between voiced and voiceless fricatives at the same place of articulation have focused on the duration of the voiced and the voiceless parts of the considered fricative[8–11]. The consonant voicing is then considered as binary. However, the amount of energy of the voicing component over that of the frication noise component is likely to vary continuously during the fricative segment. Indeed, one may assume that the motions of the articulators and the abduction movements of the vocal folds, which are relatively slow in comparison with the oscillations of the vocal folds, gradually modify both the amplitude of the frication noise source and the amplitude of the voiced source. The acoustic impact of the vocal tract configuration on frication noise sources has been widely studied[5,6], but little attention has been paid to the acoustic impact of the configuration at the glottis. In a previous study[12], it has been experimentally shown that glottal pulses characteristics are subjected to large variations at vowel-consonant transitions. More interestingly, the presented results suggest that the glottal abduction movement starts before the vowel-consonant transition, and that the glottal adduction movement ends after the consonant-vowel transition. This implies a gradual abduction of the glottis, along with sustained vocal folds oscillations, by creating a posterior membranous gap, sometimes referred to the *linked leak*[13]. Recently, the existence of such a glottal gap due to the partial abduction of the vocal folds has been proposed to be an important feature in the production of voiced fricatives[14]. Since the membranous opening directly acts on the acoustic volume velocity, it also modifies the amplitude of the frication noise source. Thus, bad coordination between the glottal opening and the geometry of the constriction may result in an uncontrolled frication noise, and is likely to produce voiced fricatives that are perceived as voiceless because of a too large amount of frication noise.

The paper develops this idea: it uses the partial abduction model introduced in our recent papers[14,15] for a numerical study about the acoustic impact of the configuration at the glottis on the production of the voiced fricatives as a function of the geometry of the supraglottal constriction and the place of articulation. It focuses on the influence of the membranous glottal opening, which is related to the degree of abduction of the glottis, on some acoustic features of the speech signal, such as the spectral centroid, the spectral spread, and the voicing quotient. The aim is to define the boundaries of the different regimes of production of fricatives in a phonatory-articulatory space spanned by the glottal abduction degree, the position, and the geometry of the supraglottal constriction. After presenting the acoustic model in Sec. II, the paper details the configurations of the vocal tract used for the numerical simulations in Sec. III as well as the acoustic features that are investigated. Results of the simulations are presented in Sec. IV, which discusses the impact of the glottal abduction degree on the first spectral moments of the simulated fricatives, and on the amount of the generated frication noise. Finally, the boundaries of the different regimes of fricative production and their impact on the phonetic strategies to contrast voiced and voiceless fricatives are discussed and compared with data from a real speaker in Sec. V.

## II. ACOUSTIC MODEL

The simulation framework used to compute the acoustic propagation inside the vocal tract is derived from the *transmission line circuit analog* (TLCA) approach[16]. 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* (RTLA) model[17,18], 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.[19]. The framework that is used in this paper considers recent improvements of TLCA-based techniques[15,20], such as the possibility of connecting self-oscillating models of the vocal folds with incomplete closure due to the presence of a membranous glottal gap.

### A. Acoustic propagation

Considering a general case, the vocal tract is seen as a waveguide network, where each waveguide represents a side cavity (the oral tract, the nasal tract, the piriform fossae, etc). TLCA-based techniques have shown[16,20] that the wave propagation inside such networks is driven by a set of linear equations. In a matrix form, it writes

$$\mathbf{f} = \mathbf{Z}\mathbf{u}, \tag{1}$$

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 tubelet, and $\mathbf{u} \in \mathbb{R}^{N+1}$ is the vector containing the volume velocities inside each tubelet.

When dealing with self-oscillating models of the vocal folds, a quadratic term accounting for the pressure drop inside the glottal constriction, due to the Bernoulli resistance, should be added to the first line of the system[15]:

$$\mathbf{f} = \mathbf{Z}\mathbf{u}_Z + \mathbf{Q}\mathbf{u}_Q, \tag{2}$$

where $\mathbf{Q}$ is a square matrix the same size as $\mathbf{Z}$ having only one non-zero element, that is $Q_{(1,1)} = R_b$, and $\mathbf{u}_Q \in \mathbb{R}^{(N+1)} = [U_1^2, U_2^2, \ldots, U_{N+1}^2]^T$ is the vector containing the square power of the volume velocities. The term $R_b$ is the Bernoulli resistance[15]. 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 domain of validity of the plane waves assumption depends on the radius of the maximal cross-sectional area of the vocal tract. Indeed, the cut-off frequency under which the assumption is considered as valid[21] is

$$f_c = \frac{0.5861 c_s}{2 r_{max}}, \tag{3}$$

where $c_s$ is the speed of sound and $r_{max}$ is the radius of the largest vocal tract section. In this study, $f_c$ differs according to the considered place of articulation: it is around 10 kHz, 8 kHz, and 7 kHz for the palato-alveolar, the alveolar, and the labiodental fricatives, respectively.

### B. Frication noise generation model

The frication noise is generated by a turbulent air flow that appears downstream of the supraglottal constriction[1]. Interactions with obstacles, such as the teeth, the lips, and the walls of the vocal tract, usually occur and impact the acoustic nature of the turbulent noise source. The aeroacoustic mechanisms that are involved in the production of the frication noise are not totally comprehended yet. Consequently, their complicated nature makes them challenging to integrate into simplified acoustic models. Indeed, small variations of geometric or biomechanic parameters, such as the jet angle, the nature of the obstacle, or wall discontinuities downstream of the supraglottal constriction, may significantly modify the spectral characteristics of the produced turbulent noise[3,6]. Simplified models for frication noise sources commonly consider them as acoustic poles, dipoles, or quadripoles, depending on the cause of the turbulence[3,22].

In this paper, the frication noise source is modeled as an acoustic dipole that is activated when the Reynolds number is above a certain critical value, arbitrarily chosen at $Re_c = 1700$, referring to the previous study by Sondhi and Schroeter[23]. The noise source is a bandpass filtered Gaussian white noise[24]. According, to Stevens[1] (p.388, 2nd edition), the amplitude of the noise source is proportional to $U_{DC}^3/a_c^{2.5}$, where $a_c$ is the cross-section area of the supraglottal constriction. Consequently, considering the $i^{th}$ section of the spatially sampled vocal tract, the amplitude of the noise source $P_{n_i}(t)$ at section $i$ and instant $t$ is

$$P_{n_i}(t) = \max\left\{0, \xi w_c(t)\left(Re^2(t) - Re_c^2\right)\frac{U_{DC}^3(t)}{a_{i-1}^{5/2}(t)}\right\}, \tag{4}$$

where $\xi$ is an arbitrarily adjustable real constant used to control the noise level, and $w_c(t)$ is a colored noise function, $U_{DC}$ is the air flow volume velocity inside the vocal tract, and $a_{i-1}$ is the area of the upstream tubelet. $Re = \frac{2\rho}{\mu}\frac{U_{DC}}{\pi\sqrt{a_c}}$ is the Reynolds number of the air flow inside the vocal tract, where $\mu$ and $\rho$ are the shear viscosity and the mass density of the air, respectively.

The computation of $U_{DC}$ follows the method proposed by Maeda[24]: it is the positive solution at each time instant $t$ of the quadratic equation of the low frequency model of the vocal tract, namely

$$\kappa\rho\left[\frac{1}{a_{gl}^2} + \frac{1}{a_c^2}\right]U_{DC}^2 + \left[12\frac{l_{gl}^2 d_{gl}}{a_{gl}^3} + 8\pi\frac{\mu d_c}{a_c^2}\right]U_{DC} - P_{Sub} = 0, \tag{5}$$

where $\kappa$ is a scaling factor, which is 1.42 for a rectangular duct[7], $a$, $l$, and $d$ denote the area, the length, and the width of the glottis (index $gl$) and the supraglottal constriction (index $c$), and $P_{sub}$ is the subglottal pressure.

The frication noise source function $w_c(t)$ is computed from a Gaussian white noise function $w_g(t)$, shaped by the impulse response $g(t)$ of a finite impulse response filter in order to simulate a colored noise, hence

$$w_c(t) = w_g(t) * g(t).$$

Following previous studies[22,25], $g(t)$ is the impulse response of a low-pass filter having a cut-off frequency of $f_c = 0.15\sqrt{\pi}\frac{U_{DC}}{a_c^{3/2}}$. The gain $\xi$ has been empirically chosen. It has been tuned so that, when vowel-fricative-vowel (VFV) pseudowords are simulated, the relative level of energy of the fricative in comparison with the surrounding vowels is similar to that observed in natural speech signals. As expected, the amplitude $\xi$ varies according to the place of articulation, and is higher for sibilant fricatives than for non-sibilant fricatives ($\xi = 4.3\times10^{-7}$ for the palato-alveolar fricatives /ʒ,ʃ/, $\xi = 0.78\times10^{-7}$ for the alveolar fricatives /z,s/, and $\xi = 0.61\times10^{-7}$ for the labiodental fricatives /v,f/).

### C. Glottis model with a membranous glottal gap

The model used to simulate the glottis is similar to the one introduced in recent extensions of the Single-Matrix Formulation paradigm[15] and used to simulate voiced fricatives[14]. It

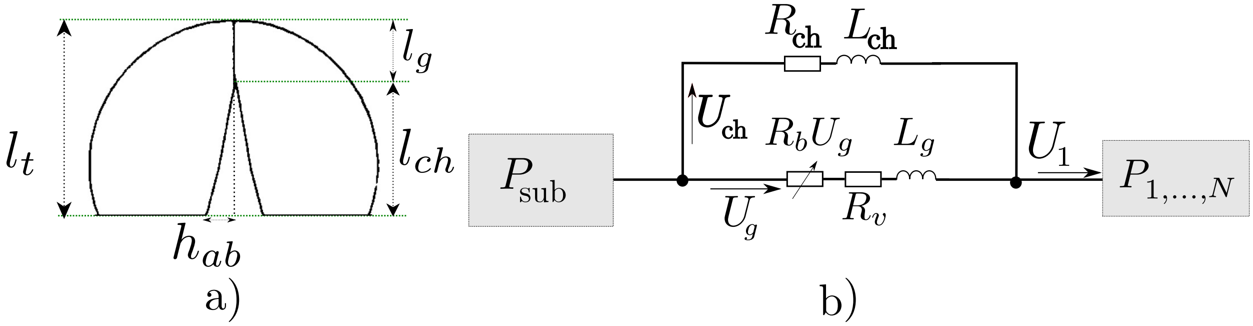


Figure 1. a) View of the glottis model, adapted from Cranen and Schroeter[13]. In this model, the incomplete closure is due to a partial abduction of the vocal folds. $l_g$ is the length of the vibrating part of the vocal fold, $l_{ch}$ is the length of the membranous gap and $l_t = l_g + l_{ch}$ is the total length of the vocal folds. The abduction of the vocal folds is denoted by $h_{ab}$. b) Electric-circuit analogy of the partially closed glottis. $U_{ch}$, $R_{ch}$, and $L_{ch}$ are the volume velocity, the energy loss, and the air inertance inside the membranous gap respectively.

considers two distinct portions of the glottis, represented in Fig. 1: an oscillating part that is computed using a classic two-mass model of the vocal folds[26], and a partially abducted part, the so-called membranous glottal gap, that allows an incomplete closure along the length of the vocal folds during the oscillation cycles. The two-mass model of the vocal folds considers recent improvements to take into account smooth contours, a mobile separation point[26], the viscous losses and the unsteady flow effects[27,28]. Note that other models of vocal folds with consideration of the membranous glottal gap have been proposed in the past[29–31]. We chose our model, which has been specifically designed to be used with the Single-Matrix Formulation paradigm. It should be noted that lumped models, such as the one used in this study, are simplifications of the realistic vocal folds dynamics, and that finite element models might provide a more realistic description of the acoustic impact of the abduction/adduction movement. Yet, lumped models are still useful for speech research as they allow qualitative investigations of the global behavior of the vocal folds to be made using a few parameters and with a very low computation cost, in comparison with finite element models. A review of existing self-oscillating models for the vocal folds may be found in Ref.[32]. Similarly to Birkholz *et al.*[30], our model assumes the angle made by the vocal folds at the membranous gap tip to be constant, hence the length $l_{ch}$ is controlled by the abduction $h_{ab}$:

