Calibration of external lighting and sensing photoglottograph

Anne Bouvet, Annemie van Hirtum, Xavier Pelorson, Shinji Maeda, Kiyoshi Honda, Angelique Amelot

CALIBRATION OF EXTERNAL LIGHTING AND SENSING PHOTOGLOTTOGRAPH

A. Bouvet[1], A. Van Hirtum[1], X. Pelorson[1], S. Maeda[2], K. Honda[2], A. Amelot[2]

[1] CNRS UMR216, Gipsa-lab, Université Grenoble-Alpes, Grenoble, France
[2] Laboratoire de Phonétique et Phonologie (LPP), Université Paris 3, Paris, France
anne.bouvet@grenoble-inp.fr

Abstract: **Observation and measurement of vocal folds vibration and glottal opening during speech requires techniques as little invasive as possible for the subject. The LPP has developed the External Photoglottograph (ePGG) system. It consists of illuminating the glottis through the neck skin with an infrared light and recording light variation intensity modulated by glottal movement with a photodiode placed across the larynx. The system is tested on two mechanical larynx replicas. The first one consists of two rigid half cylinders in forced oscillation controlled by a step motor. The second one is flow driven and uses latex tubes filled with water in order to reproduce vocal folds self-oscillation. Time-varying glottal area is measured accurately for both replicas. Experimental results are compared to ePGG recordings in order to assess the correlation between area measurements and ePGG signal. This characterization is used to propose a calibration of the glottal opening as a function of parameters affecting the ePGG signal (distance, angle, skin, tissue, system setting, etc.).**

Keywords: **ePGG, glottal area, vocal folds auto-oscillation, non-invasive measurement, mechanical replica**

I. INTRODUCTION

Observation and measurement of the glottal opening and vocal folds vibration during speech requires the measurement of small displacements (of order of a millimeter) at a high sampling rate. Further, the technique must be as little invasive as possible in order not to prevent from normal speech and/or articulation.

External lighting and sensing ElectroPhotoGlottoGraph (ePGG), developed at LPP [4], relies on transillumination technique (Fig. 1), which consists of illuminating the glottis through the neck skin with infrared light and recording the variation of light intensity modulated by the vibration of the vocal folds with a sample rate of 20 kHz. In contrast to common laryngeal illumination techniques no visible light is used to light the glottal area. Instead light in the near infrared (IR) spectral range is used since wavelengths in this range (700-1000 nm) are reported to transilluminate large sections (many centimeters) of human tissue [1-3]. Compared with other photoglottographic techniques, this device has the advantage that both the light source (IR) and the light sensor (S) are positioned on the speaker's neck as illustrated in Fig. 1. Non-invasive, it allows to perform measurements without coercion for the speaker and can be used without medical assistance.

The objective of this paper is therefore to assess ePGG measurements during auto-oscillation of a mechanical glottal replica for which the varying glottal area can be accurately quantified during oscillation. The influence of all parameters potentially affecting the outcome of an ePGG measurement is studied in a controlled and repeatable way in order to propose a calibration procedure or/and provide objective guidelines for ePGG usage.

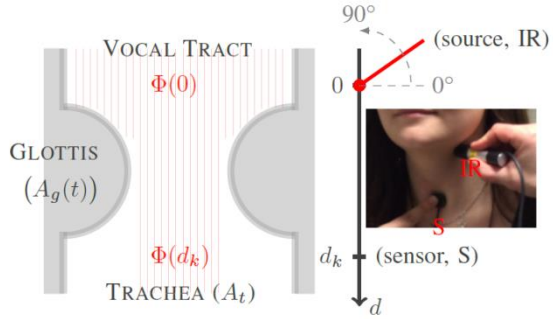
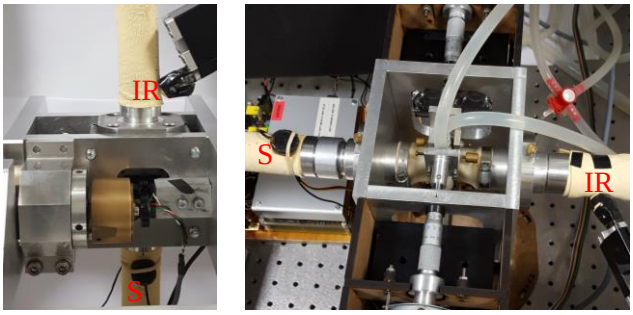


Fig. 1: Glottal transillumination principle.

