



HAL
open science

VALIDATION OF A NONINVASIVE SYSTEM TO OBSERVE GLOTTAL OPENING AND CLOSING: EXTERNAL PHOTOGLOTTOGRAPH (ePGG)

Angelique Amelot, Darshan Sathiyarayanan, Shinji Maeda, Kyioshi Honda,
Lise Crevier-Buchman

► **To cite this version:**

Angelique Amelot, Darshan Sathiyarayanan, Shinji Maeda, Kyioshi Honda, Lise Crevier-Buchman. VALIDATION OF A NONINVASIVE SYSTEM TO OBSERVE GLOTTAL OPENING AND CLOSING: EXTERNAL PHOTOGLOTTOGRAPH (ePGG). 11th International Conference on Voice Physiology & Biomechanics, Aug 2018, East Lansing, United States. halshs-02510216

HAL Id: halshs-02510216

<https://shs.hal.science/halshs-02510216>

Submitted on 17 Mar 2020

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

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

VALIDATION OF A NONINVASIVE SYSTEM TO OBSERVE GLOTTAL OPENING AND CLOSING: EXTERNAL PHOTOGLOTTOGRAPH (ePGG)

Angélique Amelot¹, Darshan Sathiyarayanan¹, Shinji Maeda¹, Kyoshi Honda², Lise Crevier-Buchman^{1,3}

¹ Université Sorbonne-Nouvelle Paris 3 - Laboratoire de Phonétique et de Phonologie (LPP)-UMR7018, Paris, France
² School of Computer Science and Technology, Tianjin, China
³ Hôpital Européen Georges Pompidou, Paris, France

Keywords: External PhotoGlottoGraph; ElectroGlottoGraphy; Video-endoscopy; Open Glottis

INTRODUCTION

Investigation of vocal fold movements required invasive instrumentations like Electromyography (EMG), Video-endoscopy or Transillumination. External lighting and sensing PGG (ePGG), recently developed at LPP [1] [2], relies on transillumination technique. It consists of illuminating the cavity above the glottis through the exterior neck skin surface with infrared LED (IR-LED) placed above the glottis and recording medially under the glottis variations of light intensity modulated by vibrations of or by abduction/adduction of the vocal folds with the 20 kHz sample rate. Compared with previous photoglottographic techniques, this device has the definite advantage that both IR-LED and the photosensor are positioned on the speaker’s exterior neck surface (see Fig. 1). Thus, the data can be noninvasively acquired.

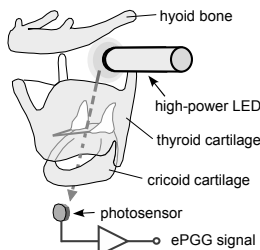


Fig. 1: External lighting and sensing photoglottography (ePGG) system: an IR-LED is placed on the skin on the neck, between hyoid bone and thyroid cartilage.

The ePGG system was recently tested on two mechanical larynx replicas [3]. Our current study is to validate and calibrate the outputs of ePGG on human speakers. We compared ePGG data with data from high-speed and standard video-endoscopies, and from electroglottography (EGG). The goal of the study is i) to validate ePGG signals by a multimode analysis and ii) to understand the relationships between the vocal folds movements and the degree of opening of the glottis indicated by ePGG and by EGG.

METHODS

Subjects

Two healthy subjects, one female (speaker 1) and one male (speaker 2) have participated in the data acquisition at the

European Hospital of George Pompidou at Paris with the help of a medical doctor.

Instrumentations, corpus, and analysis

a) ePGG and EGG (EG2PCX2 Glottal Enterprises) signals, and high-speed video-endoscopy (Richard Wolf GmbH) images (4000 images/s) are simultaneously recorded as shown in Fig. 2. Audio signal is recorded by a microphone integrated in the video-endoscopy system. An additional external microphone is used. Its signal and those from ePGG and EGG are saved in a PC via a multichannel USB data card. These two audio signals allow us to make the time-alignment of the video-images sequence, and ePGG and EGG signals. For the comparison, frame-by-frame variations of glottal aperture are detected with the help of a MATLAB script resulting in a glottal aperture signal (GA). Because of the requirement of strong light for the high-speed operation, we are obliged to employ a rigid straight endoscope (see Fig 2), which limit the corpus to an isolated stationary high vowel [i]. The point of interest, therefore, becomes the comparison of glottal pulse shapes during the vowel [i]. By the way, in both this session and the next one b), the white LED light coming from the endoscope is used as the light-source for ePGG.

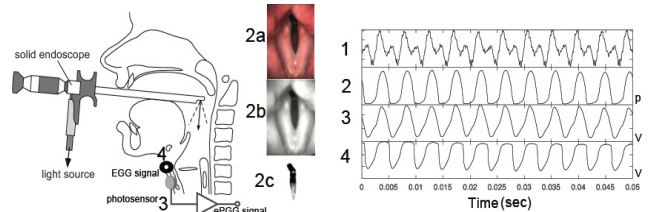


Fig. 2: Setup to acquire simultaneously (1) Acoustic signal in Volts, (2) images from high-speed video-endoscopy (4000 images/s): (2a) Original image, (2b) Filtered image, (2c) Detected GA and (2), GA, in pixels (3) ePGG signal in Volts, and (4) EGG in Volts.

b) In this session we focus our attention on slow abduction/adduction of vocal folds related to the production of vowel-consonant sequences. In such case, the high-speed video-endoscopy can be replaced by a standard video-endoscopy with the speed of 25 images/s. We use a flexible fiber-endoscopy, which is inserted through the nasal cavity. This, therefore, allows a speaker to utter any vowel-consonant sequence. The corpus in this

session consists of [VCV], [CVCV], and [CVC], where C=[s, z, p, b] and V=[a, i, u]. Each target word is repeated three times in the carrier sentence “Je dis ___ trois fois” (“I say ___ three times”). So, if a target word is [sas], then the utterance becomes “Je dis sas sas sas trois fois”.

