Biomedical Signal Processing and Control 51 (2019) 318–327

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

External lighting and sensing photoglottography: Characterization and MSePGG algorithm

A. Bouvet[a], A. Amelot[b], X. Pelorson[a], S. Maeda[b], A. Van Hirtum[a]

[a] LEGI, CNRS UMR 5519, Univ. Grenoble Alpes, France
[b] LPP, CNRS UMR 7018, Univ. Paris 3, France

ARTICLE INFO

Article history:
Received 1 June 2018
Received in revised form
22 November 2018
Accepted 4 January 2019

Keywords:
Multi-Signal-ePGG algorithm
Glottal area

ABSTRACT

Continuous observation of the time-varying glottal area lacks a direct, quantitative, non-invasive measurement method despite its relevance to study breathing, speech production, swallowing, *etc.* External photoglottography (ePGG) relies on external glottal transillumination and sensing, it is therefore suitable for non-invasive and continuous observation. Nevertheless, a formalized relationship between ePGG signal and glottal area is lacking. The current paper proposes a Multi-Signal-ePGG (MSePGG) algorithm approach based on characterization of ePGG measurements under controlled conditions using mechanical glottal replicas. MSePGG accounts for main parameters affecting the ePGG signal: glottal area to be quantified and measurement conditions such as tissue properties and signal amplification. It is shown that MSePGG enables quantitative and continuous measurement of the time-varying glottal area on mechanical replicas. Application to a human subject is illustrated and discussed.

1. Introduction

Observation, and further measurement, of the glottal area between the moving vocal folds during breathing, speech production or swallowing, has been for long a major technological challenge. Since Garcia's pioneer experiments using mirrors [1], several different techniques have been developed and optimized. Video recordings using an endoscope coupled with a stroboscopic light or a high speed camera has become a very popular technique despite of the costs of the equipment and the need for extensive post-processing in the case of high-speed recordings [2–6]. This technique might cause discomfort and is invasive due to the insertion of optic devices (through the oral cavity in the case of a rigid endoscope, or through the nasal cavity in the case of a flexible endoscope), and thus a medical environment is required and pronunciation of certain phonemes can be hindered or inhibited. Quantitative area extraction from endoscopic images remains challenging even when stereo-endoscopy or additional devices are used [3,7–9], also being due to the trade-off between spatial and temporal resolution for image acquisition among others. Non-invasive alternatives are very few [10]. Ultrasound techniques have been tested but lack of spatial resolution [11–13], and therefore ultrasound-based imaging is mostly used for innocuous visualization only [14–18].

PhotoGlottoGraphy (PGG) [10,19] consists of devices for illuminating the glottis and measuring the amount of light that passes between the vocal folds. In its original development, PGG is an invasive technique as it requires the insertion of a light source or of a light sensor through the oral or nasal cavity. In contrast, External PhotoGlottoGraphy (ePGG) [20–22] is a non-invasive technique, both light source and sensor being placed outside of the vocal tract on the exterior of the neck (Fig. 1). Another difference with the classical PGG is the use of a lightning in the near infrared (IR) instead of visible light. Indeed, IR wavelengths in this range 700–1000 nm are reported to transilluminate large sections of human tissue [23–25]. Given the non-invasive nature of ePGG, this system no longer requires a medical environment and allows to make continuous measurements with as less disturbance as possible, *e.g.* during speech production. Consequently, if a relationship between measured ePGG signals and glottal area variation $A_g(t)$ can be established, ePGG is suitable to observe variations of the glottal area non-invasively and continuously regardless of location (medical practice, laboratory, field, ...) which makes it an interesting technique for many disciplines. Therefore, the aim of this work is to investigate and formalize the relationship between ePGG measurements and glottal area.

In Section (2), the ePGG system is detailed. Next, mechanical replicas and setups are presented (Section 3) and used to experimentally characterize (Section 4) the relationship between varying glottal area and ePGG signal under controlled conditions. From this characterization, a Multi-Signal-ePGG (MSePGG) model and parameters estimation procedure is proposed (Section 5).

https://doi.org/10.1016/j.bspc.2019.01.014
1746-8094/

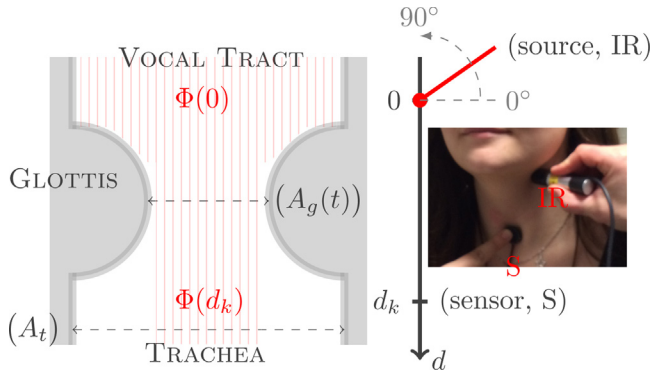


Fig. 1. External photoglottograph (ePGG) principle (mid-coronal plane) of glottal transillumination: orientation angle of light source (IR), distance from source d, light sensor (S) at position d_k, trachea area A_t, glottal area $A_g(t)$ and light flux $\Phi(d)$.

MSePGG estimated glottal areas are then validated (Section 6) on a deformable glottal replica and its application to a human subject is discussed. The general discussion and conclusion is formulated in Sections 7 and 8.

2. ePGG measurement system

The ePGG system [20–22] consists of two main elements (Fig. 1): a light source (infrared LED, LSF812N1, wavelength 810 nm, size ≤5 mm, beam angle 45 ± 5°) and a single light sensor (photo-diode, Vishay Semiconductors BP104, peak sensitivity at wavelength 950 nm, size ≤3 mm) placed in a holder. Electrical ePGG signals (between 0 V and 5 V) are acquired using a data acquisition card (Data Translation, 16 bit) and software (QuickDaq 7.8.10). In addition, the ePGG signal is amplified linearly prior to acquisition in order to compensate for *e.g.* inter-subject differences affecting light absorption (tissue, skin, *etc.*). Spurious light sources (ambient light, screen, *etc.*) are dimmed during acquisition to ensure the ePGG signal quality. In addition, both light source (IR) and sensor (S) are shielded once their position is fixed.

The light source (IR) is positioned at different supraglottal positions to illuminate the glottis through the surface of the front neck and a sensor (S) is placed at a subglottal position to record the light variation due to vocal fold displacement that modulates glottal area A_g as schematized in Fig. 1. Concretely for a human subject, in this work, the sensor (S) is fixed in the mid-sagittal plane above the suprasternal notch and the light source (IR) is positioned somewhere along the mid-coronal plane as shown in Fig. 2. Distance d (Fig. 1) indicates then the distance between transverse (horizontal) planes containing the source and sensor, respectively. Furthermore, the orientation of the IR source is varied by turning its holder in the mid-coronal plane as depicted in Fig. 1. As a convention, light emitted towards the sensor (following axis d) yields orientation angle 90° and light emitted parallel to the transverse plane (perpendicular to d) yields orientation angle 0°. The ePGG signal is sampled at 20 kHz so that the temporal resolution of 0.05 ms is excellent given the frequency range of interest, *i.e.* typical ≤250 Hz during phonation (Table 1). This temporal resolution is also high in comparison with other techniques using high-speed imaging [4].

