PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Glottal Opening Measurements in VCV and VCCV Sequences

Benjamin Elie[(1)], Angélique Amelot[(2)], Yves Laprie[(3)], Shinji Maeda[(2)]

[(1)]Loria, France, benjamin.elie@loria.fr
[(2)]LPP, France, angelique.amelot@gmail.com
[(3)]Loria, France, yves.laprie@loria.fr

Abstract

Many studies on speech acoustics and production use articulatory synthesis as a framework to investigate the relationship between articulatory gestures and acoustic features. Although supraglottal articulatory models are available, usually built from vocal tract imaging acquisitions, glottal gestures are commonly modeled with simple geometric primitives which do not necessarily reflect reality. This study is a first step towards the development of a database of realistic glottal gestures which will be used to design the glottal opening dynamics in articulatory synthesis paradigms. In the first part of this paper, we present experimental measurements of glottal opening dynamics in VCV and VCCV sequences uttered by real subjects, thanks to a specifically designed external photoglottographic device (ePGG). The corpus was designed to highlight the differences in glottis opening between fricatives and stops. The existence of different patterns of glottal opening is evidenced according to the class of the consonants. A numerical study is then used to show the influence of these patterns on the production of sounds and on the coarticulation.

Keywords: Speech, Glottis, Consonants

1 INTRODUCTION

Glottis opening is one of the parameters which define the articulation mode. It is also an essential parameter for exploring aerodynamic phenomena in the vocal tract with respect to the airflow. Recently, we have exploited the glottis opening data to explore the conditions of simultaneous existence of voicing and a frication noise for the voiced fricatives [1]. In particular, we have shown that they correspond to a production regime that is difficult to maintain, which explains the relative rarity of voiced fricatives in languages, and the final devoicing in some languages.

We therefore wanted to study glottis opening in a more systematic way, especially for the stops sounds. We acquired these glottis opening data using electroglottography, a recently developed technology and, before using this data in acoustic simulations, we first wanted to develop the acquisition methodology, and conduct evaluation of those data for several speakers.

Glottis opening is difficult to measure directly. The direct technique is fibroscopy, which involves passing a camera through the nostril to image the vocal folds from the top of the pharynx. This is an invasive technique, and one additional difficulty is that the lack of light only allows a frequency of about 30 Hz to be reached. Relative to the duration of a sound, i.e. between 50 and 100 ms for a vowel, this means that only 2 to 4 images of the vocal folds are obtained, and this is not enough to observe transitions between opening and closing of the glottis finely.

Fast imaging offers a much better temporal resolution (between 2000 and 4000 images per second) but the camera is fixed at the extremity of a rigid tube and it is therefore possible to visualize the vocal folds for vowels only. Whatever the imaging technique, it is impossible to measure the area at the glottis and only qualitative information is available.

Finally, transillumination technique [2] provides a far better sampling frequency because a photosensitive sensor is placed on the neck below the glottis, but is still invasive because it requires a source light to be introduced at the top of the pharynx through the nose.

Figure 1. Position of the emitting diods and photosensitive sensor. The subject presses each of the two diodes against her neck.

Electroglottography [3] is a non-invasive technique which provides a quantitative information about the contact between vocal folds, as well as fundamental frequency as a by-product. The higher the contact between the vocal folds, the stronger the conductance. Conversely, the conductance almost stabilizes at zero as soon as the vocal folds are separated. This means that the temporal evolution of the glottis area is not precisely known as soon as there is no longer contact. This is an important weakness since this area has a direct impact about the airflow in the vocal tract.

All these reasons motivated the development of the ElectroPhotoGlottoGraphy (EPGG) by Honda and Maeda [4], which provides information on the glottis opening area without being invasive. The principle consists in injecting light above the glottis and using a photosensitive sensor placed below the glottis to detect the light flow that depends directly on the surface of the opening. To avoid, or at least limit, the influence of visible light (natural or artificial) Honda and Maeda used infrared light. The system therefore consists of two infrared diodes to emit light above the vocal folds and a photosensitive sensor placed below the glottis. A measuring unit and a microphone complete the device. As can be seen in the photo, the system is much smaller than the one developed by Birkholz et al. [5].

1.1 First experiments and settings

The operating principle is quite simple, but it remains to verify the link between the amount of light recovered by the sensor and the surface of the glottis area, and to define an effective experimental protocol.

For the first point, Bouvet et al. [6] used a model of silicone vocal folds whose geometry is perfectly known. A source of pressure is used to vibrate these artificial vocal folds. The EPGG system was fixed on this model. The measurements show that the recovered light flow is an affine function of the amount of light injected at the entrance of the vocal folds. This validates this glottis opening measurement system.