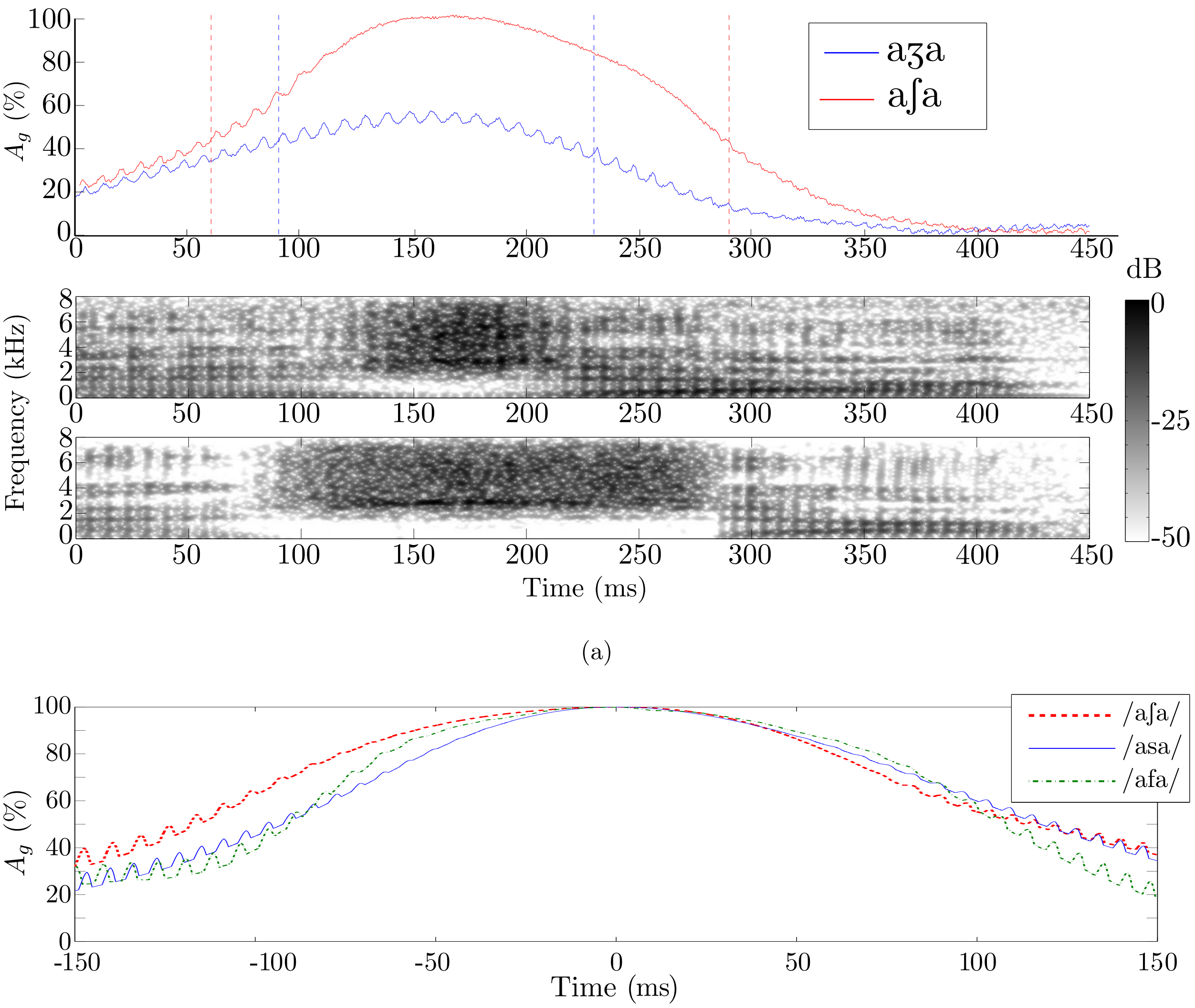
$$l_{ch} = l_t \frac{h_{ab}}{h_{max}}, \tag{6}$$

where $h_{max}$ is the maximal abduction length, i.e. the length $h_{ab}$ for the full glottal abduction. The glottal opening area is then $A_g = h_{ab}l_{ch}$. In order to present generic results, for the rest of the paper, the abduction is controlled by a parameter called *degree of abduction*, denoted by $D_{ab}$ and expressed in percent, which represents the ratio between the area of the glottal gap and that of the fully abducted glottis. That is, $D_{ab} = 0\%$ when there is no gap (full adduction), and $D_{ab} = 100\%$ when the glottis is fully abducted. Note that this ratio is the square power of the ratio $h_{ab}/h_{max}$. We chose to use an area ratio rather than a length ratio because of the possibility to compare with the experimental measurements of the glottal opening area presented in Sec. V C.

The glottal gap has been shown to have significant acoustic impact on speech production, whether on the spectral tilt[13], or on self-oscillating movements of the vocal folds[31]. Except for a few studies[12,14], the acoustic impact of the glottis configuration on the fricative production has been given little interest. However, its role during fricatives is important, since the anticipating abduction movement during the vowel preceeding the consonant modifies the voice quality, which becomes breathy at the vowel offset[12]. It is thus likely that the breathy voice at the consonant onset favors the frication noise appearance during the fricative. Besides, external lighting and sensing photo-glottography (ePGG) measurements[33] (*cf.* Fig. 2) show a glottal opening waveform which is the superimposition of two components, a low frequency component that reaches the maximum around the middle of the fricative segment, and a higher frequency component that corresponds to the oscillations of the vocal folds. The ePGG signal displayed in Fig. 2 is an example of time evolution of the glottal opening chosen among a set of VFV pseudowords recorded for several speakers. The coexistence of these two components confirms the idea of a partial abduction of the glottis, which results in a glottal configuration similar to the glottis configuration shown in Fig. 1.

We have shown that the glottal gap due to the incomplete closure of the vocal folds may be connected to the Single-Matrix Formulation as an acoustic waveguide connected in parallel to the oscillating part of the glottis[15]. One may refer to previous papers[16,20] for details about the numerical computation, and to our previous paper[15] for details about the integration of the glottal leakage side branch.

The ability of our model to reproduce the actual time evolution of the glottal opening area has been tested by running several simulations. For each simulation, the time varying parameter $D_{ab}$ is set as the low frequency component of the ePGG data. The presented



(a)

(b)

Figure 2. a) Top: example of external lighting and sensing photo-glottography (ePGG) curve of a VFV sequence. Here is the minimal pair /aʒa/-/aʃa/. The phonetic segmentation is represented by vertical dashed lines. Middle: wide band spectrogram of the acoustic signal of the utterance /aʒa/. Bottom: wide band spectrogram of the acoustic signal of the utterance /aʃa/. b) Zoom-in view (300 ms window) of the time evolution of the glottal opening area simulated from ePGG data measured for the three places of articulation. The reference time $t = 0$ is the time at the maximal glottal opening.

simulations in Fig. 2 b) correspond to 3 vowel-fricative-vowel (VFV) pseudowords where F is a voiceless fricative, at each place of articulation. Due to the lack of information about the geometric configuration of the vocal tract, it has been set as a static area function for the whole simulated utterance, as it is in the rest of the paper. The obtained time evolution of the simulated glottal opening areas in the case of gradual abduction/adduction movements is very similar to those observed in ePGG measurements, as shown in Fig. 2.

## III. DATA AND METHODS FOR THE NUMERICAL STUDY

### A. Data

### *1. Extracting the area function*

Area functions are extracted from MRI acquisitions of sustained fricatives at the three places of articulation of French fricatives: palato-alveolars (/ʃ,ʒ/), alveolars (/s,z/), and labiodentals (/f,v/). Each shape has been acquired 7 times: for each place of articulation, the subject was asked to articulate the fricative as if he had to pronounce different vowels afterward, namely /i,ɛ,a,o,u,y,ø/. The subject is a volunteer with informed consent and approval of the local ethics committee. It is a male native French speaker who was 33 years old at the time of the acquisitions. The data were collected with an 8-channel neurovascular coil array. The protocol consisted in a 3D volume of the vocal tract acquired with a custom modified Enhanced Fast Gradient Echo (EFGRE3D, TR 3.12 ms, TE 1.08 ms, matrix 256×256×76, with spatial resolution 1.02×1.02×1.0 mm$^3$).

Then, the contours of the vocal tract have been extracted by hand on the midsagittal slice to compute the midline. This line, which should be perpendicular to the propagation of a plane wave in the vocal tract, is used to decompose it into tubelets. The method used to compute the midline is based on dynamic programming to select the best path of segments connecting the larynx to the lips[34]. The determination of the midline is applied either on the whole vocal tract when there is no occlusion, which is the case with the vocal tract shapes of fricatives exploited in the paper, or on all the open sections delimited by the vocal tract extremities or occlusions.

The last step consists in dividing the vocal tract into tubelets perpendicularly to the midline. Attention is paid to the fact that two consecutive tubelets cannot cross in high curvature regions of the vocal tract. Recovering the third dimension from the 2D information, namely the midsagittal distance and the length of each tubelet, in order to estimate the area function has given rise to a number of works[35–37]. However, the improvement with respect to methods derived from that proposed by Heinz and Stevens[38] is not very marked. We thus accepted the $\alpha\ \beta$ parameters proposed by Soquet *et al.*[36]. The choice of the $\alpha - \beta$ transformation over 3D segmentation of the vocal tract is motivated by its generic nature that fits our study. Our main interest is to determine a global shape of the vocal tract, with realistic position, length, and area of the constriction.

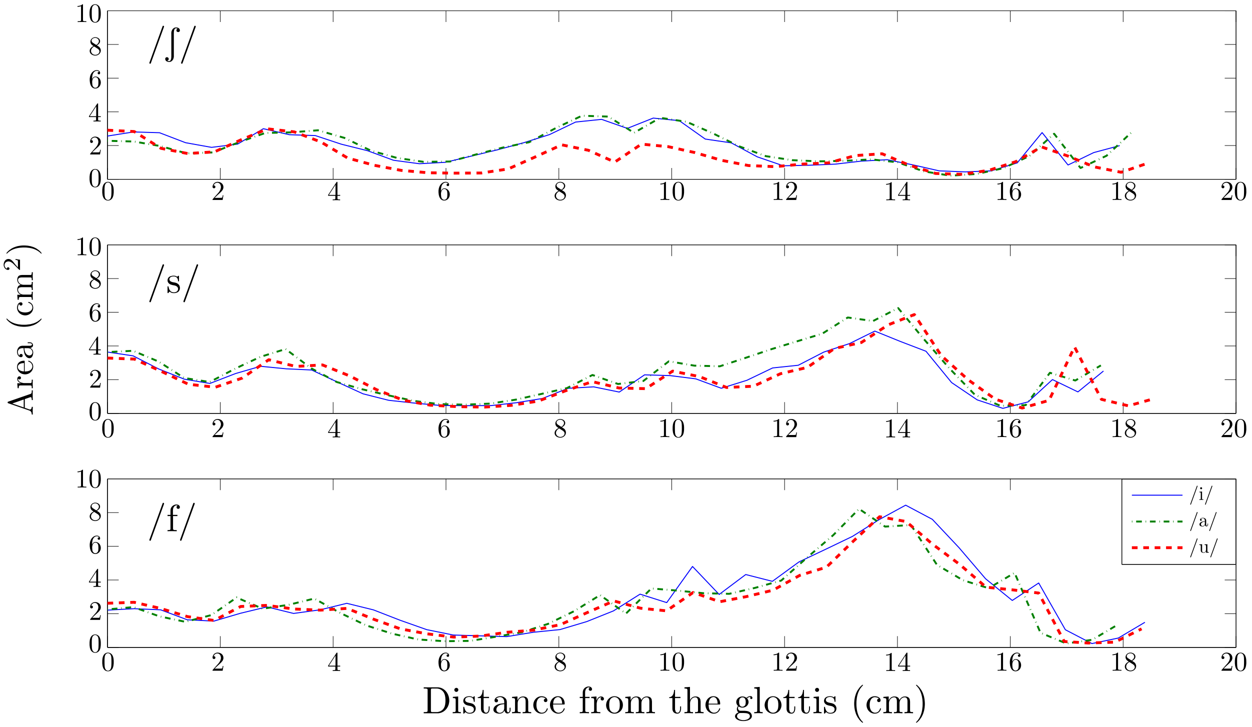


Figure 3. Area functions of the three places of articulation (from top to bottom: /ʃ,s,f/) extracted from MRI data, for 3 cardinal vowel contexts, /i,a,u/.