II. METHODS

In order to realize this objective, experiments are performed using the following experimental setups.

Firstly, to characterize the ePGG outcome, the system (source and sensor) is mounted on an optic bench, so that the emitter and receiver system can be characterized (positioning, settings) and environmental measurement disturbances can be identified. Next, a uniform Plexiglas tube (diameter 25 mm) is used to represent the human trachea and pharynx. The tube is covered with layers of lamb leather to simulate the absorption of light by the human skin. The impact of leather thickness and IR source positioning (angle, distance to sensor) are assessed.



a) **b)**

Fig. 2: ePGG system source/emitter IR and optic sensor/receiver (S) placed at a) motor-driven rigid replica, b) flow-driven auto-oscillating replica.

Secondly, the ePGG system is placed on a mechanical glottal replica [6] to which two tubes are added representing the trachea and the pharynx as illustrated in Fig. 2.a. The "vocal folds" consist of two rigid half-cylinders, one of which is forced into motion by an eccentric motor. A second glottal replica is used (Fig. 2.b), the "vocal folds" consist then of latex tubes filled with water so that the replica is able to self-oscillate in interaction with an airflow [10,11]. In both cases, the initial glottal area prior to oscillation can be imposed, so that steady geometrical configurations can be systematically quantified. Usage of both replicas ensures that slow (rigid motor-driven, frequency range < 15 Hz, aperture < 2 mm) as well as fast (auto-oscillation, frequency range ~100-300 Hz, aperture < 2 mm) vocal folds displacements can be studied accurately whereas the glottal area can be measured either using an optical sensor (type OPB700) for the first replica or using a high-speed camera (Motion BLITZ Eosens Cube 7) at 525 frames per second and standard image processing techniques validated on the auto-oscillating replica [4]. These replicas allow to study parameters in the range relevant to orders of magnitudes observed on human speakers [5-9].

III. EPGG SIGNAL MODEL

The ePGG system is assessed on mechanical replicas. Since experimental setups are equipped to measure the glottal area, the relationship between ePGG signal and glottal area can be studied on these replicas as a function of parameters potentially affecting the ePGG signal (Fig. 1). In the following, the experimental ePGG signal characterization is presented firstly for static geometrical configurations with constant glottal area and secondly for dynamic geometrical configurations with a time-varying glottal area. During all experiments, the room temperature was 21 ± 1 °C.

A. Static glottal area

Firstly, the effect of source-sensor distance *d* (Fig. 1) on the ePGG signal is sought. The ePGG system is positioned on the mechanical airway replica with constant area (A_u = 491 mm^2). The source-sensor distance *d*, is systematically varied in the range 2 mm ≤ *d* ≤ 200 mm and the orientation angle is 27°.

In addition, in order to mimic the influence of wall tissue thickness, measurements are performed adding two (thickness 1.4 mm) or three leather layers (thickness 2.1 mm). Measured mean ePGG signals are plotted in Fig. 3. The ePGG signal decreases with *d* regardless of wall thickness. Linear fitting of measured ePGG signals in the range *d* ≤ 100 mm (appropriate for human subjects) and in the range *d* ≥ 100 mm (appropriate for mechanical replicas), results in R^2 ≥ 98.9%. Consequently, a first order linear approximation can be used to characterize the evolution of ePGG signal with source-sensor distance *d*, while the negative slope depends on wall absorption (thickness) and distance *d*. All remaining experiments are done with 2 layers (thickness 1.4 mm).