For the second point we conducted a series of experiments on several speakers in order to know how EPGG can be used, and to know the influence of the anatomical characteristics of the speakers. The first observation is that the orientation of the light-emitting diodes has a very important impact on the measured light flow. Their orientation has thus to be adjusted with a good precision.

The second observation is that the vertical movement of the larynx has an important influence on the recovered light flow since the relative position of the diodes and the sensor with respect to the glottis opening changes. The third observation is that the measurements depend strongly on the speaker, probably because of the variety of tissues traversed by the light, and of course their thickness. We thus investigated a number of settings

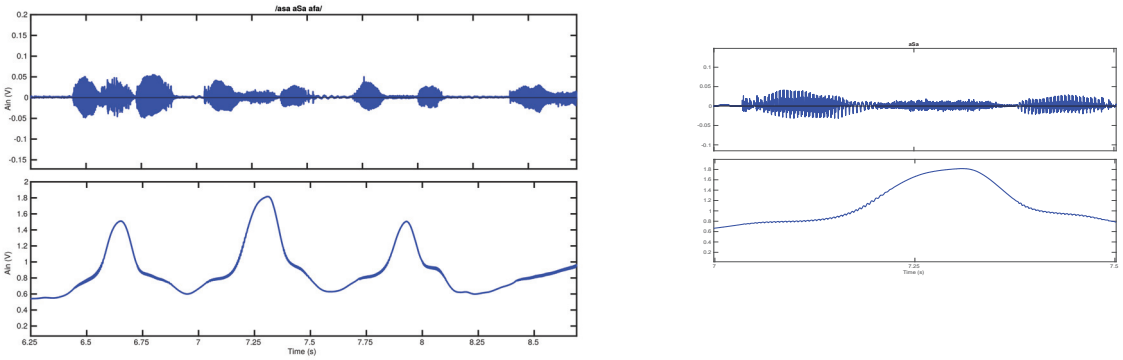

Figure 2. EPGG signal (bottom) and acoustic signal (top) for /asa aʃa afa/ and zoom on /aʃa/

to optimize the quality of measurements. One of the difficulties is to ensure good contact between the light emitting diodes and the neck. Initially we used an elastic strap band but to obtain a good contact it is necessary that the strap is very tight which is not comfortable for the subject. Finally, we ask the subject to hold the diodes himself or herself (see Fig. 1) which has the additional advantage of partially blocking the movements of the larynx, and therefore of obtaining a signal with a more stable baseline.

Figure 2 illustrates an EPGG acquisition for the 3 French unvoiced fricatives /s ʃ f/. It can be noted that the volume of light recovered is not zero when the glottis is closed, nor is the baseline constant. As also noted by Sawashima and Hirose [7], it seems that there are still some vibrations after the glottis opens. But contrary to their observations, there is still a residual voicing in the signal.

1.2 Corpus design

The first experiments we carried out concerned the production of voiced and unvoiced fricatives [1] with respect to the glottis opening strategies. The measurements made on this occasion showed that the glottis opening is stronger for unvoiced fricatives than for all other sounds, except for breathing. Since the amount of light depends on the position of the larynx, we looked for a way to normalize the measurements. Glottal opening for /asa/ presents the double advantage of being sufficiently large without being the greatest opening, and stable enough. During the construction of the corpus we therefore introduced /asa/ as a normalization sequence before and after each item. Each utterance is thus of the following form: /asa/ item /asa/ The corpus is therefore made up of the following items:

- V*C*V where V is a cardinal vowel and C belongs to {p t k b d g f s ʃ v z ʒ l m n ʁ},
- aC_1C_2a where V is a cardinal vowel, C_1 belongs to {b d g} and C_2 to {l ʁ},
- as*C*a where *C* belongs to {b d g p t k},
- geminated stops in /pap papa/ ("pape papa"), /pat tatue/ ("patte tatouée"), /sak kaʁe/ ("sac carré"), /kʁab bagaʁœʁ/ ("crabe bagarreur"), /pad dat/ ("pas de date"), /blag gaʁãti/ ("blague garantie"),
- 4 sentences of variable length.

This small corpus covers all the consonants, some of the most frequent clusters in French, especially those with /ʁ/, geminated stops and some sentences. The corpus has been recorded by 3 female and male French speakers.