3. Mechanical replicas and setups

To fully assess the potential of ePGG as a non-invasive measurement of glottal area A_g, the relationship between ePGG and A_g needs to be studied quantitatively as a function of parameters potentially affecting the ePGG signal. Therefore, mechanical replicas of laryngeal airway portions are mounted to an experimental setup developed to control and measure physical quantities in a reproducible and accurate way. An overview of variables of interest and their order of magnitudes on human adult subjects and on replicas is given in Table 1. Note that the given fundamental frequency range is associated with vocal folds self-oscillation and hence rapid A_g variations during human phonation. Slower A_g variation occurs during other phoneme production, respiration, *etc.*

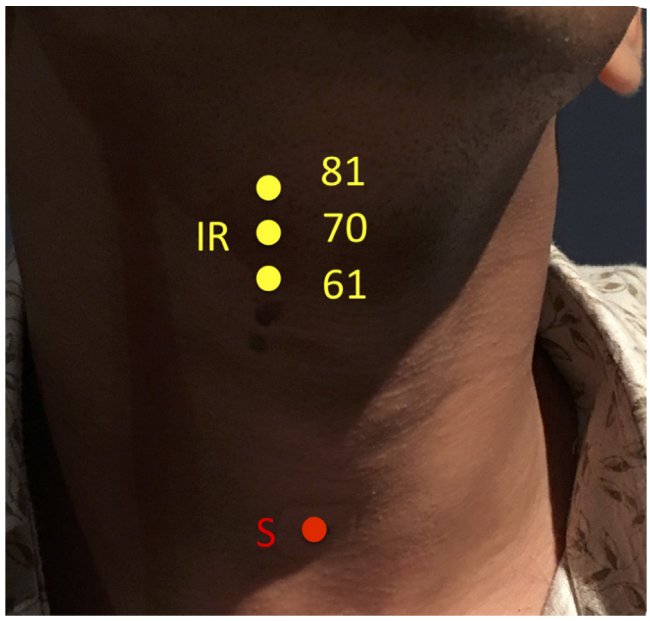


Fig. 2. Illustration of ePGG fixed subglottal light sensor (S) position and 3 supraglottal source (IR) positions along the front neck surface and source-sensor distances d [mm].

Table 1
Typical values for variables of interest for adults and mechanical replicas [26–31].

Quantity	Symbol	Human	Replicas[a,b]
Glottal area	A_g	≤270 mm^2	≤250 mm^2
Trachea radius	R_t	8–10 mm	12.5 mm
Trachea length	L_t	120 mm	150–260 mm
Frequency[c]	f_0	80–250 Hz	90–200 Hz
Subglottal pressure[c]	P_u	≤1500 Pa	≤1500 Pa

[a] Rigid mechanical replica (Section 3.2) is scaled 3:1.
[b] Deformable mechanical replica (Section 3.3) is scaled 1:1.
[c] During self-oscillation of the deformable replica.

The upper bound of glottal area A_g for human subjects (Table 1) corresponds to the maximum reported for quiet breathing [29,30]. An average value for normal subjects is disagreed on in literature due to inter subject variability, the difference between inspiration and expiration and the dependence on breathing effort. As an example, average values reported for normal subjects in [29,30] differ for both inspiration (126 ± 8 mm^2 versus 217 ± 54 mm^2) and expiration (70 ± 7 mm^2 versus 178 ± 35 mm^2). From trachea radius $R_t \approx 9 \pm 1$ mm (Table 1), it follows that trachea area $A_t = \pi R_t^2$ is approximated as $A_t \approx 254 \pm 60$ mm^2. This standard deviation results in ≤23% under- or overestimation of A_t, which reflects the limited inter- and intra-subject variability of A_t compared to. Furthermore, A_t is independent from the glottal aperture so that A_t holds for normal subjects as well as subjects suffering from a (vocal folds) pathology affecting the glottal area during respiration, closure or phonation. Next, simplified mechanical replicas and setups are presented. To study the variation of the glottal area in a controlled way two mechanical vocal fold replicas are used, a rigid (Section 3.2) one and a deformable (Section 3.3) one. This way rapid glottal area variation during vocal folds self-oscillation (deformable replica) as well as slow glottal area variation during

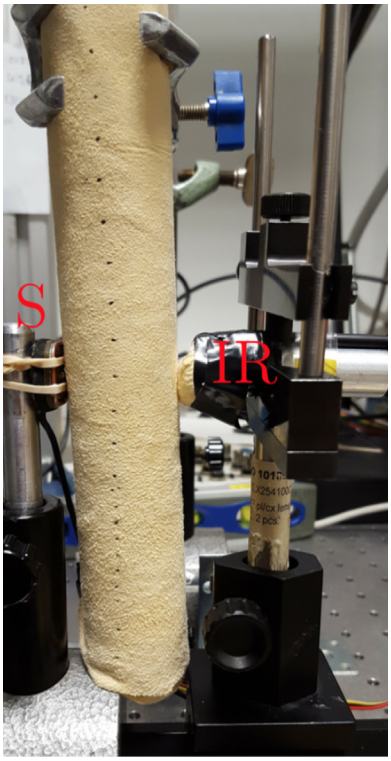


Fig. 3. Mechanical airway replica with ePGG light sensor (S) and source (IR).

glottal abduction or adduction (rigid replica) can be reproduced while the glottal area is known. To mimic the effect of airway walls, a simple mechanical airway replica (Section 3.1) is mounted to the subglottal and to the supraglottal side of each vocal fold replica so that the light sensor and source of the ePGG device can be attached on the exterior of the airway replicas.

3.1. Mechanical airway wall replica

The mechanical 'airway wall' replica, illustrated in Fig. 3, consists of a uniform tube (transparent plexiglas, internal area $A_u = 491\,\text{mm}^2$, wall thickness 2 mm, length 150–260 mm) which can be covered with layers of lamb leather (layer thickness 0.7 mm) in order to vary wall light absorption, *i.e.* representing differences in human airway wall absorption properties due to wall thickness, wall tissue composition, *etc.* [23–25]. Lamb leather is used since it easily available and allows to vary ePGG voltages within a range pertinent for human subjects [20–22]. Note that it is not attempted to reproduce realistic details of a multi-layered airway tissue since the sensor is sensible to the absorption resulting from all tissue between the source and the sensor regardless of its composition. Instead, a simple way to vary the wall absorption is proposed.

3.2. Rigid mechanical vocal fold replica

The rigid vocal fold replica [27], illustrated in Fig. 4, consists of two rigid parallel half cylinders (radius 10 mm, constant width $l_g = 25$ mm, dural), one of which is forced into motion by an eccentric motor. Sinusoidal movement of the gap between both half cylinders is imposed with a frequency up to 12 Hz while the amplitude of the motion varies in the range of 0 mm up to 1.2 mm. Time-varying glottal width $h_g(t)$ is assessed by means of a calibrated optical emitter-receptor system (OPB700, accuracy ±0.01 mm, sampling frequency 10 kHz), so that the rectangular glottal area is obtained as $A_g(t) = h_g(t) \cdot l_g$.

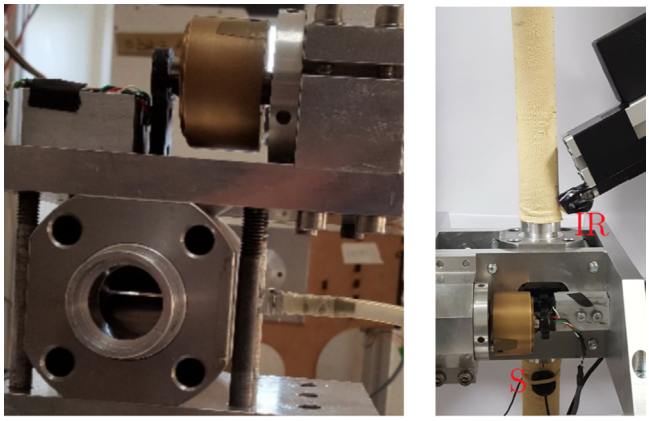


Fig. 4. Rigid mechanical vocal fold replica without (a) and with (b) subglottal and supraglottal airway replicas to which ePGG sensor (S) and source (IR) are mounted.