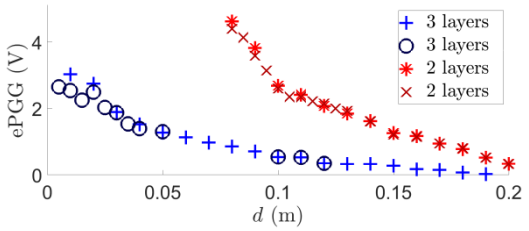


Fig. 3: Mean ePGG signal as a function of source-sensor distance d for the airway replica with 2 and 3 leather layers.

Secondly, static geometrical configurations are done to determine the effect of the source orientation angle in the mid-coronal plane (Fig. 1) on the ePGG signal. The ePGG system is again positioned on the uniform mechanical airway replica, i.e. in absence of a glottal constriction (no glottal replica). The source orientation angle is systematically varied from 0° up to 40° and the source-sensor distance is held constant to *d* = 100 mm. Measured mean ePGG signals are plotted in Fig. 4. For orientation angles up to about 15°, the ePGG signal is minimum and only marginally (< 0.3 V) affected by the orientation angle due to the source (IR) half beam angle of 22.5 ± 2.5°. Further increasing the orientation angle above 15° results in a linear (R^2 = 98.1%) increase of the mean ePGG signal. All remaining experiments are done for orientation angle 27°.

Thirdly, static geometrical configurations are performed to determine the effect of glottal area on the ePGG signal (Fig. 1). The rigid and deformable vocal

folds mechanical replicas are used to which a uniform mechanical airway wall replica is attached at each end.

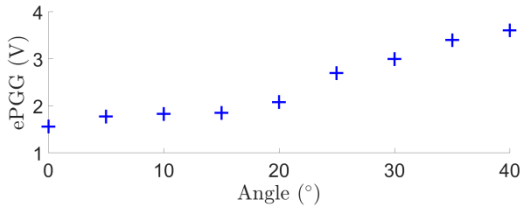


Fig. 4: Mean ePGG signal as a function of source orientation.

Source and sensor are positioned on each airway replica (trachea end and vocal tract end) so that the glottal area of the mechanical replica corresponds to the minimum area of the channel portion between source and sensor. The (minimum) source sensor distance is d = 150 mm for the rigid and d = 257 mm for the deformable glottal replica. Glottal area A_g is varied in the range 0-55 mm^2 (rigid) and 20-100 mm^2 (deformable). Measured mean ePGG signals are plotted in Fig. 5. The ePGG signal increases linearly with A_g for both the rigid (R^2 = 99.2%) and the deformable (R^2 = 98.2%) replica. So that ePGG signal and glottal area relate well using a linear approximation. Note that, in general, differences in slope and offset can occur due to 1) positioning of the ePGG system (source-sensor distance d, orientation angle) and 2) channel wall properties affecting absorption (thickness, material, etc.). In the next section, time-varying glottal areas are considered.

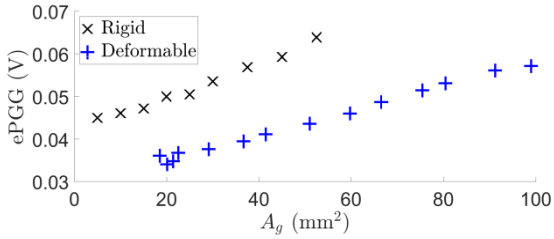
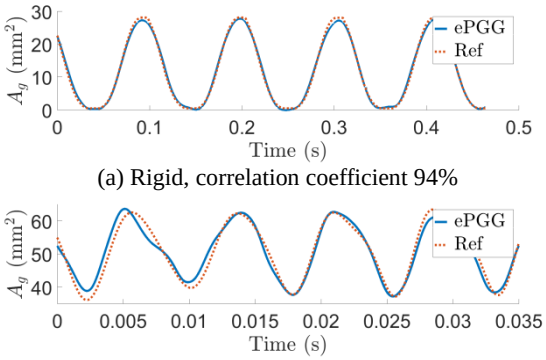


Fig. 5: Mean ePGG signal as a function of static glottal area A_g for a rigid and a deformable mechanical glottal replica.