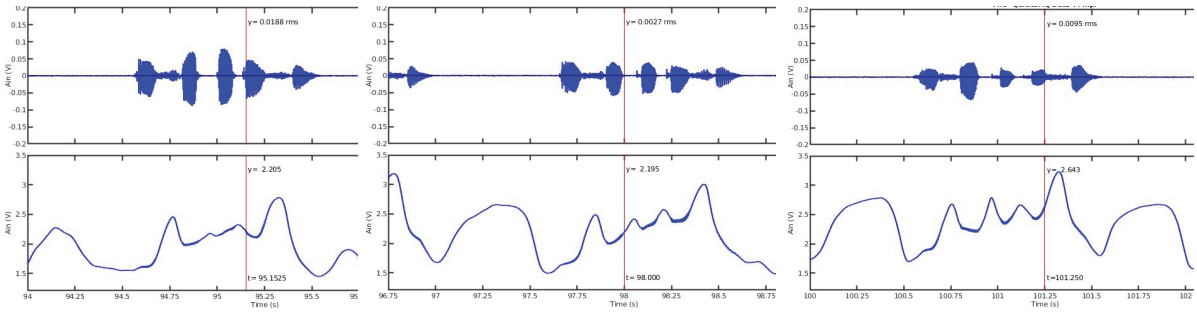

Figure 3. EPGG signal for /asa papa sa/, /asa kaka sa/ and /asa titi sa/

2 Analysis of glottis opening for stops

Figure 3 shows that the glottis opening is considerably smaller for unvoiced stops than unvoiced fricatives. The complete data confirms this trend. The first explanation is that the production of fricatives requires maintaining a turbulent flow throughout the duration of the fricative, and consequently a large opening to ensure a sufficient air flow. On the contrary, the production of unvoiced stops only requires stopping the vibration of the vocal folds and filling the cavity behind the constriction in order to achieve an overpressure with respect to atmospheric pressure.

The peak of the opening is reached approximately in the middle of the segment formed by the closure and burst, i.e. the moment during which there is no more voicing. On the 3 examples /papa kaka titi/ it can be noted that the opening is larger for /kaka/ than /papa/, and that it is bigger for /titi/ than /kaka/. This increase of the opening coincides with the existence of an increasingly intense frication noise following the release burst.

Overall, the same trend can be observed for /u/ and /i/ compared to /a/, i.e. a larger opening during the closure. From the point of view of the larger glottis opening for the vowels /i/ and /u/, the tongue has to anticipate the position it will have after the occlusion is released. The cavity volume behind the constriction is therefore larger, and a larger opening is required to bring a sufficient amount of air. Compared to /t/ and /k/, /p/ does not impose any constraint on the tongue, which can therefore fully anticipate its position for the following vowel. Since the vowel /i/, and to a lesser extent /u/, are characterized by a back cavity of a larger volume than /a/ it is therefore necessary to maintain a larger opening to allow enough air to enter. The volume of air behind the constriction, and the vocal tract shape also explains the duration of the global burst, i.e. from the transient noise corresponding to the release of the constriction to the first voiced period.

Overall, the larger the cavity behind the constriction and the smaller the area in front of the vocal tract for the next vowel, i. e. /i/ and /u/, the longer the burst duration for the overpressure to vanish. On the other hand, the back cavity of /a/ is very narrow and therefore not very voluminous, which leads to a minimal burst duration, especially for /p/ since the tongue can anticipate its back position very early.

Although the larynx moves over time, Figure 2 shows that the glottis opening is absent or very small for the voiced stops. However, there is a small opening in the case of /didi/ which gives rise to a slight frication noise in high frequency. The rest of the data confirms this trend.

3 Analysis of glottis opening for stops

We also acquired data for consonant clusters, especially when the second consonant is /ʁ/. In the case of an unvoiced stop the presence of /ʁ/ results in a glottis opening that is much larger than that of the unvoiced stop alone in order to produce a frication noise after the burst. The opening is almost zero when /ʁ/ follows a voiced stop (see Figure 5). The difference between /kʁa/ and /gʁa/ is therefore very strong since there is no glottis opening for /gʁa/. It turns out that the voicing mode is entirely determined by the consonant preceding /K/.