3.3. Deformable mechanical vocal fold replica

The deformable vocal fold replica [32,31], illustrated in Fig. 5, consists of two latex tubes (radius 5 mm) filled with water. When airflow is generated by a compressor (Atlas Copco, GA5 FF-300-8 equipped with pressure regulator and manual valve) and passes through the gap between both tubes, fluid-structure interaction leads to self-oscillation, in the same way as during human phonation. The associated time-variation of the gap is observed using a high-speed camera (Motion BLITZ Eosens Cube 7, 525 frames per second and aperture time 750 μs). Time-varying glottal area $A_g(t)$ (accuracy $\pm 0.002\,\text{mm}^2$) is then extracted from the gathered images as detailed in [31].

It is noted that absorption properties of airway wall material (mostly lamb leather) and vocal folds (mostly water) are different and that transillumination through the water-filled latex tubes occurs.

4. ePGG signal characterization

The ePGG system (Section 2) is assessed on the mechanical replicas (Section 3). Since experimental setups are equipped to measure the glottal area, the relationship between ePGG signal and glottal area can be systematically 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 (Section 4.1) and secondly for dynamic geometrical configurations with time-varying glottal area (Section 4.2). During all the experiments, the temperature was maintained at $21.5 \pm 1.0^\circ$C.

4.1. Static glottal area

In this section, static geometries are considered so that mean ePGG signals are shown. Mean values are derived from consecutive ePGG signal for 3 s and the coefficient of variation yields less than 24%.

Firstly, the effect of the source-sensor distance d (Fig. 1) on the ePGG signal is sought. The ePGG system is positioned on the mechanical airway replica with the constant area ($A_u = 491\,\text{mm}^2$, Section 3.1) as shown in Fig. 3. The source-sensor distance d is systematically varied by repositioning the source in the range $d \leq 200$ mm and the orientation angle is 27°. In addition, in order to mimic the influence of wall tissue thickness, measurements are performed for two (thickness 1.4 mm) and three leather layers

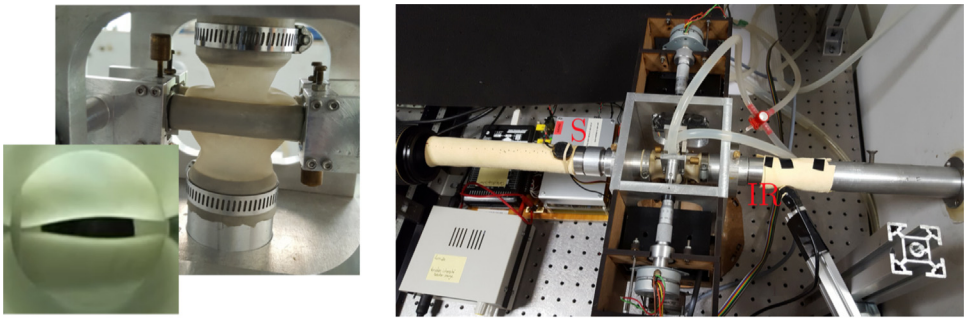


Fig. 5. Deformable mechanical vocal fold replica without (a) and with (b) subglottal and supraglottal airway replicas to which ePGG sensor (S) and source (IR) are mounted.

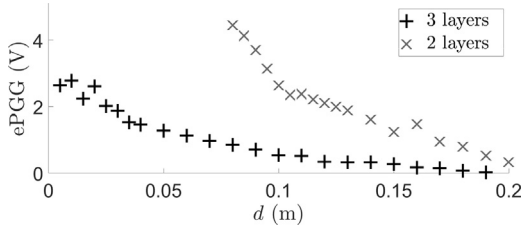


Fig. 6. Mean ePGG signal as a function of source-sensor distance d for the airway replica with 2 (×) and 3 (+) leather layers.

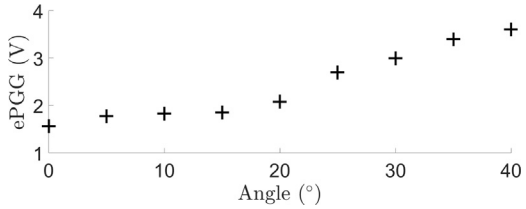


Fig. 7. Mean ePGG signal as a function of source orientation.

(thickness 2.1 mm). Measured mean ePGG signals are plotted in Fig. 6.

The ePGG signal decreases with d regardless of wall thickness. Linear fitting of measured ePGG signals in the range $d \leq 100$ mm (appropriate for human subjects) and in the range $d \geq 100$ mm (appropriate for dynamic mechanical replicas), results in coefficient of determination $R^2 \geq 98.9\%$. 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 the remaining experiments are done with 2 layers (thickness 1.4 mm). Secondly, static geometrical configurations are assessed in order 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 (Fig. 3), *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. 7.

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 $22.5 \pm 2.5°$ (Section 2). 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

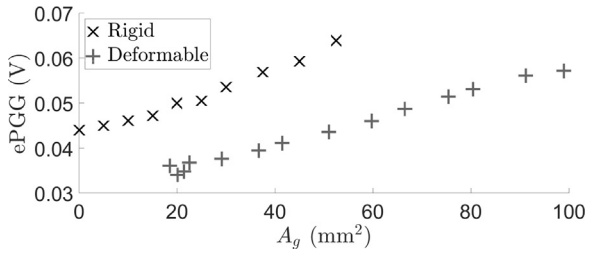


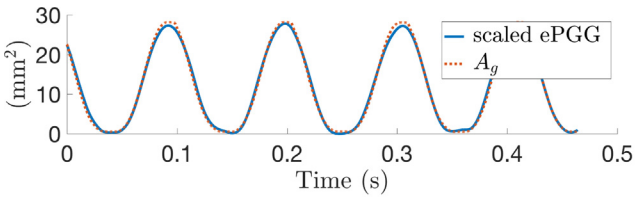
Fig. 8. Mean ePGG signal as a function of static glottal area A_g for a rigid (×) and a deformable (+) mechanical glottal replica.

rigid (Section 3.2) and deformable (Section 3.3) mechanical replicas are used and a uniform mechanical airway wall replica is attached at each end as shown in Figs. 4(b) and 5 (b). The 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 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. 8.

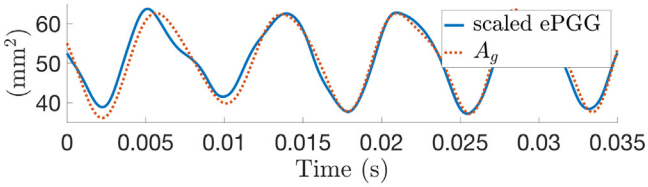
The ePGG signal increases linearly with A_g for the rigid ($R^2 = 99.2\%$) and the deformable ($R^2 = 98.2\%$) replica, indicating that ePGG signal and glottal area relate well using a linear approximation. Note that, in general, the differences in slope and offset in the figure may be due to 1) positioning of the ePGG system (source-sensor distance d, orientation angle) and 2) channel wall properties affecting light absorption (thickness, material, *etc.*).

4.2. 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 replica (Fig. 4) with source-sensor distance $d = 150$ mm and for the flow-driven deformable replica (Fig. 5) with source-sensor distance $d = 257$ mm. For the motor-driven replica, oscillations are imposed at frequencies $f_0 \in \{2, 5, 10, 12\}$ Hz and the glottal area is varied in the range $0 \leq A_g \leq 40$ mm^2. For the flow-driven deformable replica self-oscillations are observed for mean subglottal pressures $P_u \in \{500, 570, 720, 840\}$ Pa so that the fundamental frequencies yielded $f_0 \in \{113, 125, 129, 131\}$ Hz and the area varied in the range $20 \leq A_g \leq 100$ mm^2. Typical examples of correlated time signals for slow (rigid) and fast (deformable) vocal folds displacement are plotted in Fig. 9. The scaled ePGG signal is firstly normalized by accounting for its mean and standard deviation and then scaled to mm^2 by accounting for the mean and standard deviation of the shown A_g.



