Glottal opening and strategies of production of fricatives

Benjamin Elie[1], Yves Laprie[1]

[1]INRIA/CNRS/université de Lorraine, France

benjamin.elie@loria.fr, yves.laprie@loria.fr

Abstract

This work investigates the influence of the gradual opening of the glottis along its length during the production of fricatives in intervocalic contexts. Acoustic simulations reveal the existence of a transient zone in the articulatory space where the frication noise level is very sensitive to small perturbations of the glottal opening. This corresponds to the configurations where both frication noise and voiced contributions are present in the speech signal. To avoid this unstability, speakers may adopt different strategies to ensure the voiced/voiceless contrast of fricatives. This is evidenced by experimental data of simultaneous glottal opening measurements, performed with ePGG, and audio recordings of vowel-fricative-vowel pseudowords. Voiceless fricatives are usually longer, in order to maximize the number of voiceless time frames over voiced frames due to the crossing of the transient regime. For voiced fricatives, the speaker may avoid the unstable regime by keeping low frication noise level, and thus by favoring the voicing characteristic, or by doing very short crossings into the unstable regime. It is also shown that when speakers are asked to sustain voiced fricatives longer than in natural speech, they adopt the strategy of keeping low frication noise level to avoid the unstable regime.

Index Terms: Fricative, Voicing, Glottal chink, Phonetics

1. Introduction

The abduction/adduction movement of the glottis in vowel-consonant-vowel (VCV) utterances is relatively slow. It has been shown [1] that it started before the beginning of the consonant, and ended after the second vowel onset. Electroglottography (EGG) measurements on real speakers [2] show a gradual decrease of the contact area of the vocal folds at the vowel-fricative transition, and conversely, a gradual increase at the fricative-vowel transition. These results suggest that, during the abduction movement, the vocal folds open gradually along their length, leading to a membranous glottal gap, also called the *membranous glottal chink* [3, 4], and gradually close during the adduction movement.

So far, the role of the membranous gap in the production of fricatives has been given little interest. Studies about the glottal chink have focused on its effect on the vowels spectral slope [3], or on the vocal folds dynamic [4]. It has also been used to simulate different modes of phonation [5, 6], since the glottal leakage is known to impact the voice quality and more specifically the breathiness [7]. Recently, we proposed a glottal chink model for simulating voiced fricatives [8]. A numerical study (not published yet [9]) has shown that it plays a major role in the production of fricatives.

Indeed, this numerical study has investigated the acoustic impact of the membranous chink opening on several acoustic features of the simulated fricative, including the first spectral moments and a voicing index, called the *voicing quotient*. Simulations have revealed the presence of three different regimes

of fricative production, called $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$, characterized by the balance between the contributions of the voiced source and of the frication noise. They are delimited by the membranous chink opening. Basically, $\mathcal{A}$ is stable, and shows high voicing and low frication noise, $\mathcal{B}$ shows mixed frication noise and voiced source, but is very unstable: small perturbations of the chink opening leads to large variations of the frication noise level. Besides, the opening range of $\mathcal{B}$ is very small in comparison with other regimes. $\mathcal{C}$ is stable, and corresponds to a voiceless signal: the frication noise is predominant over the voiced source. This regime occurs when the membranous glottal gap is big, or when the glottis is fully abducted.

Consequently, it is unlikely that speakers sustain regime $\mathcal{B}$ for producing voiced fricatives since it may be very difficult. Going through the unstable zone $\mathcal{B}$ for a short moment before sustaining the stable regime $\mathcal{C}$ is, however, the unique solution for producing a voiceless fricative in intervocalic contexts. This results in a few voiced frames in the fricative segment, i.e. short segments of voiceless fricatives are voiced. When uttering voiced fricatives, the speaker may have several strategies to ensure the good contrast with its voiceless counterpart: i) staying in $\mathcal{A}$ in order to favor 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. In contrast, voiceless fricatives are made longer in order to minimize the proportion of voiced frames in the fricative segment [10].