Fig. 3 shows the extracted area functions. For the sake of clarity, only the cardinal vowel contexts, /i,a,u/, are displayed for each place of articulation.

## *2. Trachea*

The incomplete closure of the glottis during the oscillation cycle of the vocal folds leads to a constant coupling between the vocal tract and the trachea. It is therefore important to account for this coupling by connecting a subglottal waveguide upstream of the glottis. Such connection with the single-matrix formulation has been previously proposed by Ho *et al.*[39], but it uses a very complex geometry of the subglottal system, modeled as a tree-like structure that accounts for all the branching patterns of the bronchial airways. Achieving such a fine degree of modeling is not in the scope of the paper, since it is hard to get such information for real speakers. Consequently, the subglottal system is modeled here as a single waveguide connected to a pressure source located at the input. The area function that is used is borrowed from a previous study by Story[18].

### *3. Glottal parameters*

The input parameters used for the glottis model are the same than the ones used in our previous paper[15]. Values of the mass and stiffness of the vocal folds model have been chosen so that the fundamental frequency of the simulated voiced signal is approximately 150 Hz. This corresponds roughly to the resonance frequency of the lumped mass-spring system. The maximal abduction length $h_{max} = 1$ mm, and the total length of the vocal folds $l_t$ is 2.2 cm, so that the maximal glottal opening $A_{max} = l_t h_{max} = 0.22$ cm$^2$. The length of the vocal folds is set to a rather high value, in comparison with the typical value for males ($\simeq$ 1.6 cm), in order to fit with our subject, who has a low pitched voice and a long vocal tract ($\simeq$ 18 cm, see Fig. 3).

## B. Simulated configurations

The numerical study investigates the acoustic impact of various phonatory and articulatory configurations on the simulated signals, namely the subglottal pressure, the position, and the geometry of the supraglottal constriction, each of them as a function of the glottal abduction degree $D_{ab}$. It consists in simulating voice signals with static vocal tract area functions. For each simulated configuration, $D_{ab}$ varies from 0 to 100% with an increment step of 2%, which results, for each of the configurations, in 51 simulated signals with various glottal openings. Each simulated signal is 200 ms long, with a simulation frequency of 60 kHz, and with an abduction degree $D_{ab}$ that remains constant. The purpose of setting a simulation frequency to 60 kHz, which is far above the frequency domain of interest, is to avoid strong frequency warping[16].

#### *1. Subglottal pressure*

In this paper, the subglottal pressure refers to the constant value assigned to the pressure source connected to the trachea input. In the simulations presented in Sec. IV A, it varies from 500 to 1500 Pa, with an increment step of 100 Pa. These values correspond to subglottal pressures encountered in normal[40] and loud or singing[41] voices. For the simulations presented in Sec. IV B, it is set to a nominal value $P_{Sub} = 1000$ Pa.

#### *2. Supraglottal constriction*

The acoustic impact of the geometry of the supraglottal constriction is studied by modifying its cross-sectional area, denoted by $a_c$, in the area functions extracted from MR images. For the simulations presented in Sec. IV B, $a_c$ is set to different values ranging from 0.1 cm$^2$ to 0.5 cm$^2$. For those presented in Sec. IV A, $a_c$ is taken as the minimal cross-section area extracted from MR images.

### C. Investigated features

#### *1. Regimes of fricatives: voicing quotient and minimal abduction of the vocal folds*

In a previous study[14], it has been shown that, for a given articulatory condition corresponding to a fricative, the simulated speech signal could exhibit three regions with distinct acoustic properties according to the glottal opening: i) an almost purely voiced signal when the glottis is almost entirely adducted (little frication noise is generated, similar to an approximant consonant), ii) a mixed voiced/noisy signal when the level of both voiced component and the frication noise share a similar order of magnitude, and iii) a purely noisy signal, similar to the voiceless fricative, when the voiced component is negligible in comparison with the frication noise. These three regions, respectively denoted $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ in the rest of the paper, are studied as a function of the glottal abduction degree $D_{ab}$. This is done by using a voicing index, named the *voicing quotient*[42]. The voicing quotient (denoted $VQ$ in this paper) is defined as the proportion, expressed in percent, of the energy of the periodic component in the speech signal, hence

$$VQ(\%) = 100 \times \frac{||s_p||_2^2}{||s_p + s_n||_2^2}, \tag{7}$$

where $s_p$ is the periodic (or voiced) component of the signal, and $s_n$ is the frication noise signal. Both periodic and aperiodic components of the simulated signals are computed thanks to a specifically designed periodic/aperiodic decomposition technique of the speech signal, that has been proven to be robust in the case of colored noise such as in voiced fricatives[42]. A value of 100 % indicates a speech signal containing only periodic components, and a value of 0 % indicates a speech signal containing only noisy components. Note that the voicing quotient is directly related to the Harmonics-to-Noise Ratio (HNR) that quantifies the harmonicity of the signal, but the definition of the voicing quotient is more adapted to our study, since it directly quantifies the amount of voicing in fricatives.

For a given area function, varying the glottal abduction degree leads to a decreasing voicing quotient as the glottis opens up. A typical curve, shown in Fig. 4, highlights the presence of these regions. In $\mathcal{A}$, the voicing quotient is constantly high, around 90%, since just a little frication noise is generated so that the voiced components dominate in the simulated speech signal. In $\mathcal{B}$, the frication noise is generated along with a voiced source, the latter being produced by the oscillating part of the vocal folds. The amount of frication noise that is generated is very sensitive to small perturbations of the abduction degree, hence large variations of the voicing quotient. Then, in $\mathcal{C}$, the voicing quotient vanishes. In this regime, the noise source is predominant over the voiced source, leading to a voiceless fricative. This can be seen in the spectrum in Fig. 4, where the harmonics at $f = nF_0$, with $F_0 = 150$ Hz, are visible in the low frequency domain, and vanish in regime $\mathcal{C}$. This observation suggests the existence of two stable regimes ($\mathcal{A}$ and $\mathcal{C}$) and an unstable transition regime, i.e. $\mathcal{B}$.

From the variation of $VQ$ as a function of the abduction degree, it is possible to define two quantities, called the minimal abduction, denoted $D_1$ and $D_2$, and corresponding to the boundaries between regime $\mathcal{A}$ and $\mathcal{B}$, and between $\mathcal{B}$ and $\mathcal{C}$, respectively. They are computed as the two values $D_{ab} = \{D_1, D_2\}$ so that it leads to the best linear fitting in the three regions of the function $VQ = f(D_{ab})$, as shown in Fig. 4.

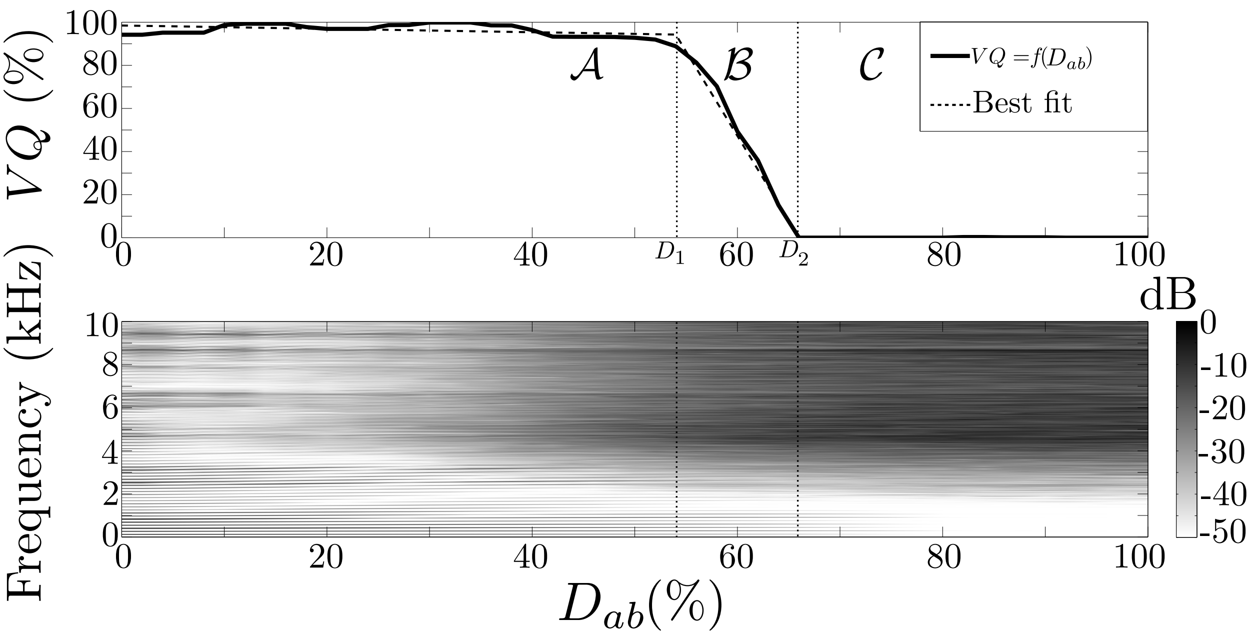


Figure 4. Top: example of curve representing the voicing quotient $VQ$ as a function of the glottal abduction degree $D_{ab}$. The area function is that of the alveolar fricative in the /i/ context. The subglottal pressure value is $P_{sub} = 1000$ Pa, and $a_c = 0.15$ cm$^2$. The optimized linear fitting used to estimate the minimal abductions $D_1$ and $D_2$ is represented by the dashed line. Bottom: Spectrogram representing the evolution of the simulated signal spectrum as a function of $D_{ab}$.

### *2. Spectral characteristics*

The studied spectral characteristics investigated in this paper are the first two spectral moments, namely the spectral centroid, and the spectral spread.

The spectral centroid is a measure of the balance between the low and high frequency contributions in a spectral distribution, it writes

$$S_1 = \frac{\sum_{n=1}^{N} f_n A_n}{\sum_{n=1}^{N} A_n}, \tag{8}$$

where $f_n$, with $n = 1, 2, \ldots, N$, indicates the positive frequency bins of the $2N$-order FFT, and $A_n$ are the corresponding magnitude.

The spectral spread is a measure of the square root of the variance of the spectral distribution. A small spectral spread indicates a spectrum that concentrates its energy in a small frequency range, located in the vicinity of its centroid. It writes

$$S_2 = \sqrt{\frac{\sum_{n=1}^{N} A_n (f_n - S_1)^2}{\sum_{n=1}^{N} A_n}}. \tag{9}$$

Both $S_1$ and $S_2$ are expressed in Hz and are computed in the frequency domain ranging from 50 Hz to 10 kHz.