(a) Rigid, correlation coefficient 94%



(b) Deformable, correlation coefficient 90%

Fig. 9. Correlated time signals of glottal area $A_g(t)$ (dashed line) and scaled ePGG (full line): (a) rigid replica (f_0 = 10 Hz), (b) deformable replica (f_0 = 129 Hz).

Correlation coefficients between ePGG signals and glottal area $A_g(t)$ yield 90% for the rigid and 85% for the deformable glottal folds replica.

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

5. Multi-signal-ePGG (MSePGG)

In Section 4, 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 sensor and 3) the measurement condition determined by the combination of wall properties (*e.g.* absorption), environment (*e.g.* light) and ePGG system settings (*e.g.* amplification outlined in Section 2) and positioning (*e.g.* orientation angle). In the following, a Multi-Signal-ePGG (MSePGG) approach is proposed accounting for each of these factors. The underlying model (Section 5.1), parameters estimation (Section 5.2.1) and their initialization (Section 5.2.2) is outlined. The sought relationship between ePGG signal and glottal area $A_g(t)$ is detailed (Section 5.3) and an overview of the resulting MSePGG workflow is summarized (Section 5.4).

5.1. Model

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,

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

where transmitted light intensity $I(d_k)$ at sensor position d_k is then expressed using light flux Φ as

$$I(d_k) = \oiint_{A_{\min}(d_k)} \Phi(d_k) dA, \tag{2}$$

where

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

is the minimum area encountered by the transmitted light flux between the source and sensor positions. Furthermore, in Section 4 it was shown that for the ranges of interest (Table 1), 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

$$\Phi_m(d) = \alpha_d d + \beta_d, \tag{4}$$

with slope $\alpha_d < 0$ and offset $\beta_d > 0$ (see Section 4). From (2), $I(d_k)$ is now modeled as

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

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

$$\begin{aligned} U_m(d_k) &= \gamma \left((\alpha_d d_k + \beta_d) \cdot A_{\min}(d_k) \right) + \eta \\ &= \left(\alpha_v d_k + \beta_v \right) \cdot A_{\min}(d_k) + \eta \end{aligned} \tag{6}$$

where $\eta > 0$ is the 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 this latter model only holds for positive coefficient $\alpha_v d_k + \beta_v$: in other words source-sensor distance d_k must satisfy $d_k \leq -\beta_v/\alpha_v$. Additionally, it is noted that $A_{min} = 0$ corresponds to glottal closure for which no direct light is transmitted through the vocal folds although light transmitted due to tissue transillumination might remain. Therefore, η is independent of d and A_{min} so that η reflects solely the measurement condition. Considering now the time-variation of the glottal opening, model (6) can be directly extended as

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

Consequently using model (7), extracting area $A_{min}(d_k, t)$ from measured ePGG signals $U(d_k, t)$ reduces to a problem of parameter estimation as assessed in the next section.

5.2. Calibration

5.2.1. Parameter estimation

Let vector $\boldsymbol{U} = \begin{bmatrix} U(d_1) & \cdots & U(d_K) \end{bmatrix}^\dagger \in \mathbb{R}^K$ denote the concatenation of K mean ePGG measurements and $.^\dagger$ the transpose operator. Vectorizing model (6) yields

$$\boldsymbol{U}_m = [\alpha_v \boldsymbol{d} + \beta_v] \odot [\boldsymbol{A_{\min}}] + \eta, \tag{8}$$

where $\boldsymbol{d} = \begin{bmatrix} d_1 & \cdots & d_K \end{bmatrix}^\dagger \in \mathbb{R}^K$ is the sensor position vector, $\boldsymbol{A_{\min}} = \begin{bmatrix} A_{\min}(d_1) & \cdots & A_{\min}(d_K) \end{bmatrix}^\dagger \in \mathbb{R}^K$ is the minimum area vector and $\odot$ is the Hadamard product. It is worth noting that this latter model relies on the $K+3$ parameters $\left\{ \{A_{\min}(d_k)\}_{1 \leq k \leq K}, \alpha_v, \beta_v, \eta \right\}$ that should be estimated from K ePPG measurements $\{U(d_k)\}_{1 \leq k \leq K}$. So, without additional constraints, model (8) is not identifiable. A solution to overcome this inherent problem is to reduce the number of parameters to be estimated by measuring for several distances the ePGG voltage associated with the same A_{min}.

To simplify the notations, and without loss of generality, let us assume that $\tilde{K}$ measurements have been performed for each of the N different A_{min} values (so that $N \times tildeK = K$). Consequently, model (8) can be recast as

$$\boldsymbol{U}_m = \left[\alpha_v \boldsymbol{d} + \beta_v \right] \odot \left[\tilde{\boldsymbol{A}}_{\boldsymbol{\min}} \otimes \mathbb{1}_{\tilde{K}} \right] + \eta, \tag{9}$$

where $\tilde{\boldsymbol{A}}_{\boldsymbol{\min}} = \begin{bmatrix} A_{\min_1} & \cdots & A_{\min_N} \end{bmatrix}^\dagger \in \mathbb{R}^N$, $\mathbb{1}_{\tilde{K}} \in \mathbb{R}^{\tilde{K}}$ is the identity vector and $\otimes$ is the Kronecker product. Using such a constraint, the number of parameters is reduced to $N+3$, leading thus to an identifiable model as soon as $N+3 \leq K$.

In practice, the set of parameters $\left\{\tilde{\boldsymbol{A}}_{\mathbf{min}}, \alpha_v, \beta_v, \eta\right\}$ is estimated by minimizing mean square error $J()$ between the K ePGG measurements $\boldsymbol{U}$ and the related ePGG model $\boldsymbol{U}_m$ provided by (9):

$$\left\{\hat{\tilde{\boldsymbol{A}}}_{\mathbf{min}}, \hat{\alpha}_v, \hat{\beta}_v, \hat{\eta}\right\} = \underset{\substack{\tilde{\boldsymbol{A}}_{\mathbf{min}} \geq 0,\ \alpha_v \leq 0,\\ \beta_v \geq 0,\ \eta \geq 0}}{arg\ min} J\left(\tilde{\boldsymbol{A}}_{\mathbf{min}}, \alpha_v, \beta_v, \eta\right) \tag{10}$$

where

$$J(\tilde{\boldsymbol{A}}_{\mathbf{min}}, \alpha_v, \beta_v, \eta) = \frac{1}{K}(\boldsymbol{U} - \boldsymbol{U}_m)^{\dagger}(\boldsymbol{U} - \boldsymbol{U}_m). \tag{11}$$

To solve (10), a gradient descent iterative method is used [33]. The initialization of this iterative optimization algorithm is discussed in the next section.

5.2.2. Initialization

To initialize parameters set $\left\{\tilde{\boldsymbol{A}}_{\mathbf{min}}, \alpha_v, \beta_v, \eta\right\}$ a three-step procedure is applied.

Firstly, η is initialized as $\eta^{(0)}$, the mean ePGG voltage when no direct light is transmitted. Indeed, this condition is obtained when $A_{min} = 0$, so that (7) resumes to $U_m(d_k, t) = \eta$, $\forall(d_k, t)$. Thus, $\eta^{(0)}$ is estimated as the average value of all the measurements for which $A_{min} = 0$. Note that trained subjects can produce glottal closure on instruction, for other subjects closure is mimicked by supplying no excitation light. The effect of the subject instruction on η^0 and resulting parameter and area estimations needs to be assessed in future studies.