These observations open new perspectives for understanding the strategies available to implement the voice/unvoiced contrasts of fricatives, or for explaining some phenomena, such as the final fricative devoicing [11–13]. In order to complete the numerical observations, this paper aims at experimentally verifying the existence of such an unstable zone in fricatives, and also the existence of these potential different strategies. In this paper, we chose to focus on French sibilant fricatives. For that purpose, we used simultaneous recordings of the audio speech signal, and measurements of the glottal opening with an external lighting and sensing photo-glottography (ePGG) device [14]. Thus, it is possible to relate the time evolution of the voicing quotient, derived from the audio recording, to the glottal opening given by ePGG data. This is a first attempt to experimentally study the presence of this unstable zone.

2. Numerical observations

In [9], simulations used static configurations of the vocal tract. They were derived from static MRI acquisitions on a subject, asked to sustain French fricatives at the three places of articulation, i.e. labiodental (/f,v/), alveolar (/s,z/), and palato-alveolar (/ʃ,ʒ) fricatives. Then, simulations consisted in varying the membranous chink length, denoted l_{ch}, from 0 (no glottal chink), to 100% (fully abducted glottis).

The effect of the membranous opening on the voicing quotient gave interesting results. Voicing quotient (VQ) is defined

as the ratio between the energy of the voiced contributions in the speech signal and the total energy of the signal. 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. Both periodic and aperiodic components of the simulated signals are computed thanks to a specifically designed periodic/aperiodic decomposition technique of the speech signal [15].

Results of the numerical study in [9] revealed that, for a given area function, varying the membranous chink opening leads to a decreasing voicing quotient as the chink opens up. A typical curve, shown in Fig. 1, highlights the presence of the aforementioned three different regimes of fricative production, $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$. 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 $\mathcal{B}$, frication noise is generated along with voiced source due to the oscillating part of the vocal folds. The amount of frication noise generated is very sensitive to the value of the chink opening, hence large variations in the value 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. 1, where the harmonics at $f = nF_0$, with $F_0 = 150$ Hz, are visible in the low frequency domain, and vanish in regime $\mathcal{C}$.
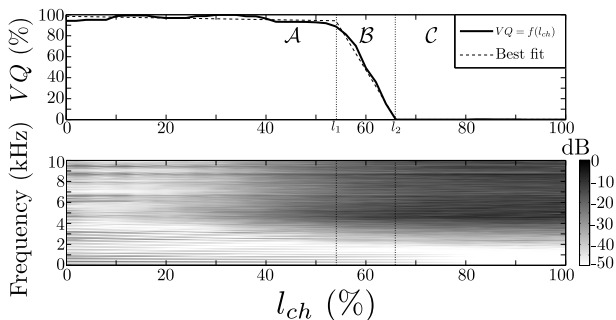

Figure 1: *Top: example of curves representing the voicing quotient VQ as a function of the membranous chink opening l_{ch}, extracted from [9]. The area function is the alveolar fricative in the /i/ context. The critical lengths l_1 and l_2 represent the boundaries between the regimes. Bottom: Spectrogram representing the evolution of the simulated signal spectrum as a function of the chink opening.*

From the variation of VQ as a function of the chink opening, it is possible to define two quantities, called the critical lengths, denoted l_1 and l_2, and corresponding to the boundaries between regimes $\mathcal{A}$ and $\mathcal{B}$, and between regimes $\mathcal{B}$ and $\mathcal{C}$, respectively. They are computed as the two values $l_{ch} = \{l_1, l_2\}$ which lead to the best linear fitting in the three regions of the function $VQ = f(l_{ch})$, as shown in Fig. 1. Note the unstable nature of $\mathcal{B}$, which also has a very small range of range chink opening: the computed values of $\Delta l = l_2 - l_1$, i.e. the length of the transition zone $\mathcal{B}$, are very small, namely around 10 % in this example.

3. Experimental study: data and methods

In the experiments, both ePGG and audio signals are sampled at 20 kHz. The ePGG device [14] 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. It has been previously successfully used to investigate glottal opening gestures (e.g. in [16]).

A French native speaker was asked to produce vowel-fricative-vowel (VFV) pseudowords, where both vowels are the same, chosen among the three cardinal vowels /i,a,u/, and the fricatives are the French sibilant fricatives /s,z,ʃ,ʒ/. The speaker was sometimes asked to exaggeratedly lengthen the fricative. The purpose is to be able to compare the strategies of producing voiced fricatives in the case of short and long phonetic segments. For each fricative, and for each configuration (short and long), we obtained 4 realizations of each intervocalic context (/aFa/, /iFi/, and /uFu/), this results in 2 realizations of 3 intervocalic contexts of 4 fricatives in 2 length configurations, hence 48 utterances.