## IV. ACOUSTIC FEATURES

### A. Effect of the subglottal pressure

#### *1. Voicing quotient*

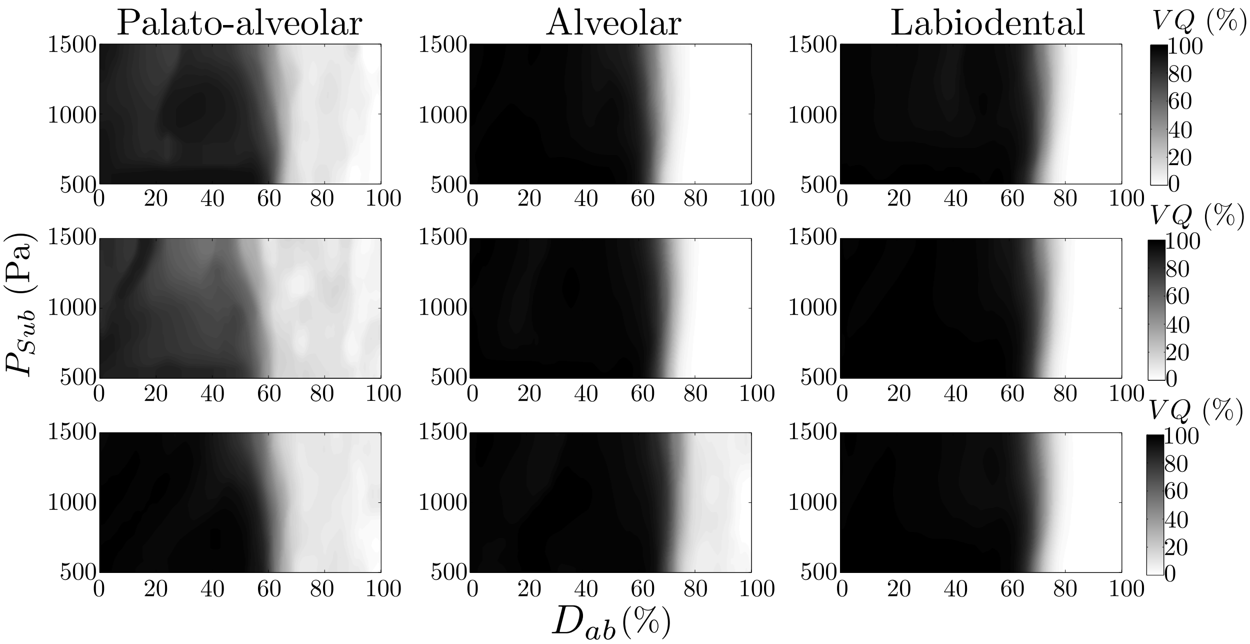


Figure 5. Voicing quotient of simulated voice signals as a function of the subglottal pressure $P_{Sub}$ and the glottal abduction degree $D_{ab}$. Each column of figures corresponds to a place of articulation. From left to right: palato-alveolar fricatives /ʃ,ʒ/, alveolar fricatives /s,z/, and labiodental fricatives /f,v/. Each row of figures corresponds to a vowel context, /i,a,u/, from top to bottom.

Fig. 5 shows the voicing quotient as a function of the glottal abduction degree and the subglottal pressure. For the sake of clarity, it does not show results of all of the 21 configurations (7 contexts for each of the 3 places of articulation), but only those corresponding to the context of cardinal vowels, namely /i,a,u/. The same simplification applies in the rest of the paper. The general behavior of the voicing quotient is similar to the typical curve displayed in Fig. 4: it is constantly high for weak abductions, then it suddenly plunges at a certain point and vanishes at a second critical point. For palato-alveolar fricatives, these critical abduction degrees tend to be smaller when the subglottal pressure rises. Indeed, when the subglottal pressure increases, this raises the Reynolds number due to an higher DC component of the airflow volume velocity inside the vocal tract. Consequently, the Reynolds critical number $Re_c$ above which the frication noise is generated is reached at smaller glottal openings. In a $D_{ab} - P_{Sub}$ plane, as shown in Fig. 5, it results in the left part exhibiting high values of voicing quotient, and the right part exhibiting low values, both parts being separated by a small transition area. Note that there is no significant variations with the subglottal pressure for the alveolar and labiodental fricatives.

Also, for palato-alveolar fricatives, the glottal abduction degree for which the voicing quotient starts to significantly decrease is smaller than for other places of articulation. The lowering of this critical abduction seems to be more important for palato-alveolar fricatives. The vowel context seems to have little influence on the voicing quotient. Indeed, for each place of articulation, the shapes of the voicing quotient values for the three vowel contexts are very similar.

### *2. Spectral centroid*

The spectral centroid as a function of phonatory conditions shows distinct areas, as evidenced in Fig. 6. At the bottom left, where the abduction degree and the subglottal pressure are small, the spectral centroid is constantly between 1000 and 2000 Hz. At the top right, when both the abduction degree and the subglottal pressure are large, the spectral centroid is much higher, and lies around 4500 Hz for palato-alveolar fricatives, and up to 6000 Hz for alveolar fricatives. In these areas, the spectral centroid is relatively stable in regards with variations of the phonatory configurations. The values in this region depend on the place of articulation considered: alveolar fricatives have the highest values of spectral centroid, around 6 kHz, then the labiodental fricatives have a spectral centroid around 5.6 kHz, and finally, the palato-alveolar fricatives have a spectral centroid around 4.5 kHz. These values are in agreement with data observed in natural speech by Jongman *et al.*[2] for English,

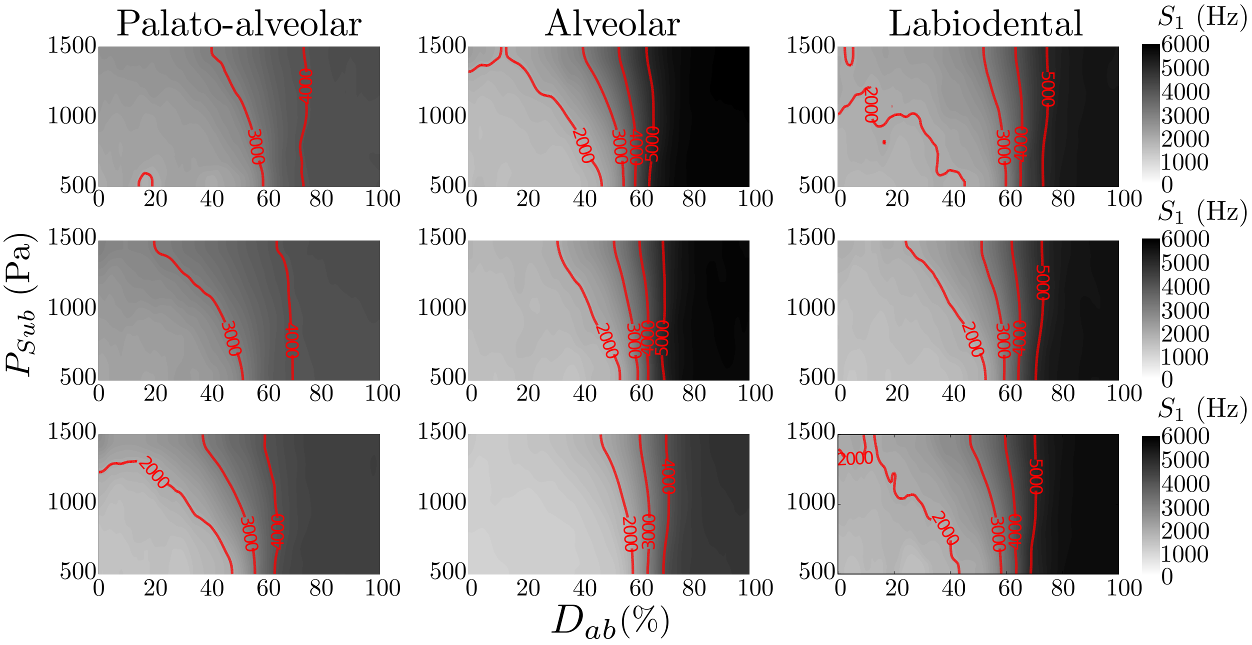


Figure 6. Spectral centroid of simulated voice signals as a function of the subglottal pressure $P_{Sub}$ and the glottal abduction degree $D_{ab}$. Each column of figures corresponds to a place of articulation. From left to right: palato-alveolar fricatives /ʃ,ʒ/, alveolar fricatives /s,z/, and labiodental fricatives /f,v/. Each row of figures corresponds to a vowel context, /i,a,u/, from top to bottom. Contour lines at a 1000 Hz-interval are represented by solid lines.

and Lonchamp for French[43]. The alveolar fricatives of English and French speakers exhibit the higher spectral centroid than other fricatives, whereas palato-alveolar fricatives have the lowest spectral centroid. Same data also showed that, like in the presented simulations, voiced fricatives exhibit lower spectral centroids than their voiceless counterpart. A small transition area may be seen. In this range, small perturbations of the glottal opening or of the subglottal pressure lead to large variations of the spectral centroid.

Similarly to the voicing quotient, the vowel context seems to have no significant influence on the spectral centroid of the simulated fricative. However alveolar and palato-alveolar fricatives simulated in the /u/ context exhibit slightly different patterns of spectral centroid values than for /i/ and /a/ context. For palato-alveolar fricatives, the high values (above 4 kHz) are reached for weaker abductions and smaller subglottal pressures than for other places of articulation. Alveolar fricatives in the /u/ context show smaller spectral centroids (around 4.8 kHz, while it is around 6 kHz for /i,a/ contexts). This may be due to the strong lip protusion and the small lip opening of the speaker in that case, as shown by the area functions in Fig. 3, in comparison with the other contexts. The main effect is then to significantly lower the formant frequencies, and consequently the spectral centroid.

Variation patterns of the spectral centroid as a function of the glottal abduction may be explained by the fact that when the subglottal pressure and the glottal opening are small, the aerodynamics conditions are not fulfilled to generate a frication noise, so that the simulated speech signal contains only periodic, or voiced components. In that case, the energy is mainly concentrated in the low frequency range, hence a low spectral centroid. Then, when the glottal opening and the subglottal pressure are sufficiently large to generate a frication noise (central region), noisy components arise in the simulated speech signals. This has for effect to enhance the high frequency range of the spectrum. Hence the rise of the spectral centroid. A slight increase of either the subglottal pressure or the glottal opening in this regime leads to a decrease of the voiced component level, and an increase of the noise level, hence the rise of the spectral centroid. When the voiced components have disappeared, at high values of $D_{ab}$ and $P_{Sub}$, and when the simulated signals contain only noise, the spectral centroid reaches its maximal value, since the low frequency domain of the spectrum is significantly weakened by the absence of the voiced contributions. This is confirmed by the similarities between the pattern of voicing quotient values in Fig. 5 and that of spectral centroid values in Fig. 6.

### *3. Spectral spread*

The plot of the spectral spread as a function of $P_{Sub}$ and $D_{ab}$, shown in Fig. 7, also highlights the presence of areas that are related to both the voicing quotient and the spectral centroid. Again, the palato-alveolar fricative exhibits a different behavior than other places of articulation. For alveolar and labiodental fricatives, the spectral spread values at the bottom left corner is low, and then suddenly increases to reach a maximum around 3 kHz. At the top right corner, the spectral spread is also small, with values at the same order of magnitude as at the bottom left corner. They are also relatively stable. These three areas are approximately located at the same place as the three areas of the spectral centroid. Palato-alveolar fricatives present smaller values of the spectral spread, and they are relatively constant independently of the position in the $D_{ab} - P_{Sub}$ plane.