Secondly, it is sought to initialize α_v and β_v by measuring ePGG signals $\boldsymbol{U}^{(0)}$ associated with a constant reference area $A_0 > 0$ (detailed in the next section) for at least two different sensor positions $\boldsymbol{d} = [d_1 \ldots d_P]^{\dagger} \in \mathbb{R}^P$, with $P \geq 2$ (see[1]) and $\tilde{\boldsymbol{A}}_{\min} = A_0\, \mathbb{1}_P \in \mathbb{R}^P$. From model (8) and assuming that A_0 is known, one can estimate $\alpha_v^{(0)}$ and $\beta_v^{(0)}$ by the following least mean square (LMS) minimization

$$\alpha_v^{(0)},\quad \beta_v^{(0)} = \underset{\alpha_v,\, \beta_v}{arg\ min} \left\| \boldsymbol{U}^{(0)} - \boldsymbol{U}_m^{(0)} \right\|_2^2, \tag{12}$$

where

$$\boldsymbol{U}_m^{(0)} = \quad A_0\,[\alpha_v \boldsymbol{d} + \beta_v] + \eta^{(0)}.$$

Consequently, $\alpha_v^{(0)},\quad \beta_v^{(0)}$ are expressed in a closed form as

$$\begin{pmatrix} \alpha_v^{(0)} \\ \beta_v^{(0)} \end{pmatrix} = \left(\boldsymbol{D}^{\dagger}\boldsymbol{D}\right)^{-1}\boldsymbol{D}^{\dagger}\frac{\left(\boldsymbol{U}^{(0)} - \eta^{(0)}\right)}{A_0}, \tag{13}$$

with $\boldsymbol{D} = [\boldsymbol{d}, \mathbb{1}_P] \in \mathbb{R}^{P\times 2}$.

Thirdly, from $\left\{\alpha_v^{(0)}, \beta_v^{(0)}, \eta^{(0)}\right\}$ and measured ePGG signals $\boldsymbol{U}$, the N elements of $\tilde{\boldsymbol{A}}_{\mathbf{min}}$ $\left(\left\{A_{\min_n}\right\}_{1\leq n\leq N}\right)$ are initialized by a LMS minimization obtained by approximating ePGG measurements $\boldsymbol{U}$ with model (9), leading to

$$\forall\, n,\quad A_{\min_n}^{(0)} = \frac{[\alpha_v^{(0)}\boldsymbol{d}_n + \beta_v^{(0)}]^{\dagger}\left(\boldsymbol{U}_n - \eta^{(0)}\right)}{[\alpha_v^{(0)}\boldsymbol{d}_n + \beta_v^{(0)}]^{\dagger}[\alpha_v^{(0)}\boldsymbol{d}_n + \beta_v^{(0)}]}, \tag{14}$$

where $\boldsymbol{U}_n \in \mathbb{R}^{\tilde{K}}$ (resp. $\boldsymbol{d}_n \in \mathbb{R}^{\tilde{K}}$) is the nth subvector of $\boldsymbol{U}$ (resp. $\boldsymbol{d}$) so that $\boldsymbol{U} = [\boldsymbol{U}_1^{\dagger}, \ldots, \boldsymbol{U}_N^{\dagger}]^{\dagger}$ (resp. $\boldsymbol{d} = [\boldsymbol{d}_1^{\dagger}, \ldots, \boldsymbol{d}_N^{\dagger}]^{\dagger}$).

[1] Note that $P \geq 2$ is a theoretical criterion. In practice, $P \geq 3$ is used.

5.3. Estimation of glottal area $A_g(t, k)$

When the glottis is situated in between the source (IR) and the sensor (S) (Fig. 1) glottal area $A_g(t, k)$ can be estimated, where $A_g(t, k)$ indicates the glottal area as a function of time t using the signal gathered through the sensor at position d_k. Indeed in this case, minimum area A_{min} encountered by the light flux corresponds to glottal area A_g and thus

$$A_g(t, k) = A_{\min}(d_k, t). \tag{15}$$

Consequently, glottal area $A_g(t, k)$ is estimated in the same way as $A_{min}(d_k, t)$ outlined in the previous section. This means that firstly the parameter set $\left\{\tilde{\boldsymbol{A}}_{\boldsymbol{g}}, \alpha_v, \beta_v, \eta\right\}$ is estimated using $\tilde{K}$ different sensor positions for N different glottal areas $\boldsymbol{A}_g \in \mathbb{R}^N$ so that the model is identifiable, *i.e.* $N + 3 \leq K$ with $K = N \times \tilde{K}$.

Once the parameter set estimation is assessed and assuming that the measurement condition is not altered – *i.e.* unchanged combination of subject, source positioning, environment and hardware settings – it is argued (and shown in Section 6) that any additional ePGG measurement $U(d_i, t)$ with a sensor at any source-sensor distance d_i results in glottal area estimation $\hat{A}_g(t, i)$ using (7) and substituting $U_m(d_i, t)$ with $U(d_i, t)$ and $\left\{\alpha_v, \beta_v, \eta\right\}$ with $\left\{\hat{\alpha}_v, \hat{\beta}_v, \hat{\eta}\right\}$ so that

$$\hat{A}_g(t, i) = \frac{U(d_i, t) - \hat{\eta}}{\hat{\alpha}_v d_i + \hat{\beta}_v}. \tag{16}$$

It is worth noting that the estimated area $\hat{A}_g(t, i)$ does not depend on source-sensor distance d_i, but depends on the choice of reference area A_0 during initialization. Setting $A_0 = 1$ results in a normalized area estimation relative to A_0. When a quantitative area estimation is needed, it is proposed to use trachea area ($A_0 \approx 254\,\mathrm{mm}^2$) or glottal area during quiet inspiration ($A_0 \approx 127\,\mathrm{mm}^2$) as a reference area (Section 3, [29]).

Consequently, once MSePGG calibration is done, resulting $\{U(d_i, t)\}_{1\leq i\leq I}$ associated with $I \geq 1$ sensors results in I estimations of the minimum area, *i.e.* $\{\hat{A}_g(t, i)\}_{1\leq i\leq I}$. Thus a single or multiple sensor positions can be used to estimate the glottal area once calibration is achieved. For $I > 1$ a set of estimations is obtained, which can be further post-processed, *e.g.* to further optimize the estimation by applying a sensor-selection criteria (*e.g.* signal-to-noise ratio) or to further characterize the estimation (*e.g.* determine an uncertainty on the resulting minimum area estimation). Such post-processing is not considered in this work.

5.4. Workflow

To summarize, the MSePGG workflow of minimum area estimation is schematized in Fig. 10. The first two steps – parameter initialization and estimation (Section 5.2) – lead to MSePGG calibration in order to account for the specific measurement condition (subject, environment, amplification, *etc.*). Once calibration is achieved the minimum area as a function of time $\hat{A}_{\min}(t, i)$ can be estimated for any experimental protocol and for any of $I \geq 1$ sensor positions as long as calibration parameters $\left\{\hat{\alpha}_v, \hat{\beta}_v, \hat{\eta}\right\}$ remain valid, *i.e.* measurement conditions are unaltered. Different post-processing strategies can be applied to the resulting I estimations of minimum area.