Fig. 2 shows an example of recorded data, with ePGG measurements, and computed VQ curves from the audio recordings, for 2 pseudowords, /aʃa/ and /aʒa/. Differences between a voiceless fricative and its voiced counterpart are clearly highlighted: the voiceless fricative is longer, its voicing quotient is null for most of its duration, and the glottal opening is larger than for the voiced fricative. Interestingly, the existence of the unstable zone revealed by the numerical simulations is also visible. Indeed, for each fricative, the voicing quotient suddenly drops 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. For the voiceless fricative, the fall of VQ from 80% to 6% and its rise from 0% to 70% occur when the glottal opening varies of around 20% of the maximal opening. For the voiced fricative, there is a drop from 80% to 20% of the voicing quotient with very little glottal opening variation.
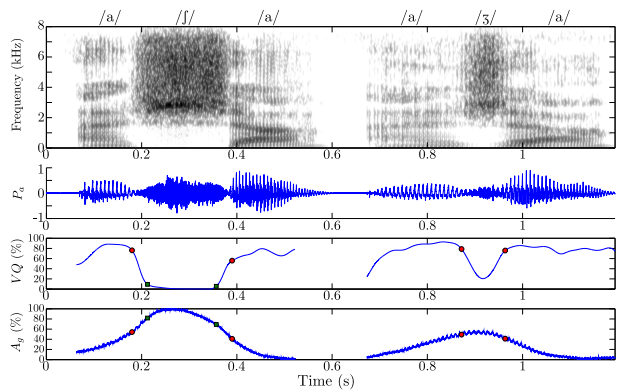

Figure 2: *From top to bottom: spectrogram, acoustic pressure waveform P_a, voicing quotient VQ, and relative glottal area A_g, for two pseudowords, /aʃa/ and /aʒa/. Circle marks denote the boundaries between $\mathcal{A}$ and $\mathcal{B}$, and square marks denote the boundaries between $\mathcal{B}$ and $\mathcal{C}$.*

4. Experimental study: results

4.1. The different strategies for producing voiced fricatives

Fig. 3 shows examples of the different strategies used by the speaker to realize the sibilant voiced fricatives of French in

intervocalic contexts. For each fricative, the two mentioned strategies are realized. In the /aFa/ context where the fricative is short, a quick incomplete abduction/adduction of the glottis leads to a dip in the voicing quotient, which falls at around 15% for /ʒ/, and slightly under 50% for /z/. The other strategy consists in staying in regime $\mathcal{A}$ by reducing the glottal opening. This results in a constantly high voicing quotient ($VQ \geq 80\%$). The second strategy is used by the speaker when the fricative is longer. This is certainly due to the fact that he is unable to stay in $\mathcal{B}$ due to its unstable nature. Consequently, the unique solution to contrast the fricative with the voiceless counterpart is to stay in $\mathcal{A}$ in order to emphasize the fricative voicing.
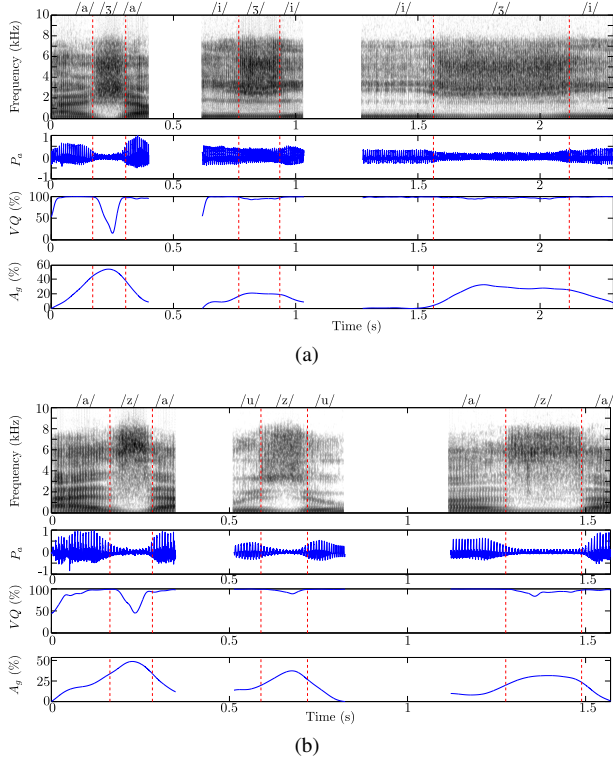