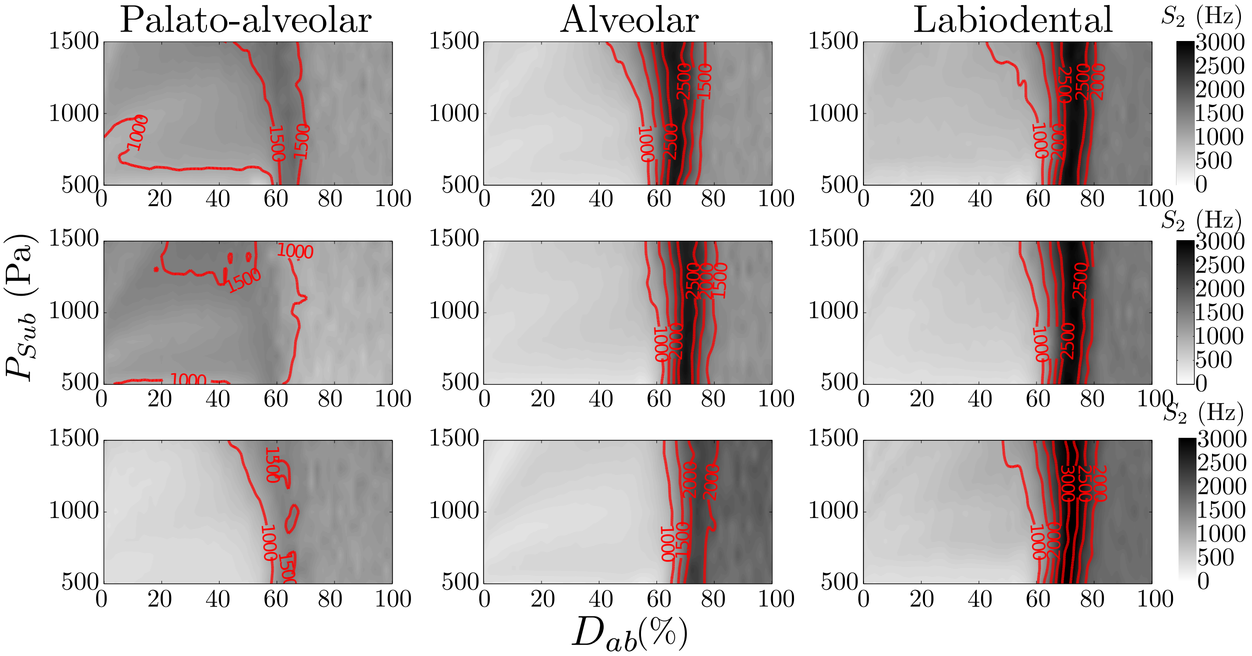


Figure 7. Spectral spread of simulated voice signals as a function of the subglottal pressure $P_{Sub}$ and the glottal abduction degree $D_{ab}$. Each column of figures corresponds to a place of articulation. From left to right: palato-alveolar fricatives /ʃ,ʒ/, alveolar fricatives /s,z/, and labiodental fricatives /f,v/. Each row of figures corresponds to a vowel context, /i,a,u/, from top to bottom. Contour lines at a 500 Hz-interval are represented by solid lines.

Similarly to the spectral centroid, the existence of three different regions, as well as their characteristics may be explained by the degree of contribution of the frication noise. In weak abduction and low subglottal pressure conditions, the absence of frication noise yields to an almost purely harmonic spectrum with a marked spectral slope. Hence a small spectral spread. As long as no frication noise is generated, the spectral shape remains the same independently of the values of the glottal opening and the subglottal pressure. As soon as the frication noise is generated, the spectrum of the simulated fricative is enhanced in the high frequency domain. At the same time, the voiced component is still predominant in the low frequency range. As a consequence, the spectral spread is large, since the signal contains energy both in the low frequency domain (voiced component) and in the high frequency domain (noisy component). The spectral spread value is very sensitive to the balance between the voiced component energy level and the noise level, hence quick changes in the spectral spread values in the central area. Finally, when values of both $P_{Sub}$ and $D_{ab}$

are large enough so that there is no voiced components in the simulated fricative any longer, the spectrum in the low frequency domain is significantly weakened so that the energy is concentrated in the mid- and high-frequency domains, hence smaller values of the spectral spread.

In the transition regime, where the spectral spread is unstable, the alveolar fricative in the /u/ context (around 2150 Hz) shows smaller spectral spreads than in other contexts (around 2600 Hz). However, in the stable region corresponding to the maximal glottal abduction, the spectral spread is then larger in the /u/ context (around 1800 Hz) than in other contexts (around 1300 HZ). There is no significant differences among vowel contexts for the labiodental fricative. The differences observed for the /u/ context are certainly due to the fact that the lip protrusion is stronger, and that the lips are closer each other than in other contexts. However, labiodental fricatives are characterized by small lip opening whatever the vowel context, hence little acoustic differences among contexts.

## B. Influence of the cross-sectional area of the supraglottal constriction

### *1. Voicing quotient*

The voicing quotient as a function of the glottal abduction degree $D_{ab}$ and the supraglottal constriction area $a_c$ is shown in Fig. 8. As expected, the voicing quotient is high for weak abductions, then dramatically decreases at a critical abduction degree, and vanishes for large values of $D_{ab}$. The influence of the supraglottal constriction area is limited when $a_c$ is larger than 0.3 cm$^2$. Under this value, $a_c$ has a more marked influence: the value of $D_{ab}$ above which the voicing quotient vanishes exhibits a local minimum for $a_c$ around 0.1 cm$^2$.

### *2. Spectral centroid*

Plots of the spectral centroid of the simulated signals as a function of $D_{ab}$ and $a_c$, shown in Fig. 9, exhibit complicated patterns. A strong glottal abduction leads to higher values of the spectral centroid. The supraglottal constriction area has also a significant influence on the spectral centroid. Basically, a small $a_c$ leads to a high value of the spectral centroid. It results in low values of the spectral centroid at the top left corner, and high values at the bottom right corner of the $D_{ab} - a_c$ plane. There is no significant differences according to the

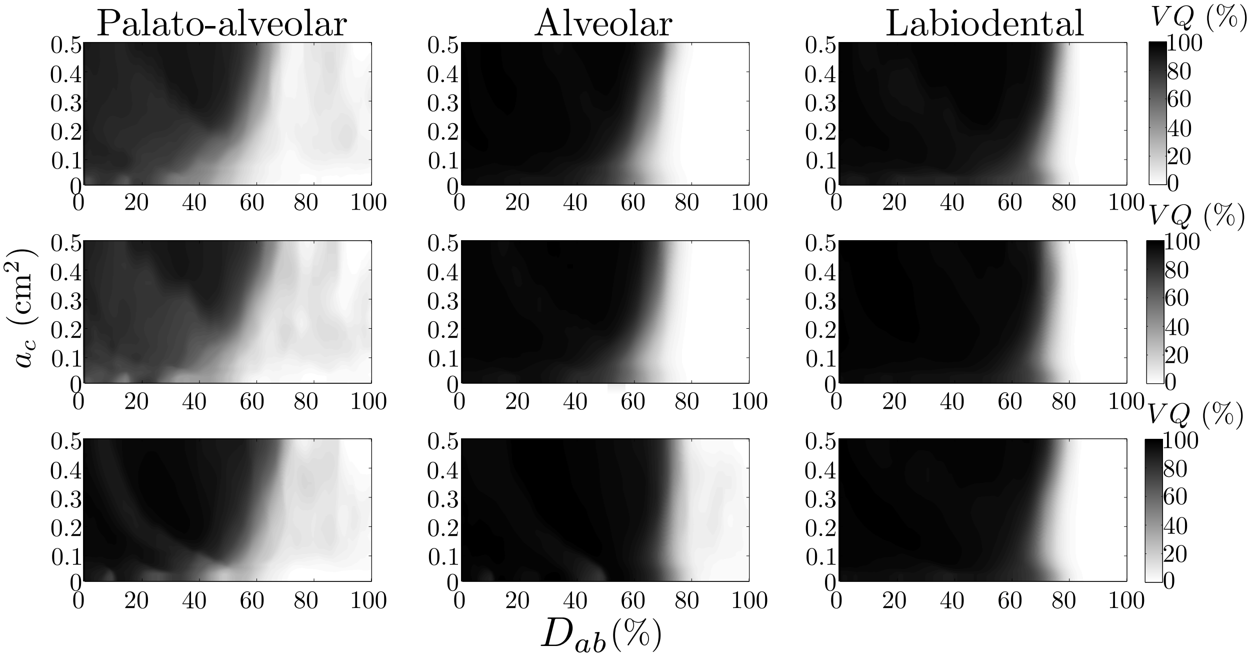


Figure 8. Voicing quotient of simulated voice signals as a function of the supraglottal constriction area $a_c$ and the glottal abduction degree $D_{ab}$. Each column of figures corresponds to a place of articulation. From left to right: palato-alveolar fricatives /ʃ,ʒ/, alveolar fricatives /s,z/, and labiodental fricatives /f,v/. Each row of figures corresponds to a vowel context, /i,a,u/, from top to bottom.

vowel context, except for /su/, which exhibits smaller spectral centroids, as already observed in Fig. 6. It is also worth noting than for a given $a_c$, the spectral centroid suddenly rises for $D_{ab}$ values that roughly correspond to quick variations of the voicing quotient in Fig. 8. The local minimum at $a_c = 0.1$ cm$^2$ is then clearly visible for alveolar and labiodental fricatives. It is less marked for palato-alveolar fricatives, as they exhibit more complicated patterns. This suggests that the cross-sectional area of the supraglottal constriction in postalveolar fricatives seems to have a more significant impact on the spectral properties.

### *3. Spectral spread*

Fig. 10 shows the spectral spread as a function of $D_{ab}$ and the constriction area $a_c$. For a given constriction area, one can observe the same behavior as that in Fig. 7: the spectral spread is the smallest at both extremities of the glottal abduction degree, and reaches a maximum in a central area. The spectral spread is also larger for narrow constrictions.

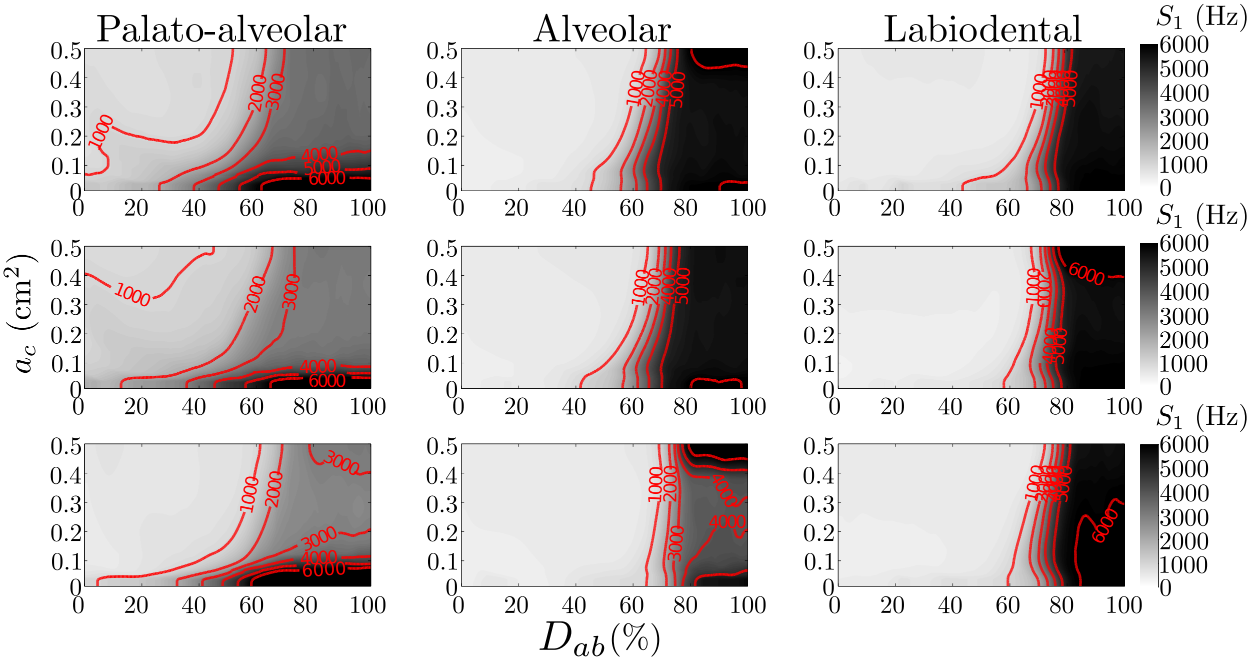


Figure 9. Spectral centroid of simulated voice signals as a function of the supraglottal constriction area $a_c$ and the glottal abduction degree $D_{ab}$. Each column of figures corresponds to a place of articulation. From left to right: palato-alveolar fricatives /ʃ,ʒ/, alveolar fricatives /s,z/, and labiodental fricatives /f,v/. Each row of figures corresponds to a vowel context, /i,a,u/, from top to bottom. Contour lines at a 1000 Hz-interval are represented by solid lines.