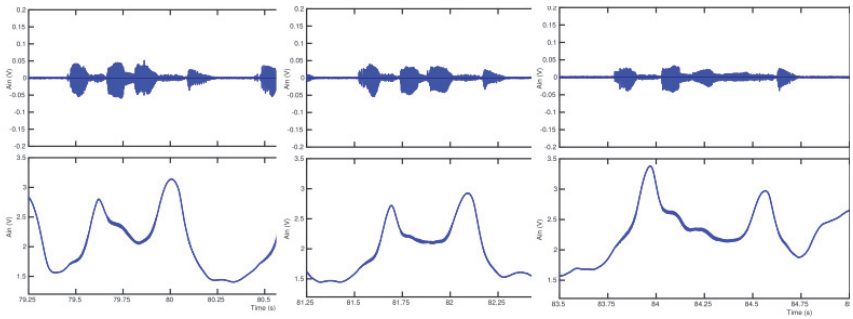

Figure 4. EPGG signal for /asabasa/, /asagasa/ and /asa didi sa/

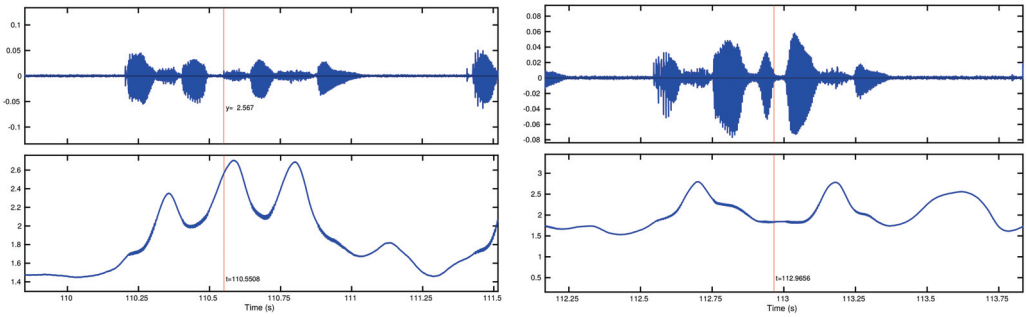

Figure 5. EPGG signal for /asa kʁa sa/ and /asa gʁa sa/

Despite this absence of opening, the vocal folds stop vibrating during /ʁ/ in the case of /gʁa/ but not in the other voiced cases /bʁa/ and /dʁa/, either because the constriction is almost complete at the constriction level /gʁa/, or because the volume of the cavity before the constriction is sufficient for /bʁa/ and /dʁa/ to maintain a pressure difference that allows the vibration of the vocal folds.

4 CONCLUDING REMARKS

Apart from the consonants that motivated this work, these EPGG data illustrate the speaker strategy at the level of a sentence. The speaker begins with a breath that results in a long glottis opening. He/she then closes it completely just before starting a voiced sound (about 100 ms) in order to reach a sufficient subglottal pressure. There is also an almost complete closure when the sentence begins with an unvoiced fricative but the closing gesture does not completely finish until a new opening gesture corresponding to the fricative starts. We used those data to design the algorithm that sets the opening to the glottis for our copy synthesis experiments [8].

ACKNOWLEDGEMENTS

This work is supported by the ANR grant “ArtSpeech” (2015-2019).

REFERENCES

[1] B. Elie and Y. Laprie, “Acoustic impact of the gradual glottal abduction on the production of fricatives: A numerical study”, *Journal of the Acoustical Society of America*, vol. 142, no. 3, pp. 1303–1317, Sep. 2017. DOI: 10.1121/1.5000232. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01423206.

[2] A. Löfqvist and H. Yoshioka, "Laryngeal activity in Swedish obstruent clusters", *Journal of the Acoustical Society of America*, vol. 68, no. 3, pp. 792–801, 1980. DOI: 10.1121/1.384774.

[3] E. Abberton and A. Fourcin, in *Instrumental Clinical Phonetics*, M. J. Ball and C. Code, Eds. London, Apr. 2008, ch. Electrolaryngography, pp. 119–148, ISBN: 9780470699119. DOI: 10.1002/9780470699119.ch5.

[4] K. Honda and S. Maeda, "Glottal-opening and airflow pattern during production of voiceless fricatives: A new non-invasive instrumentation.", *Journal of the Acoustical Society of America*, vol. 123, no. 5, p. 3788, 2008.

[5] E. Suthau, P. Birkholz, A. Mainka, and A. P. Simpson, "Non-invasive photoglottography for use in the lab and the field", in *proc. of Speech Communication; 12. ITG Symposium*, Paderborn, 2016.

[6] A. Bouvet, A. V. Hirtum, X. Pelorson, S. Maeda, K. Honda, and A. Amelot, "Calibration of external lighting and sensing photoglottograph", in *Proc. 10th Int. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)*, Firenze, Italy, 2017.

[7] M. Sawashima and H. Hirose, "Laryngeal gestures in speech production", in *The Production of Speech*, P. F. MacNeilage, Ed. New York: Springer Verlag, 1983, ch. 2, pp. 11–38.

[8] B. Elie and Y. Laprie, "Copy synthesis of running speech based on vocal tract imaging and audio recording", in *22nd International Congress on Acoustics (ICA)*, Buenos Aires, Argentina, Sep. 2016. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01372310.