(a)

(b)

Figure 3: *Examples of the different strategies used for producing voiced fricatives in intervocalic context. From top to bottom: wide band spectrogram, acoustic pressure waveform, voicing quotient, and glottal opening, for (a) palato-alveolar and (b) alveolar voiced fricatives. The phonetic segmentation of the fricatives are displayed as vertical dashed lines.*

4.2. Critical glottal opening

Since the experiments deal with glottal area opening measurements, the critical values to define the $\mathcal{A}/\mathcal{B}$ and $\mathcal{B}/\mathcal{C}$ boundaries are called critical areas, A_1 and A_2, respectively. For each voiceless fricative, the critical opening A_1 and A_2 have been estimated as the glottal opening at the different time points $t_{1,1}$, $t_{2,1}$, $t_{2,2}$, and $t_{1,2}$. We consider $t_{1,1}$ as the moment when the voicing quotient starts to suddenly plummet, $t_{2,1}$ is the moment when VQ vanishes, $t_{2,2}$ is the moment when VQ starts to rise again, and $t_{1,2}$ is the moment when VQ finishes its sudden ascent at the end of the fricatives. A_1 is then the glottal opening at instants $t_{1,1}$ and $t_{1,2}$, and A_2 is the glottal opening at instants $t_{2,1}$ and $t_{2,2}$. A_1 and A_2 are then normalized by the maximal
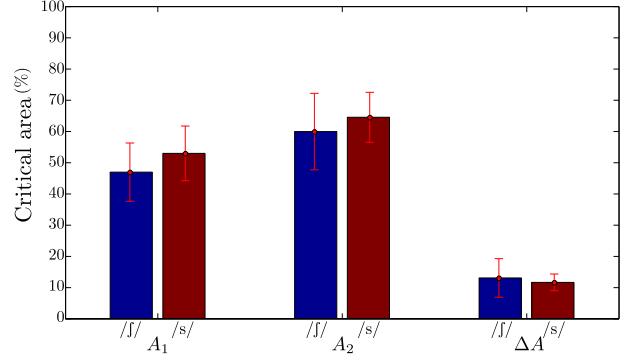

Figure 4: *Mean values of the estimated critical areas A_1, A_2 and ΔA for the palato-alveolar fricative /ʃ/, and the alveolar fricative /s/.*

glottal opening and expressed in percent. They have been carefully estimated by hand.

Fig. 4 shows the mean values of A_1, A_2, and Δ_A, as well as the standard deviation, for each place of articulation. Results confirm the numerical simulations obtained in [9]: the range Δ_A is very short, between 10 and 20% for each place of articulation. A_1 is around 40-50%, and A_2 around 60-70%. A_1 and A_2 are smaller for the palato-alveolar fricative than for the alveolar fricative. The existence of the unstable region between the two stable regimes $\mathcal{A}$ and $\mathcal{C}$, which has been first revealed by the numerical simulations, is now evidenced experimentally by these results.

4.3. Voicing quotient in voiced fricatives

This section compares the strategies in the case of a short and a long realization of the fricative. For that purpose, voicing quotient values of the time frames of the fricative segment are stored in a single vector corresponding to the short fricative realization and another one corresponding to the long realization of the fricative. It is done separately for each place of articulation. If the speaker realizes the strategy of keeping a low voicing quotient, the resulting vector should contain mainly small coefficients. Conversely, if the speaker uses the strategy of keeping a high voicing quotient, the vector should contain mainly large coefficients.
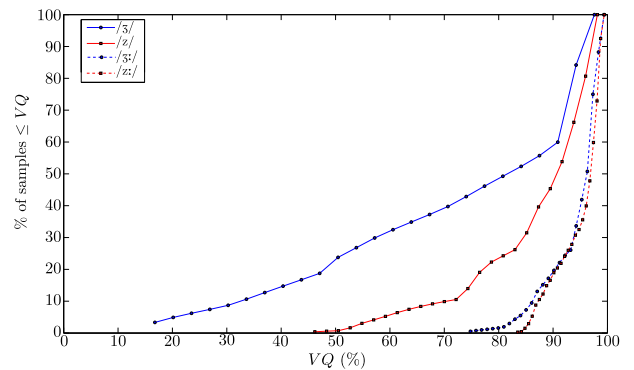