Similarly to what has been observed in Fig. 7, palato-alveolar fricatives have smaller spectral spreads than other places of articulation. This is certainly due to the fact that postalveolar fricatives present a predominant peak$^2$, located between 2 and 3 kHz, which concentrates the energy of the spectrum in this frequency domain. Like for the spectral centroid, there are no significant differences in the spectral spread patterns among the different vowel contexts.

## V. MINIMAL ABDUCTION DEGREES

In this section, two non-dimensional quantities are defined according to the method explained in Sec. III C: $D_1$ is the minimal abduction degree from which frication noise is generated, namely the boundary between $\mathcal{A}$ and $\mathcal{B}$, and $D_2$ is the minimal abduction degree from which the noise component is predominant over the voiced component, namely the boundary between $\mathcal{B}$ and $\mathcal{C}$. A third quantity is $\Delta D = D_2 - D_1$, which is the width of the

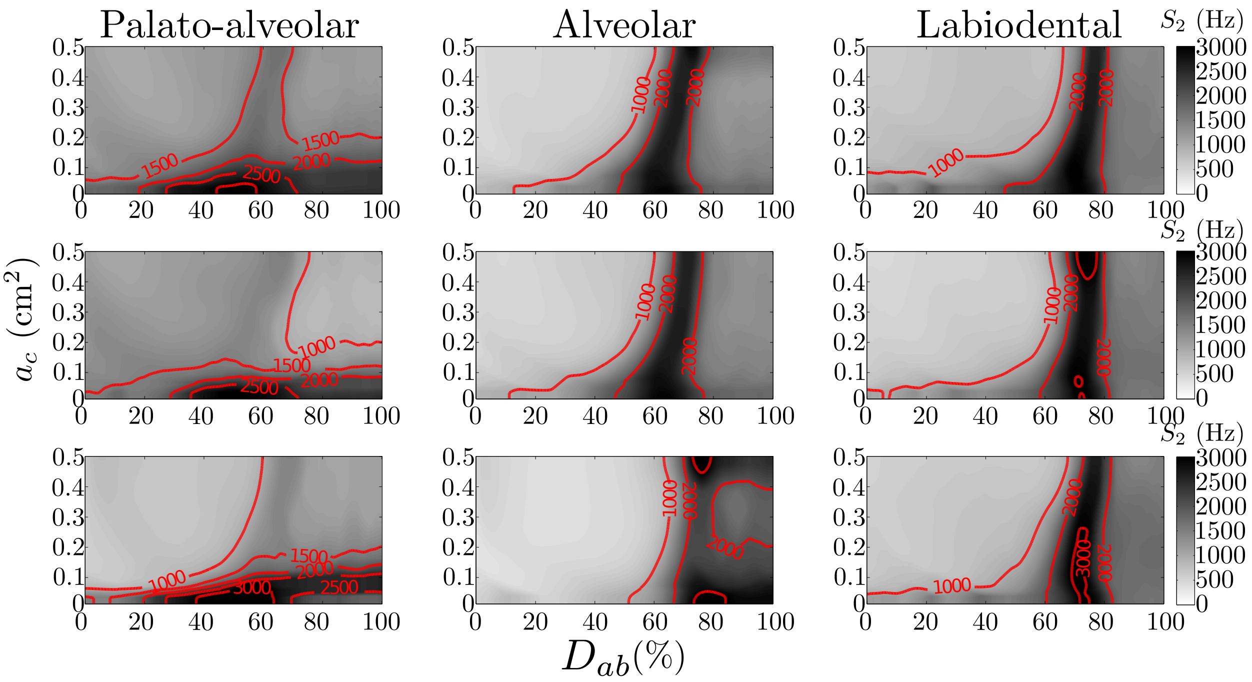


Figure 10. Spectral spread of simulated voice signals as a function of the supraglottal constriction area $a_c$ and the glottal abduction degree $D_{ab}$. Each column of figures corresponds to a place of articulation. From left to right: palato-alveolar fricatives /ʃ,ʒ/, alveolar fricatives /s,z/, and labiodental fricatives /f,v/. Each row of figures corresponds to a vowel context, /i,a,u/, from top to bottom. Contour lines at a 500 Hz-interval are represented by solid lines.

transition regime $\mathcal{B}$.

## A. Effect of the subglottal pressure

Fig. 11 shows the median value of $D_1$, $D_2$, and $\Delta D$ as a function of $P_{Sub}$ for the three places of articulation, as well as the median absolute deviation. Increasing the subglottal pressure has for main effect to decrease $D_1$, while there is no significant impact on $D_2$, which remains constant. For $D_1$, this is due to the fact that when $P_{Sub}$ is high, the low-frequency component of the acoustic volume velocity inside the vocal tract increases as the pressure drop between the subglottal region and the mouth increases. This results in the rise of the Reynolds number, hence smaller minimal abduction degrees. The evolution of $D_1$ and $D_2$ as a function of $P_{Sub}$ are qualitatively similar for all places of articulation. The median value among the group of palato-alveolar fricatives is the lowest for both $D_1$ and

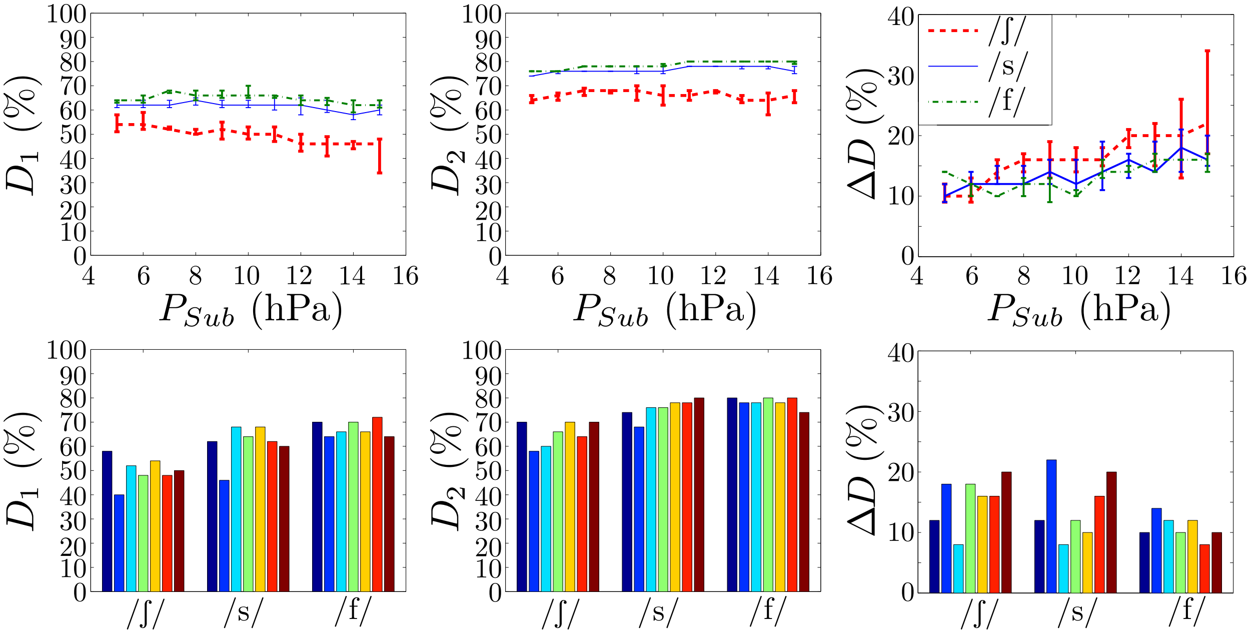


Figure 11. Top: median values of $D_1$, $D_2$, and $\Delta D$ as a function of $P_{Sub}$ for the three fricatives. Bottom: values of $D_1$, $D_2$, and $\Delta D$ at $P_{Sub} = 1000$ Pa for each extracted area function. For each group of 7 area functions, they correspond to the following vowels /i,ɛ,a,o,u,y,ø/, respectively.

$D_2$. This suggests that small variations of the vocal tract geometry, and especially at the supraglottal constriction, may significantly modify the acoustic features of the produced fricative. Interestingly, $\Delta D$ slightly increases with $P_{Sub}$: it is 10% for $P_{Sub} = 500$ Pa, and goes up to approximately 20% for $P_{Sub} = 1500$ Pa. It is also interesting to note that the values are globally very similar for each place of articulation and for each realization. This result suggests that for any articulatory configuration, the abduction range required to produce voiced fricatives depends mostly on the subglottal pressure and is not dependent on the phonological context.

## B. Effect of the cross-section area of the supraglottal constriction

Fig. 12 represents the minimal abduction degrees $D_1$, $D_2$, and $\Delta D = D_2 - D_1$, as a function of the cross-sectional area of the supraglottal constriction $a_c$. The values of the minimal abduction degree are still systematically lower for the palato-alveolar fricatives than for the other fricatives. Although very close, they are also smaller for alveolar fricatives than for labiodental fricatives. $D_1$ admits a local minimum at $a_c = 0.1$ cm$^2$. $D_2$ does not show significant modifications after the local minimum. As a consequence, $\Delta D$ admits a local maximum at $a_c = 0.1$ cm$^2$ for the palato-alveolar and the labiodental fricatives ($\Delta D = 22\%$ and 18%, respectively), and at $a_c = 0.05$ cm$^2$ for the alveolar fricative ($\Delta D = 26\%$).

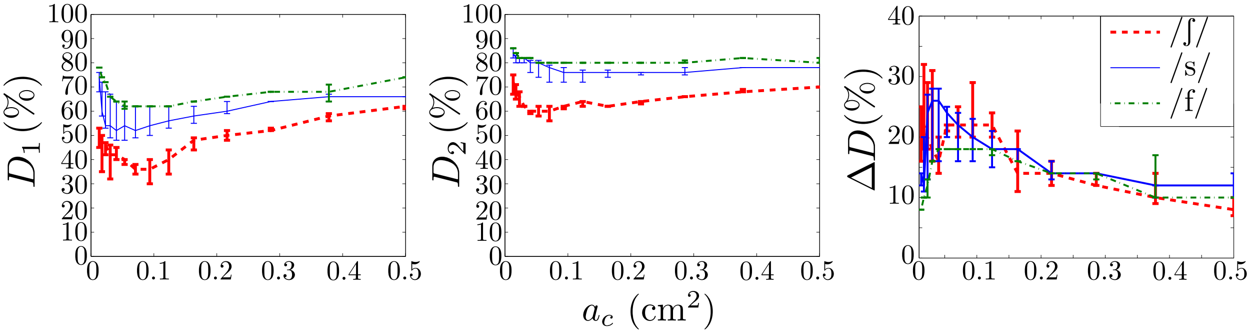


Figure 12. From left to right: $D_1$, $D_2$, and $\Delta D$, as a function of the supraglottal constriction area $a_c$, and for the three places of articulation.

## C. Experimental observations

This section reports experimental observations of the relationship between acoustic features of speech signal and glottal opening. For that purpose, simultaneous acquisitions of audio speech signals and glottal opening measurements, via the external lighting and sensing photo-glottography (ePGG)[33], have been performed on a series of vowel-fricative-vowel (VFV) pseudowords uttered by a French native, male speaker.