B. Time-varying glottal area

The correlation between the time-varying ePGG signal and the time-varying glottal area is quantified for the motor driven rigid (oscillation frequency $f_0 \in \{2, 5, 10, 12\}$ Hz, $0 \leq Ag \leq 40$ mm^2, d = 150 mm) and the flow-driven deformable (fundamental frequency $f_0 \in \{113; 125; 129; 131\}$ Hz for mean upstream glottal pressures $Pu \in \{500, 570, 720, 840\}$ Pa, $20 \leq Ag \leq 100$ mm^2, d = 257 mm) mechanical replica. Typical examples of correlated time signals for slow (rigid) and fast (deformable) vocal folds displacement are plotted in Fig. 6. Correlation coefficients between ePGG signals and glottal area $A_g(t)$ yield > 90% for the rigid and > 85% for the deformable glottal folds replica.



(a) Rigid, correlation coefficient 94%

(b) Deformable, correlation coefficient 90%

Fig. 6: Illustration of correlated time signals of scaled ePGG (full line, ePGG) and glottal area Ag (dashed line, Ref): a) rigid replica at f_0 = 10 Hz, b) deformable replica at f_0 = 129 Hz.

Consequently, the ePGG signal and glottal area are correlated at all times during the oscillation. In the following section, it is aimed to model the relationship between ePGG signal and glottal area $A_g(t)$ accounting for the different variables affecting the ePGG signal.

IV. EPGG SIGNAL MODELING

In Section III, it was shown that the ePGG signal is mainly determined by 1) the source-sensor distance, 2) the minimum area of the channel portion between the source and the sensor and 3) the measurement condition determined by the combination of wall properties (e.g. absorption), environment (*e.g.* light) and ePGG system settings and positioning (*e.g.* orientation angle). In the following, an ePGG signal model is proposed accounting for each of these factors. Next, its parameters estimation and initialization is outlined. Finally, the sought relationship between ePGG signal and glottal area $A_g(t)$ is discussed.

Following the transillumination principle shown in Fig. 1, ePGG sensor voltage U is proportional to light intensity I at distance d_k from the light source,

$$U(d_k) \propto I(d_k) \tag{1}$$

Transmitted light intensity $I(d_k)$ at sensor position d_k is then expressed using light flux ϕ as

$$I(d_k) = \iint_{A_{min}} \phi(d_k) dA, \tag{2}$$

where

$$A_{min}(d_k) = \min_{d \in [0,d_k]}(A(d)) \quad (3)$$

is the minimum area encountered by the transmitted light flux between the source position and the sensor position. Furthermore, in Section III was shown that the dependence on d and A_{min} can be described using a first order linear approximation. Consequently, light flux $\phi(d) > 0$ can be approximated by model $\phi_m(d)$ defined as

$$\phi_m(d) = \alpha_d d + \beta_d, \quad (4)$$

with slope $\alpha_d < 0$ and offset $\beta_d > 0$ (see Section IV). From (2), $I(d_k)$ is now modeled as

$$\begin{aligned} I_m(d_k) &= A_{min}(d_k) \cdot \phi(d_k), \\ &= A_{min}(d_k) \cdot (\alpha_d d_k + \beta_d). \end{aligned} \quad (5)$$