Concretely, in Section 6.2 where the MSePGG approach is applied to measurements on a human speaker in order to estimate the glottal area, the same three sensors positions are considered during all the steps so that $\tilde{K} = 3$, $P = 3$ and $I = 3$. Reference area A_0 corresponds to the area during quiet inspiration. It is noted that the choice of A_0 needs to be addressed carefully when an extended study on human subjects is aimed since an error in A_0 might affect

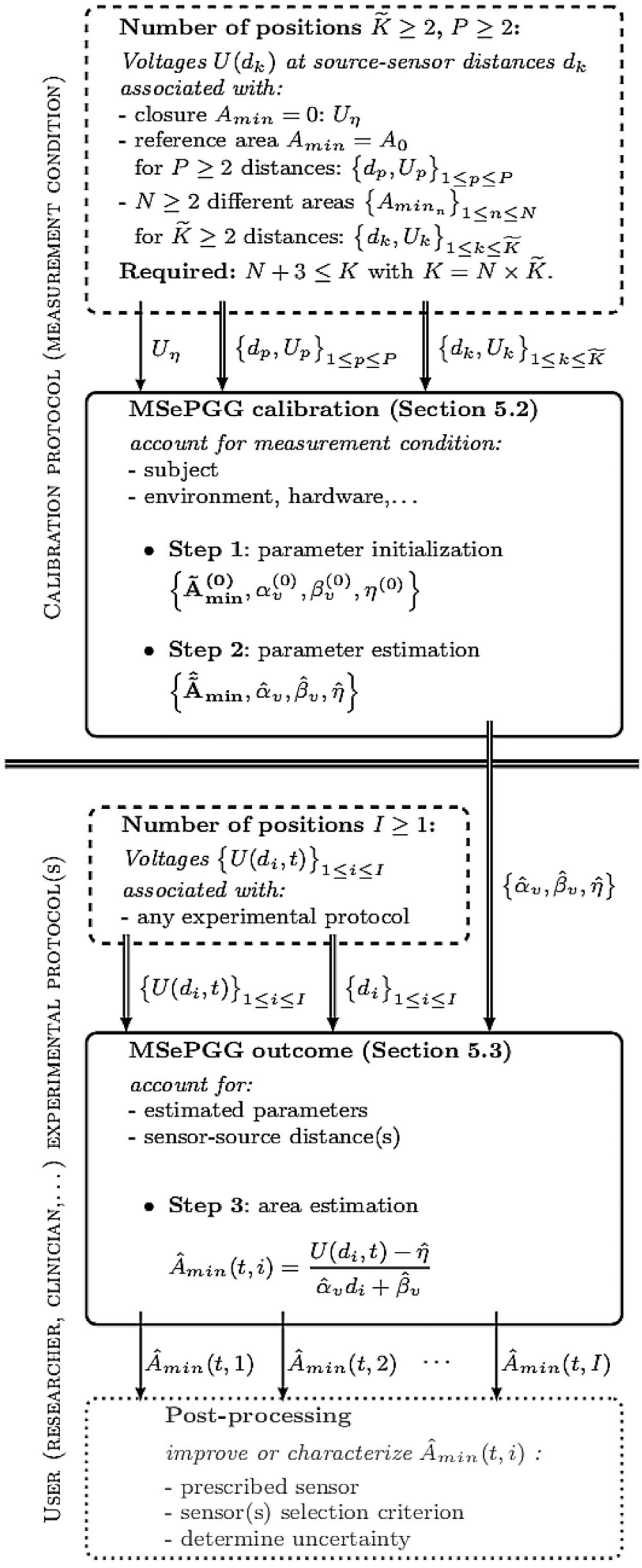
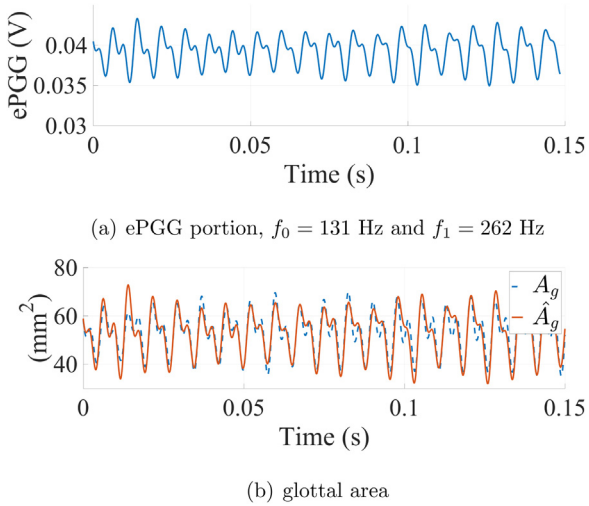


Fig. 10. MSePGG workflow overview.

estimated parameters α_v and β_v and hence estimated glottal areas $\hat{A}_g(t, i)$ following (16). Furthermore, voltages associated with $N = 3$ minimum areas were determined as the mean values associated with the maximum, minimum and averaged voltage during 16 subsequent periods of vowel /a/. No post-processing is performed so that glottal area estimations $\hat{A}_g(t, i)$ are directly shown.



(a) ePGG portion, $f_0 = 131$ Hz and $f_1 = 262$ Hz

(b) glottal area

Fig. 11. Deformable mechanical replica: (a) ePGG signal portion, (b) measured A_g (dashed line) and estimated $\hat{A}_g$ (full line) area. Mean error yields 5.4%.

6. MSePGG results

MSePGG outlined in Section 5 is applied following the workflow shown in Fig. 10. Measurements on the deformable mechanical replica (Section 6.1) and on a human speaker (Section 6.2) are assessed.

6.1. Deformable mechanical replica

Time-varying ePGG signals due to time-varying glottal area during self-oscillation are measured on the same deformable mechanical replica and under the same measurement conditions (environmental and instrumental) as described in Section 3.3. MSePGG parameter estimation is assessed using data discussed in Section 4 following the procedure outlined in Section 5.2.1. Resulting values of parameters $\hat{\alpha}_v$, $\hat{\beta}_v$ and $\hat{\eta}$ can be used directly in (16) to estimate glottal area $\hat{A}_g$ since estimated parameters ($\hat{\alpha}_v$, $\hat{\beta}_v$ and $\hat{\eta}$) depend solely on the replica and measurement conditions and not on the glottal area (Section 5.1).

Mechanical replica properties [31] are set asymmetrical between both vocal folds. The resulting area variation during self-oscillation and hence associated ePGG signal (Fig. 11(a)) are characterized by a high harmonic distortion rate (58%) due to the presence of the fundamental frequency ($f_0 = 131$ Hz) and the second harmonic ($f_1 = 262$ Hz). High harmonic distortion rates might occur in the case of voice disorders characterized by diplophonia [34,35,5,6]. Estimated ($\hat{A}_g$) glottal areas by (16) and glottal areas (A_g) measured by imaging are shown in Fig. 11(b). Quantitative comparison between estimated and measured glottal areas results in a mean error of 5.4%. This suggests that the MSePGG approach can be applied to assess the glottal area in the case of a normal glottal variation and in the case of a voice disorder related to diplophonia. In the next section, MSePGG is applied to a human speaker.