For the experiments, both ePGG and audio signals are sampled at 20 kHz. The ePGG device[33] is a non-invasive technique that consists in converting the light going through the glottis, from a light source located on the surface of the side neck, into electric current, thanks to a photosensor unit located on the speaker's front neck. The glottal aperture is then deduced from the electric current generated by the photosensor, given the reasonable assumption that it is in its linear domain. Consequently, by normalizing the glottal opening by its maximal value, the glottal opening $A_g$ provided by the ePGG data can be directly related to the glottal abduction degree $D_{ab}$ used in the numerical study. The ePGG device has been previously successfully used to investigate glottal opening gestures (e.g. in Ref.[44]). The speaker is asked to utter VFV pseudowords where the vowels are the same, chosen among the three cardinal vowels /i,a,u/, and the fricatives are the 6 French fricatives /v,f,s,z,ʃ,ʒ/.

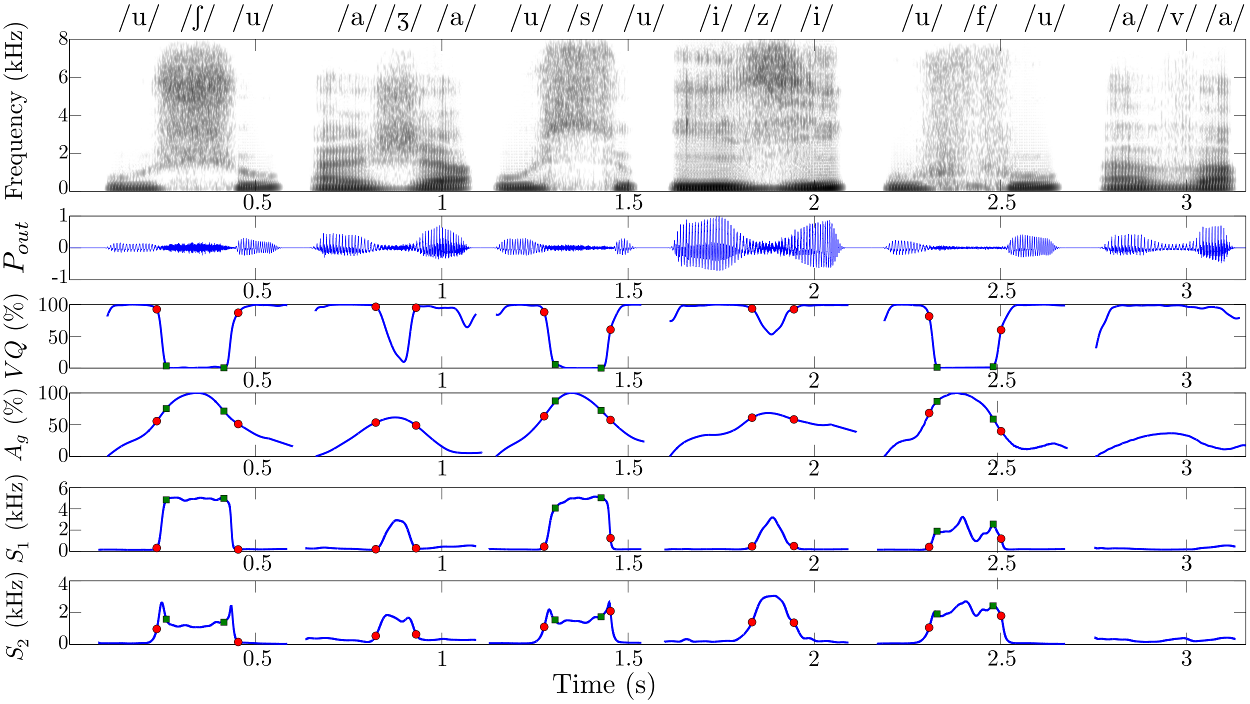


Figure 13. From top to bottom: spectrogram, acoustic pressure waveform $P_{out}$, voicing quotient $VQ$, relative glottal area $A_g$, and the spectral moments $S_1$ and $S_2$, computed from the audio recordings of 6 pseudowords, /uʃu/, /aʒa/, /usu/, /izi/, /ufu/, /ava/. Circle marks denote the boundaries between $\mathcal{A}$ and $\mathcal{B}$, and square marks denote the boundaries between $\mathcal{B}$ and $\mathcal{C}$.

Fig. 13 shows an example of recorded data, with ePGG measurements, the voicing quotient $VQ$, and the first spectral moments $S_1$ and $S_2$ computed from the audio recordings. It corresponds to 6 pseudowords representing one example of each fricative: /uʃu/, /aʒa/, /usu/, /izi/, /ufu/, /ava/. Interestingly, the acoustic features computed from the audio recording behave accordingly to the observations from the numerical study: the voicing quotient plunges at the beginning of the fricative, and then suddenly increases again at its end. During this very fast evolution of the voicing quotient, the glottal opening does not exhibit such inflections. From the $VQ$ curves, it is possible to estimate the moment where the speaker is in the transition regime $\mathcal{B}$: it starts when the voicing quotient starts to suddenly drop, and finishes at the moment when $VQ$ vanishes, namely when the speaker enters into regime $\mathcal{C}$. Symmetrically, the speaker quits $\mathcal{C}$ to produce $\mathcal{B}$ when $VQ$ starts to rise again.

Experimental minimal areas $A_1$ and $A_2$ can then be estimated by reporting the values of the glottal opening at these instants. They have been carefully estimated by hand and are shown in Fig. 13 with circle ($A_1$), and square ($A_2$) marks. Again, with the sole exception of the /f/ offset, the experimental values $A_1$ and $A_2$ are similar to those from the numerical study: $A_1$ lies between 49 and 65%, and $A_2$ is between 71 and 85%. The difference $\Delta A = A_2 - A_1$ is constantly between 14 and 20%. The critical areas $A_1$ and $A_2$ are systematically lower at the fricative offset than at the fricative onset, which suggests a hysteresis effect, not investigated by our numerical study. This is particularly true for /f/, where $VQ$ rises at the fricative offset almost at the end of the glottal adduction movement. It is also worth noting that the glottal opening during the /ava/ utterance has not reached the critical glottal opening areas, hence a constantly high $VQ$, and that it denotes an alternative way to pronounce voiced fricatives, in comparison with the realization of the other fricatives shown in Fig. 13.

The behavior of the spectral moments is also interesting in regards with our numerical simulations. The spectral centroid and the spectral spread are low and constant during the vowel, as in regime $\mathcal{A}$. Once the glottal opening reaches the critical opening $A_1$, both spectral moments suddenly rise: the spectral centroid constantly rises up to a maximal value when the glottal opening reaches the critical value $A_2$, while the spectral spread reaches a local maximum, then decreases until a constant value when $A_g = A_2$. This corresponds to the behavior of the spectral centroid and the spectral spread in regime $\mathcal{B}$, as shown in Figs. 6 and 7. This phenomenon occurs symmetrically at the fricative offset. For voiceless fricatives, when $A_g$ is larger than $A_2$, i.e. in regime $\mathcal{C}$, both spectral moments are stable.

### D. Discussion in regards to articulatory strategies

Our simulations give a new point of view about strategies available to implement the voice/unvoiced contrasts of fricatives. Here we exploit the results of our simulations together with the observations of French speakers and German learners of French[45]. In this previous work dedicated to language learning we recorded sentences with voiced fricatives in a word final position uttered by French and German speakers, and the same words (in an isolated condition) by French speakers to investigate voicing without the influence of the initial vowel of the next word. The final devoicing of voiced fricatives of French by German learners of French allows the differences in the realization of voiced fricatives to be highlighted.

In our simulations, the computed values of $D_1$ and $D_2$, whether as a function of the subglottal pressure or the cross-section area of the supraglottal constriction, lead to very small values of $\Delta D$, namely under 20 %. In this range, corresponding to the unstable transition regime $\mathcal{B}$, small perturbations of the vocal tract and the glottal configurations lead to significant modifications of the acoustic characteristics. Consequently, sustaining $\mathcal{B}$ to produce voiced fricatives may be very difficult for the speaker, and falling into regime $\mathcal{C}$ is very likely. On the contrary, voiceless fricatives is much easier because it requires to open the glottis sufficiently to be in regime $\mathcal{C}$, which is very stable. However, since the abduction/adduction movement of the vocal folds is relatively slow, the speaker goes from $\mathcal{A}$ to $\mathcal{C}$ by transiting through $\mathcal{B}$ during a short moment, leading to voiced frames in the fricative segment. Then, in order to contrast voiced and voiceless fricatives, the speaker may have several strategies. A first strategy would consist in producing a weak frication noise, i.e. maintaining regime $\mathcal{A}$ for the whole duration of the voiced fricative, as for the /ava/ utterance in Fig. 13. Secondly, the speaker may stay in the transition regime $\mathcal{B}$. Since the boundary between the voiced and unvoiced regimes is almost vowel independent for each of the places of articulation (see Fig. 11), that probably enables the control of this regime to be mastered by a speaker. However, this requires to reduce the duration of the fricative, since $\mathcal{B}$ is very unstable. Our acoustic measurements[45] show that some French speakers (approximately one third for postalveolar fricatives) sustain voicing for the whole fricative even in a final position. This corresponds to one of these two strategies. However, there is a substantial part of French speakers who fail to produce voicing for the whole fricative even if perception tests confirmed that the corresponding fricatives are perceived as voiced. The third production strategy resorts on a less precise control of the glottal opening which results in a brief excursion in regime $\mathcal{C}$, hence some unvoiced frames. The abduction/adduction movement corresponding to the devoicing and then revoicing of speech is all the easier since there is a vowel after the fricative as exhibited by Fig. 2. This probably explains why French speakers realize a release vocal schwa (/ə/) after a voiced fricative in final position. This also explains that short voiceless segments of fricatives have been shown to be an acoustic clue for voicing perception[46].

The first and second strategies have been observed in some French speakers, among a corpus of 45 speakers, for producing voiced fricatives in final position[45]. Interestingly, these strategies are used by 27 % of the speakers for /ʒ/, by 62 % of the speakers for /z/, and by 75% of the speakers for /v/. In regards with the results presented in this paper, this makes sense since palato-alveolar is the place of articulation that systematically presents the shorter region $\mathcal{A} \cup \mathcal{B}$ (smaller values of $D_1$ and $D_2$), followed by alveolars, and by labiodentals. Hence difficulties to maintain these regimes.

## VI. CONCLUSION

The paper has presented a numerical study about the influence of the existence of a membranous glottal gap due to the gradual abduction/adduction movement of the vocal folds on several acoustic features of produced fricatives. Simulations used a recent glottis model that is connected with the classic 1D wave solver based on a transmission line circuit analog framework. It is, to the best of our knowledge, the first study about the acoustic impact of fine glottal configurations, such as the partial abduction and the incomplete closure of the glottis, on the production of voiced fricatives.

Simulations have highlighted the existence of three distinct regimes of fricative production, depending on the amount of frication noise that is generated. The first regime, labeled $\mathcal{A}$ in this paper, corresponds to an almost purely voiced signal, corresponding to an approximant consonant, where the DC component of the volume velocity in the vocal tract is too low to generate a frication noise with a significantly high acoustic level. In this regime the spectral centroid and the spectral spread are relatively low, and perturbations of the speaker configurations, such as the glottal abduction degree, or the geometry of the supraglottal constriction, do not significantly modify the spectral features. The second one, labeled $\mathcal{B}$, is an unstable transition regime that corresponds to the situation where the voiced and the frication noise components of the speech signal have similar energy. In the transition regime $\mathcal{B}$, the spectral centroid and the spectral spread are higher than in the first regime, because of the presence of the noise component that enforces the high frequency domain of the uttered voice. Unlike regime $\mathcal{A}$, small perturbations of the speaker configurations significantly modify the acoustic features. Finally, the third regime, labeled $\mathcal{C}$, corresponds to voiceless fricatives, i.e. when the frication noise component is predominant over the voiced component. In regime $\mathcal{C}$, like in regime $\mathcal{A}$, the spectral features are very stable in regards with variation of the speaker configurations. However, unlike in regime $\mathcal{A}$, the spectral centroid is high and the voicing quotient is almost null. The existence of these regimes has been evidenced for each of the three places of articulation of French fricatives. The presence and the characteristics of these regimes have also been observed and evidenced experimentally thanks to simultaneous audio and glottal opening recordings on a real speaker.