Inserting (5) in (1) results in modeling the ePGG voltage $U(d_k)$ as $U_m(d_k)$ given by

$$\begin{aligned} U_m(d_k) &= \gamma\big((\alpha_d d_k + \beta_d) \cdot A_{min}(d_k)\big) + \eta \\ &= (\alpha_v d_k + \beta_d) \cdot A_{min}(d_k) + \eta. \end{aligned} \quad (6)$$

where $\eta > 0$ is the remaining signal measured for $A_{min}(d_k) = 0$ and $\gamma > 0$ is the scaling factor of (1). For sake of simplicity, let us denote $\alpha_v = \gamma\alpha_d < 0$ and $\beta_v = \gamma\beta_d > 0$. It is worth noting that $A_{min} = 0$ corresponds to glottal closure so that no direct light is transmitted. Therefore, η is independent of d and A_{min} so that η reflects solely the measurement condition. Considering now a time-variation of the glottal opening, model (6) can be directly extended as

$$U_m(d_k, t) = (\alpha_v d_k + \beta_v) \cdot A_{min}(d_k, t) + \eta. \quad (7)$$

Consequently using model (7), extracting area $A_{min}(d_k,t)$ from measured ePGG signals $U(d_k,,t)$ reduces to a parameter estimation problem which is solved using the iterative gradient descent method with appropriate initialization [12].

V. CONCLUSION

An experimental systematic characterization of the ePGG system has been assessed. Main parameters affecting the ePGG outcome are identified. Based on this result, glottal area modulation is studied experimentally on a motor-driven and on a flow driven mechanical larynx replica. It is seen that the correlation between ePGG signal and area measurement yields more than 85%. Next, a calibration method for the ePGG system is proposed. Further research focuses on a systematic validation of the proposed method.

ACKNOWLEDGMENT

Partly funded by ArtSpeech (ANR-15-CE23-0024).

REFERENCES

[1] D. Delpy and M. Cope, "Quantification in tissue near-infrared spectroscopy," *Phil. Trans. R. Soc. Lond. B*, vol. 352, pp. 649–659, 1997.

[2] T. Lister, P. Wright, and P. Chappell, "Optical properties of human skin," *J. Biomed. Optics*, vol. 17, pp. 1–15, 2012.

[3] S. Jacques, "Optical properties of biological tissues: a review," *Phys. Med. Biol.*, vol. 58, pp. R37–R61, 2013.

[4] H. Kim, K. Honda, and S. Maeda, "ePGG, airflow and acoustic data on glottal opening in Korean plosives," *Proc. 18th Int. Conf. of Phonetics Sciences (ICPhS)*, Glasgow, UK, 2015, p. 4.

[5] D. O'Shaughnessy, Speech Communication Human and Machine. *Addison-Wesley Publishing Company*, 1987.

[6] J. Cisonni, A. Van Hirtum, X. Pelorson, and J. Willems, "Theoretical simulation and experimental validation of inverse quasi one-dimensional steady and unsteady glottal flow models," *J. Acous. Soc. Am.*, vol. 124, pp. 535–545, 2008.

[7] J. Burnard West, Bio-engineering aspects of the lung, Volume 3, *M. Dekker, Ed. Marcel Dekker, INC.*, 1977.

[8] T. Brancatisano, P. Collett, and L. Engel, "Respiratory movements of the vocal cords," *J. Appl. Physiol.*, vol. 54, pp. 1269–1276, 1983.

[9] A. Scheinherr, L. Bailly, O. Boiron, A. Lagier, T. Legou, M. Pichelin, G. Caillebotte, and A. Giovanni, "Realistic glottal motion and airflow rate during human breathing," *Med. Eng. Physics*, vol. 37, pp. 829–839, 2015.

[10] A. Van Hirtum and X. Pelorson, "High-speed imaging to study an auto-oscillating vocal fold replica for different initial conditions," *Int. J. Applied Mech.*, Accepted.

[11] J. Haas, P. Luizard, X. Pelorson, and J. Willems, "Study of the effect of a moderate asymmetry on a replica of the vocal folds," *Acta Acustica*, vol. 102, pp. 230–239, 2016.

[12] J. Nocedal and J. Wright, Numerical Optimization. Springer, 2006.