6.2. Human speaker

In this section, ePGG measurements are obtained on a human speaker (21-year-old male, Telugu native speaker, source orientation angle 1°) without any voice disorder (self-reported). Measurement of ePGG data were performed while the speaker pronounced a sustained vowel /a/, during quiet respiration and during glottal closure. The light sensor and source are positioned as outlined in Section 2 and illustrated in Fig. 2. The three source positions resulting in the largest source-sensor distances ($d = 61$ mm,

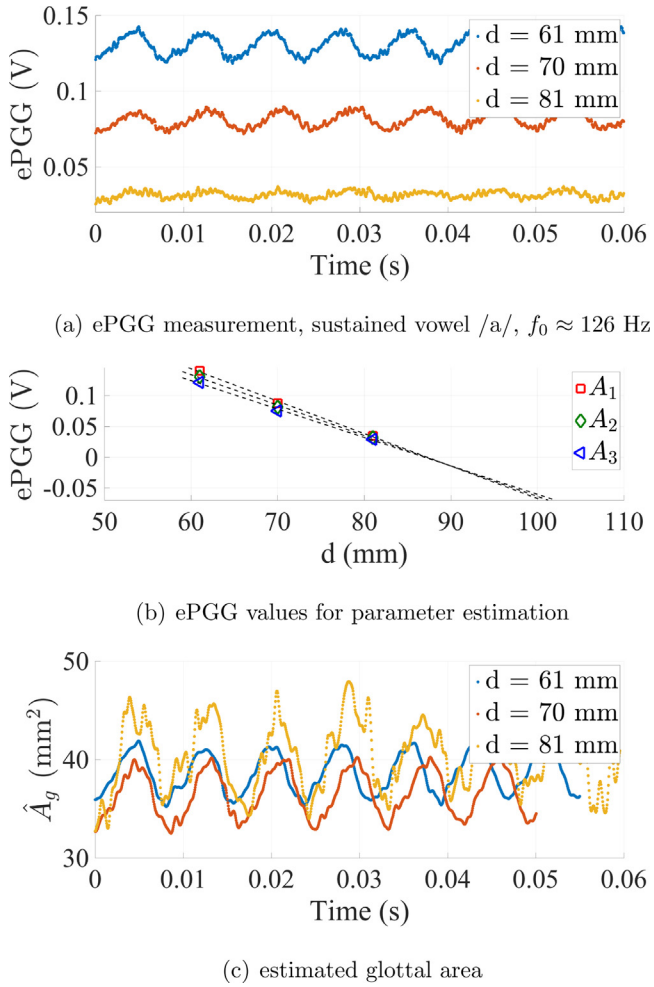


Fig. 12. Sustained vowel /a/ pronounced by a human speaker for 3 (*tildeK* = 3) source-sensor distances d: (a) ePGG data as a function of time, (b) extracted ($\tilde{K} \times N = 9$) ePGG values for MSePGG parameter estimation (symbols) as a function of d and their linear fit (dashed line) for three ($N = 3$) areas ($A_1 > A_2 > A_3$), (c) estimated glottal areas $\hat{A}_g(t)$.

$d = 70$ mm and $d = 81$ mm) were assessed as (4) is verified to hold: $P = 3$, $\tilde{K} = 3$ and $I = 3$. The ePGG measurements for each source-sensor distance during vowel /a/ are shown in Fig. 12(a).

Next, ePGG values for MSePGG parameters estimation are selected. For all three source-sensor distances d, it is aimed to determine ePGG voltages associated with three different glottal areas ($N = 3$) during vowel production so that the condition $N + 3 \leq K$ (with $K = \tilde{K} \times N$) holds and the procedure for MSePGG parameter estimation (Section 5.2.1) can be applied. For each distance d, the highest value and lowest value of the range of ePGG voltages in each period are determined. The mean value of the 16 maxima of 16 consecutive periods is considered as the ePGG voltage associated with the largest (A_1) glottal area during vowel /a/ production and this for each distance d. The same way the mean value of the 16 minima of 16 consecutive periods is considered as the ePGG voltage associated with the smallest (A_3) glottal area and this again for each distance d. The ePGG voltage corresponding to the average of these extrema for each distance d is then associated with a third glottal area (A_2) and $A_3 < A_2 < A_1$ holds. Resulting K ($K = \tilde{K} \times N$ so $K = 9$) ePGG data associated with each of these areas are plotted in Fig. 12(b) as a function of source-sensor distance d. It is observed that in agreement with (6) the ePGG level decreases linearly for each area with respect to source-sensor distance d. The linear fits yield zero voltage level at nearly the same source-sensor distance ($d = 89.2 \pm 0.2$ mm)

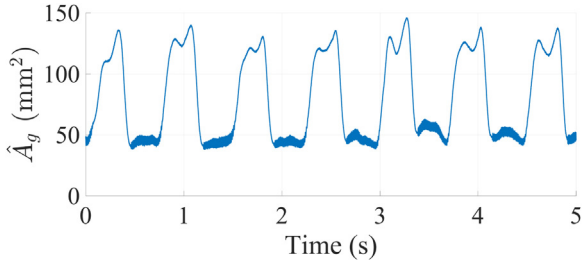


Fig. 13. Estimated glottal areas $\hat{A}_g$ for consecutive (5 s) utterance of /sa/ by a human speaker.

and thus the linear fits intersects near −0.012 V at $U_m(d_k) - \eta = 0$ following (6) for all d_k. Therefore, the intersection point's magnitude (0.012 V) provides a rough estimation of the order of magnitude of the estimated ePGG voltage associated with closed glottal area $\hat{\eta} = 0.01$ V. The extracted ePGG voltages plotted in Fig. 12(b) ($\boldsymbol{U}_{1,2,3}$) are then used to initialize MSePGG parameters following the procedure described in Section 5.2.2, *i.e.* step 1 of the calibration protocol shown in Fig. 10. Concretely, at first, $\alpha_v^{(0)}$ and $\beta_v^{(0)}$ are initialized from (13) using ePGG voltages associated with maximum glottal opening during quiet inspiration and taking glottal area $A_0 = 127$ mm^2 as a known reference area value (Section 5.3, [29]). Next, $\alpha_v^{(0)}$, $\beta_v^{(0)}$ and extracted $\boldsymbol{U}_{1,2,3}$ voltages associated with $A_{1,2,3}$ (shown in Fig. 12(b)) are used in (14) in order to initialize $A_1^{(0)} \approx 41$ mm^2, $A_2^{(0)} \approx 38$ mm^2 and $A_3^{(0)} \approx 34$ mm^2. Once all MSePGG parameters are initialized their value is estimated from (10) as described in Section 5.2.1, *i.e.* step 2 of the calibration protocol shown in Fig. 10. It is noted that the initialized ($\tilde{\boldsymbol{A}}_{\mathbf{min}}^{(\mathbf{0})} = \left[A_1^{(0)} \quad A_2^{(0)} \quad A_3^{(0)}\right]^\dagger$ resulting from (14)) and estimated ($\hat{\tilde{\boldsymbol{A}}}_{\mathbf{min}} = \left[\hat{A}_1 \quad \hat{A}_2 \quad \hat{A}_3\right]^\dagger$ resulting from (10)) area values are in good agreement (≤ 1 mm^2) and thus the proposed initialization procedure results in reasonable area values.

Once MSePGG parameters are estimated, glottal areas associated with the time-varying ePGG signals shown in Fig. 12(a) are estimated from (16). Estimated glottal areas $\hat{A}_g(t)$ for all three assessed sensor-source distances d are shown in Fig. 12(c). The mean absolute relative error between estimated glottal areas obtained from ePGG measurements at different source-sensor distances yields less than 12%. It follows that estimated areas are of the same order of magnitude as they vary in the same range, *i.e.* between 32 mm^2 and 48 mm^2. Given that ePGG signals for each source-sensor distance d are measured during a different utterance of the vowel /a/ some variation between the areas is expected. Consequently, the agreement between estimated areas is satisfying. Furthermore, this maximum amplitude of 16 mm^2 approximates the order of magnitude mentioned in literature for phonation of 12 mm^2 [9,36]. It is noted that the slight overestimation of the amplitude is obtained for the ePGG signal at $d = 81$ mm for which the signal quality is less, illustrating that sensor positioning can be further addressed. The lack of glottal closure during vowel production observed in Fig. 12(c) is confirmed by endoscopic imaging on the same subject.