In the articulatory-phonatory space, the transition regime $\mathcal{B}$ is the one with the smallest extent, confirming the fact that voiced fricatives are a difficult-to-produce class of consonants, because of the very specific aeroacoustic conditions required. Simulations have shown that the range of glottal abduction for producing voiced fricatives do not vary significantly with the place of articulation, nor with the phonological context, and not significantly with the geometry of the supraglottal constriction. On the contrary, this range varies with the subglottal pressure: the greater the pressure the longer the range of the glottal abduction degree.

In terms of articulatory strategy, $\mathcal{B}$ is not suitable to sustain fricatives since it is an unstable transition regime, hence several alternatives to contrast voiced fricative with voiceless fricatives: i) staying in $\mathcal{A}$ by favoring the voicing over the frication noise, or ii) reducing the length of the fricative segment to maximize the proportion of the amount of time in $\mathcal{B}$ in relation to the whole fricative segment. The latter strategy, although used by many speakers[46], may lead to inappropriate coordination between the articulatory configurations and the configurations at the glottis. This may explain the presence of final devoicing in fricatives, which is frequently observed in many languages[10,47], or the partial devoicing of voiced fricatives[45].

The presented study also shows that the consideration of the partial glottis abduction in articulatory synthesis is important to simulate natural running speech. However, there is still a lack of experimental measurements of the precise time evolution of the glottal opening, which limits the use of realistic time scenarios of the coordination between the vocal tract and the glottis. This paper provides a few examples of External lighting and sensing photoglottography (ePGG) data that confirm the observations made from the numerical study. Additional data, in addition to articulatory gestures, should be acquired in the next future in order to thoroughly study the tract-glottis coordination.

Finally, the proportion of direct and fine control by the speaker on the glottal abduction degree, due to active muscle control, over that of the fluid-structure interactions resulting from the aeroacoustic conditions is not fully comprehended yet. Possible future improvements of glottis models that account for the potential uncontrolled glottal abduction, due to fluid-structure interactions at the vicinity of the glottis, would be a step forward the full comprehension of the production of voiced fricatives.

## 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, and to Dr. Shinji Maeda, Dr. Angélique Amelot, and Pr. Didier Demolin for the ePGG measurements.

## REFERENCES

[1]K. Stevens, *Acoustic Phonetics* (MIT Press, Cambrige, MA), 543–555 (1998).

[2]A. Jongman, R. Wayland, and S. Wong, "Acoustic characteristics of english fricatives", The Journal of the Acoustical Society of America **108**, 1252–1263 (2000).

[3]C. H. Shadle, "Articulatory-acoustic relationships in fricative consonants", in *Speech production and speech modelling*, 187–209 (Springer) (1990).

[4]G. Ramsay and C. Shadle, "The influence of geometry on the initiation of turbulence in the vocal tract during the production of fricatives", in *Proc. 7th Int. Seminar on Speech Production (ISSP7)*, 8 (2006).

[5]C. H. Shadle, M. I. Proctor, K. Iskarous, and M. A. Berezina, "Revisiting the role of the sublingual cavity in the /s/–/ʃ/ distinction.", J. Acoust. Soc. Am. **125(4)**, 2569–2569 (2009).

[6]S. S. Narayanan, A. A. Alwan, and K. Haker, "An articulatory study of fricative consonants using magnetic resonance imaging", The Journal of the Acoustical Society of America **98**, 1325–1347 (1995).

[7]K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords", Bell Syst. Tech. J. **51(6)**, 1233–1268 (1972).

[8]K. N. Stevens, S. E. Blumstein, L. Glicksman, M. Burton, and K. Kurowski, "Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters", The Journal of the Acoustical Society of America **91(5)**, 2979–3000 (1992).

[9]L. M. Jesus and C. H. Shadle, "A parametric study of the spectral characteristics of european portuguese fricatives", Journal of Phonetics **30**, 437–464 (2002).

[10]O. Dmitrieva, A. Jongman, and J. Sereno, "Phonological neutralization by native and non-native speakers: The case of russian final devoicing", Journal of phonetics **38**, 483–492 (2010).

[11]D. Pape, L. M. Jesus, and P. Birkholz, "Intervocalic fricative perception in European Portuguese: An articulatory synthesis study", Speech Communication **74**, 93 – 103 (2015).

[12]A. Löfqvist, L. L. Koenig, and R. S. McGowan, "Vocal tract aerodynamics in /aCa/ utterances: Measurements", Speech Communication 49–66 (1995).

[13]B. Cranen and J. Schroeter, "Modeling a leaky glottis", Journal of Phonetics **23**, 165 – 177 (1995).

[14]B. Elie and Y. Laprie, "A glottal chink model for the synthesis of voiced fricatives", in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5240–5244 (2016).

[15]B. Elie and 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**, 85–96 (2016).

[16]S. Maeda, "A digital simulation method of the vocal-tract system", Speech Communication **1**, 199–229 (1982).

[17]J. L. Kelly and C. C. Lochbaum, "Speech synthesis", in *Proceedings of the Fourth International Congress on Acoustics*, 1–4 (1962).

[18]B. H. Story, "Phrase-level speech simulation with an airway modulation model of speech production", Computer Speech & Language **27(4)**, 989–1010 (2013).

[19]B. J. Kröger and P. Birkholz, "Articulatory synthesis of speech and singing: State of the art and suggestions for future research", in *Multimodal Signals: Cognitive and Algorithmic Issues*, 306–319 (Springer) (2009).

[20]P. Mokhtari, H. Takemoto, and T. Kitamura, "Single-matrix formulation of a time domain acoustic model of the vocal tract with side branches", Speech Communication **50(3)**, 179 – 190 (2008).

[21]L. Eriksson, "Higher order mode effects in circular ducts and expansion chambers", The Journal of the Acoustical Society of America **68(2)**, 545–550 (1980).

[22]S. Narayanan and A. Alwan, “Parametric hybrid source models for voiced and voiceless fricative consonants”, in *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, volume 1, 377–380 (IEEE) (1996).

[23]M. M. Sondhi and J. Schroeter, “A hybrid time-frequency domain articulatory speech synthesizer”, IEEE Trans. Acoust. Speech Sig. Process. **35(7)**, 955–967 (1987).

[24]S. Maeda, “Phoneme as concatenable units: VCV synthesis using a vocal tract synthesizer”, in *Sound Patterns of Connected Speech: Description, Models and Explanation, Proceedings of the symposium held at Kiel University, Arbeitsberichte des Institut für Phonetik und digitale Spachverarbeitung der Universitaet Kiel:31*, 145–164 (1996).

[25]P. Birkholz, “Enhanced area functions for noise source modeling in the vocal tract”, in *10th International Seminar on Speech Production, Köln*, 1–4 (2014).

[26]N. J. C. Lous, G. C. J. Hofmans, R. N. J. Veldhuis, and A. Hirschberg, “A symetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prothesis design”, Acta Acustica **84**, 1135–1150 (1998).

[27]C. Vilain, X. Pelorson, C. Fraysse, M. Deverge, A. Hirschberg, and J. Willems, “Experimental validation of a quasi-steady theory for the flow through the glottis”, J. of Sound and Vibration **276(3-5)**, 475 – 490 (2004).

[28]L. Bailly, X. Pelorson, N. Henrich, and N. Ruty, “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).

[29]B. Kröger, “Simulation of vocal fold oscillation behaviour by a self-oscillating glottis model”, Le Journal de Physique IV **4**, C5–457 (1994).

[30]P. Birkholz, B. J. Kröger, and C. Neuschaefer-Rube, “Articulatory synthesis of words in six voice qualities using a modified two-mass model of the vocal folds”, in *First International Workshop on Performative Speech and Singing Synthesis*, volume 370, 1–8 (Vancouver, BC, Canada) (2011).

[31]M. Zañartu, G. E. Galindo, B. D. Erath, S. D. Peterson, G. R. Wodicka, and 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)**, 3262–3271 (2014).

[32]B. D. Erath, M. Zañartu, K. C. Stewart, M. W. Plesniak, D. E. Sommer, and S. D. Peterson, “A review of lumped-element models of voiced speech”, Speech Communication

**55**, 667–690 (2013).

[33] K. Honda and S. Maeda, "Glottal-opening and airflow pattern during production of voiceless fricatives: a new non-invasive instrumentation", The Journal of the Acoustical Society of America **123(5)**, 3738–3738 (2008).

[34] Y. Laprie, M. Loosvelt, S. Maeda, E. Sock, and F. Hirsch, "Articulatory copy synthesis from cine X-ray films", in *Interspeech 2013 (14th Annual Conference of the International Speech Communication Association)*, 1–5 (Lyon, France) (2013).

[35] R. S. McGowan, M. T.-T. Jackson, and M. A. Berger, "Analyses of vocal tract cross-distance to area mapping: An investigation of a set of vowel images", J. Acoust. Soc. Am. **131(1)**, 424–434 (2012).

[36] A. Soquet, V. Lecuit, T. Metens, and D. Demolin, "Mid-sagittal cut to area function transformations: Direct measurements of mid-sagittal distance and area with MRI", Speech Communication **36(3)**, 169–180 (2002).

[37] P. Perrier, L.-J. Boë, and R. Sock, "Vocal tract area function estimation from midsagittal dimensions with two sets of coefficients", Journal of Speech, Language and Hearing Research **35**, 53–87 (1992).

[38] J. M. Heinz and K. N. Stevens, "On the relations between lateral cineradiographs, area functions, and acoustic spectra of speech", in *Proc. 5th Int. Congress of Acoustics*, volume 44, A44 (1965).

[39] J. C. Ho, M. Zañartu, and G. R. Wodicka, "An anatomically based, time-domain acoustic model of the subglottal system for speech production", J. Acoust. Soc. Am. **129(3)**, 1531–1547 (2011).

[40] A. Giovanni, D. Demolin, C. Heim, and J.-M. Triglia, "Estimated subglottic pressure in normal and dysphonic subjects", Annals of Otology, Rhinology & Laryngology **109(5)**, 500–504 (2000).

[41] J. Sundberg, N. Elliot, and P. Gramming, "How constant is subglottal pressure in singing", Speech Transmission Laboratory Quarterly Progress Scientific Report **32(1)**, 53–63 (1991).

[42] B. Elie and G. Chardon, "Robust tonal and noise separation in presence of colored noise, and application to voiced fricatives", in *Proceedings of the 22th International Congress on Acoustics*, 1–10 (Buenos Aires, Argentina) (2016).

[43] F. Lonchamp, "Description acoustique (acoustic description)", in *La parole et son traitement automatique (Speech and its automatic processing)*, edited by Calliope and J. Tubach,

chapter 3 (Masson, Paris) (1989).

[44] T. Kamiyama, B. Kühnert, and J. Vaissière, "Do French-speaking learners simply omit the English /h/?", in *The 17th International Congress of Phonetic Sciences (ICPhS XVII)*, 1010–1013 (Hong Kong, Hong Kong SAR China) (2011).

[45] S. Ghosh, C. Fauth, A. Sini, and Y. Laprie, "L1-L2 interference: The case of final devoicing of french voiced fricatives in final position by german learners", in *Proc. of the Interspeech 2016*, 3156–3160 (2016).

[46] K. N. Stevens, S. E. Blumstein, L. Glicksman, M. Burton, and K. Kurowski, "Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters", The Journal of the Acoustical Society of America **91(5)**, 2979–3000 (1992).

[47] J. Charles-Luce, "Word-final devoicing in german and the effects of phonetic and sentential contexts", The Journal of the Acoustical Society of America **77**, S85–S85 (1985).