RESULTS

a) ePGG, High-Speed video-endoscopy and EGG

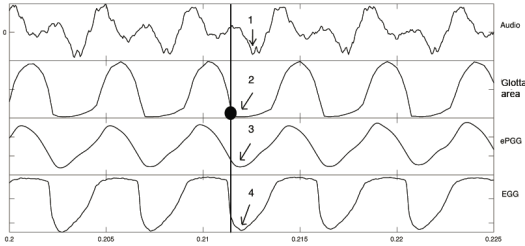


Fig. 3: Extract of the production of [i] for the speaker 1, From top to bottom: Acoustic signal, GA, ePGG, and EGG in which the polarity is inverted.

At the black dot (pointed by arrow 2) in Fig. 3, GA exhibits a strong discontinuity as if the closing vocal folds are suddenly blocked, creating a glottal closure. This occurs in every glottal cycles. The closure seems to be sustained for a while before a rapid opening of the glottis. According to the acoustic theory, the vocal tract must be excited at the closure onset by a negative pulse. Indeed, the audio signal indicates a sudden shoot (pointed by the arrow 1), which occurs about .5ms after the glottal closure onset. This .5ms delay must be the sound propagation time from the glottis to the microphone (about 20 cm), reinforcing the notion of the glottal closure onset and the closed glottal quotient.

However, we don't see those at all in ePGG and in EGG. In Fig. 3, both ePGG and EGG vary continuously exhibiting a turning point, at the arrow 3 and 4, respectively, from glottal closing to opening. We hypothesize, during the closure, that the tissues inside vocal fold are continuously deforming which is invisible to the endoscopic imaging but “visible” to ePGG and to EGG. The deformation of the tissues causes variations in the light transmission across closed glottis (ePGG sensitive) and that in the surface contact area of the vocal folds (EGG sensitive).

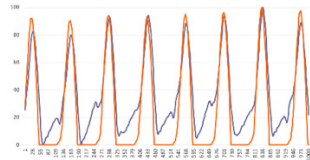


Fig. 4: Extract of normalizations of ePGG data and GA during the production of [i] for the speaker 2 (blue line is the ePGG data red line is GA from high-speed video-endoscopy (2000 images/sec).

The Fig. 4 shows how ePGG can be comparable to GA. In the ePGG curve we can recognize the closure onset as well as the closed glottal quotient, which are absent in the ePGG in Fig.3. The possible explanation is the gender difference.

Presumably on the one hand, the male subject has the thick and dense vocal-fold tissues and tight closure of the VF that tend to suppress the light going through the tissues. On the other hand, the female speaker has the thin and low dense vocal-fold tissues that leads to the result already described above.

b) We observe that ePGG can respond not only to the oscillation of the vocal-fold vibrations but also to abduction/adduction of the glottis for the consonant production. For example, during the production of an unvoiced fricative, ePGG can provide interpretable data, but EGG cannot as shown in Fig.5.

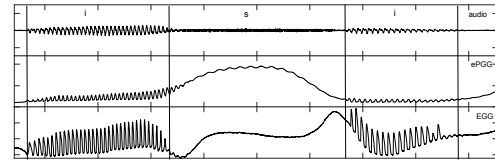


Fig. 5: From top to bottom, Audio signal [isi], ePGG, and EGG speaker

CONCLUDING REMARKS

We feel that ePGG is a versatile method for observing both glottal vibrations during voicing sounds and large glottal abduction/adduction movements in transitions between a consonant and a vowel. We could not identify glottal closure quotients on the ePGG. This appears to be because the light can go through thin vocal-fold tissues during the closure. EGG has shown a similar response in the same situation. In such a case, we need to resort to an invasive high-speed video-endoscope. However, simple and clean looking data obtained could be due to its blindness to the deformation of internal vocal-fold tissues behind the “closure”.

ACKNOWLEDGMENTS

This work was partly supported by two public grants overseen by the French National Research Agency (ANR) as part of the program “Investissements d’Avenir” (ANR-10-LABX-0083) and ArtSpeech project (ANR-15-CE23-0024).

REFERENCES

- [1] K. Honda and S. Maeda, “Glottal-opening and airflow pattern during production of voiceless fricatives: {A} new non-invasive instrumentation,” *J. Acoust. Soc. Am.*, vol. 123, no. 5, p. 3788, 2008.
- [2] J. Vaissière, K. Honda, A. Amelot, S. Maeda, and L. Crevier-Buchman, “Multisensor platform for speech physiology research in a phonetics laboratory,” *J. Phonetic Soc. Japan*, vol. 14, no. 2, pp. 65–78, Sep. 2010.
- [3] A. Bouvet, A. Van Hirtum, X. Pelorson, S. Maeda, K. Honda, and A. Amelot, “Calibration of external lighting and sensing photoglottograph,” Dec. 2017.