Figure 5: *Proportion of coefficients under VQ as a function of VQ, in short (solid line) and long (dashed line) fricative realizations.*

Fig. 5 shows the proportion of coefficients in the stored vec-

tor that are less than or equal to the VQ value in abscissa, for short and long realizations of the fricative, at each place of articulation. Long fricatives contain mainly high voicing quotient values: only 17% are less than or equal to 90%, while it is between 54% (alveolar) and 60% (palato-alveolar) for short fricatives. The lowest VQ coefficient for long alveolars is 84% (46% for short alveolars) and 75% for long palato-alveolars (17% for short ones). This confirms that speakers adopt the strategy of keeping high voicing quotient values in long fricatives. In that case, there is no significant differences between both places of articulation: they show similar patterns of voicing quotient coefficients. Results also show that palato-alveolar fricatives present lower voicing quotient coefficients than alveolar fricatives.

5. Conclusion

The study has experimentally evidenced the existence of an unstable zone during the production of fricatives, which had been previously observed only via a numerical study [9]. In this region, small perturbations of the glottal opening leads to large variations of the voicing quotient, which quantifies the proportion of the voiced contributions in the speech signals. It appears between two critical values of the glottal opening that are very close to each other. When the glottal opening is below the first critical value, the voicing quotient is very high (the signal contains almost no frication noise), and is stable. Beyond the second critical value, when the glottal opening is sufficiently large, the voicing quotient vanishes, which results in a voiceless sound.

The different possible strategies used by the speaker for producing fricatives have also been experimentally observed. They consist in either making a fast transition through the unstable region, resulting in a very short mixed voiced/noisy fricative, or staying in the low noise region in order to favor the voicing and to stay in a stable regime. Our experiments revealed that the second strategy is systematically used by our subject when the fricative is longer.

For our research project, we are currently collecting additional data, in order to investigate the occurrence of each strategy depending on the fricative and the phonological context for a speaker, and also among speakers.

6. Acknowledgements

7. References

[1] A. Löfqvist, L. L. Koenig, and R. S. McGowan, "Vocal tract aerodynamics in /aCa/ utterances: Measurements," *Speech Communication*, no. 16, pp. 49–66, 1995.

[2] V. K. Mittal, B. Yegnanarayana, and P. Bhaskararao, "Study of the effects of vocal tract constriction on glottal vibration," *The Journal of the Acoustical Society of America*, vol. 136, no. 4, pp. 1932–1941, 2014.

[3] B. Cranen and J. Schroeter, "Modeling a leaky glottis," *Journal of Phonetics*, vol. 23, no. 1–2, pp. 165 – 177, 1995.

[4] 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.*, vol. 136(6), pp. 3262–3271, 2014.

[5] P. Birkholz, B. J. Kröger, and C. Neuschaefer-Rube, "Synthesis of breathy, normal, and pressed phonation using a two-mass model with a triangular glottis." in *Interspeech*, 2011, pp. 2681–2684.

[6] ——, "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*, vol. 370, 2011.

[7] D. H. Klatt and L. C. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am.*, vol. 87(2), pp. 820–857, 1990.

[8] 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)*, March 2016, pp. 5240–eli–5244.

[9] ——, "Acoustic impact of the glottal chink on the production of fricatives: A numerical study," Dec. 2016, working paper or preprint.

[10] 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*, vol. 91(5), pp. 2979–3000, 1992.

[11] J. Charles-Luce, "Word-final devoicing in german and the effects of phonetic and sentential contexts," *The Journal of the Acoustical Society of America*, vol. 77, no. S1, pp. S85–S85, 1985.

[12] O. Dmitrieva, A. Jongman, and J. Sereno, "Phonological neutralization by native and non-native speakers: The case of russian final devoicing," *Journal of phonetics*, vol. 38, no. 3, pp. 483–492, 2010.

[13] 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*, 2016, pp. 3156–3160.

[14] 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*, vol. 123(5), pp. 3738–3738, 2008.

[15] 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*, 2016.

[16] 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)*, Hong Kong, Hong Kong SAR China, Aug. 2011, pp. 1010–1013.