Next, the same MSePGG parameters are applied to estimate (16) the time-varying glottal area during consecutive utterances of /sa/ (up to 5 s) by the same speaker in the same conditions at $d = 61$ mm as illustrated in Fig. 13. Rapid glottal area variations are observed during vowel (/a/) whereas slow variations associated with glottal opening and closing occurs to pronounce sibilant fricative /s/. As expected the glottal area widens during frication noise. It is seen that the temporal resolution allows to study the transition between vowels and fricatives in detail.

7. Discussion

Results shown in Section 6 illustrate that MSePGG provides an estimation of the time-varying minimum area on a mechanical replica and on a human speaker following the workflow summarized in Section 5.4. The MSePGG algorithm and workflow provides an elegant and innocuous method relying on 3 parameters to be estimated simultaneously. The MSePGG algorithm avoids dealing with the complexity of the composing tissue layers and anatomy. This way some restrictions related to the use of other techniques mentioned in the introduction are removed. Indeed, the MSePGG algorithm combined with the ePGG device strives to a non-invasive, non-expensive, continuous and quantitative measurement with minimum discomfort for the subject and a straightforward interpretation following an elegant and short calibration protocol at a low computational and data storage cost. Since the ePGG device is external, the danger for infections is minimal and there is no need for medical supervision. The MSePGG algorithm and workflow is a suitable candidate to provide a quantitative metric of the glottal area outside of a medical practice for clinical or non-clinical studies (health care, research, field studies) using ePGG measurements.

Some remarks can be made considering the proposed MSePGG method in Section 5. Firstly, it is noted that the accuracy of MSePGG on a human subject can be improved using a sensor array, instead of moving a single sensor to different positions to avoid intra-subject variability as well as to further simplify the parameter estimation protocol. Secondly, it is seen that the same algorithm and workflow can potentially be applied to obtain a quantitative area estimation from invasive PGG measurements as well since PGG relies on transillumination as well. Thirdly, the accuracy of MSePGG estimations is determined against area measurements on mechanical replicas for which accurate area values are available. MSePGG outcome on human subjects is merely done as an illustration since no quantitative glottal area values were available. Therefore, the influence of the parameter initialization procedure for human subjects, *i.e.* combination of instructions and parameter initialisation as well as sensor-source distances, on the parameter and area estimation needs to be assessed in future studies. In addition, thorough validation for human subjects needs to be assessed in future for a large variety of anatomical and clinical conditions. This way the applicability, advantages and limitations of the proposed MSePGG approach can be determined, *e.g.* to consider the impact of different positions of the epiglottis. It is expected that such studies will lead to further improvement of the MSePGG algorithm and workflow, the ePGG device as well as in term to a standardized protocol with respect to device positioning and settings.

8. Conclusion

Following characterization of ePGG measurements on mechanical replicas, the MSePGG algorithm and workflow is proposed in order to provide a quantitative estimation of the time-varying glottal area following a brief calibration protocol exploiting several source-sensor distances. The good quantitative agreement obtained on mechanical replicas (mean error 5.4%) and preliminary observations on a human subject (estimations within 12%) suggests that MSePGG is a promising technique to estimate the glottal area during normal as well as pathological vocal folds configurations. Future research is needed to consider the use of a sensor array ePGG device to fully investigate the MSePGG algorithm and workflow for different human subjects in order to further improve, validate and facilitate MSePGG on human subjects.

Acknowledgements

Partly funded by ArtSpeech project (ANR-15-CE23-0024). Human ePGG data registration was approved by ethics committee 1922081 (dated 02/02/2016). Thanks to D. Sathiyanarayanan for his contribution to measurements on a human speaker.

References

[1] M. Garcia, Observations on the human voice, Proc. R. Soc. Lond. 7 (1855) 399–410.
[2] J. Lohscheller, U. Eysholdt, H. Toy, M. Döllinger, Phonovibrography: mapping high-speed movies of vocal fold vibrations into 2-d diagrams for visualizing and analyzing the underlying laryngeal dynamics, IEEE Trans. Med. Imaging 27 (2008) 300–309.
[3] T. Wurzbacher, I. Voigt, R. Schwarz, M. Dollinger, U. Hoppe, J. Penne, U. Eysholdt, J. Lohscheller, Calibration of laryngeal endoscopic high-speed image sequences by an automated detection of parallel laser line projections, Med. Image Anal. 12 (2008) 300–317.
[4] D. Deliyski, R. Hillman, State of the art laryngeal imaging: research and clinical implications, Curr. Opin. Otolaryngol. Head Neck Surg. 18 (2011) 147–152.
[5] P. Aichinger, I. Roesner, M. Leonhard, B. Schneider-Stickler, D.-M. Denk-Linnert, W. Bigenzahn, A.K. Fuchs, M. Hagmüller, G. Kubin, Comparison of an audio-based and a video-based approach for detecting diplophonia, Biomed. Signal Process. 31 (2017) 576–585.
[6] P. Aichinger, M. Hagmüller, I. Roesner, B. Schneider-Stickler, J. Schoentgen, F. Pernkopf, Tracking of multiple fundamental frequencies in diplophonic voices, IEEE Trans. Audio Speech Lang. Process. 26 (2018) 330–341.
[7] M. Sawashima, S. Miyazaki, Stereo-fiberscopic measurement of the larynx: a preliminary experiment by use of ordinary laryngeal fiberscopes, Ann. Bull. RILP 8 (1974) 1–12.
[8] M. Sawashima, H. Hirose, S. Hibi, H. Yoshioka, N. Kawase, M. Yamada, Measurements of the vocal fold length by use of stereoendocope – a preliminary study, Ann. Bull. RILP 15 (1981) 9–16.
[9] H. Imagama, K. Sakakibara, I. Tokuda, M. Otsuka, N. Tayama, Estimation of glottal area function using stereo-endoscopic high-speed digital imaging, Proc. Interspeech (2010) 1–4.
[10] R. Baken, Clinical Measurement of Speech and Voice, Allyn and Bacon, USA, 1987, pp. 197–240.
[11] C. Hertz, K. Lindström, B. Sonesson, Ultrasonic recording of the vibrating vocal folds, Arch. Otolaryngol. 69 (1970) 223–230.
[12] S. Halmet, J. Reid, Transmission of ultrasound through the larynx as a mean of determining vocal-fold activity, IEEE Trans. Biomed. Eng. 19 (1972) 34–37.
[13] T. Kaneko, K. Uchida, H. Suzuki, K. Komatsu, T. Kanesaka, N. Kobayashi, J. Naito, Ultrasonic observation of vocal folds vibration, in: Vocal Folds Physiology, Univ. of Tokyo, USA, 1981, pp. 107–117.
[14] E. Friedman, Role of ultrasound in the assessment of vocal cord function in infants and children, Ann. Otol. Rhinol. Laryngol. 106 (1997) 199–209.
[15] A. Arutti, M. Poumayrac, Larynx ultrasonography: an alternative technique in the evaluation of the aero-digestive crossroad, Rev. Imagenol. 14 (2010) 30–36.
[16] S. Parangi, Editorial: translarygeal vocal cord ultrasound: ready for prime time, Surgery 159 (2016) 67–69.
[17] C. Krausert, A. Olszewski, L. Taylor, J. McMurray, S. Dailey, J. Jiang, Mucosal wave measurement and visualization techniques, J. Voice 25 (2011) 395–405.
[18] M. Masood, B. Huang, A. Goins, T. Hackman, Anatomical factors affecting the use of ultrasound to predict vocal fold motion: a pilot study, Am. J. Otolaryngol. 39 (2018) 413–417.
[19] B. Sonesson, A method for studying the vibratory movements of the vocal folds, J. Laryngol. Otol. 73 (1959) 732–737.
[20] K. Honda, S. Maeda, Glottal-opening and airflow pattern during production of voiceless fricatives: a new non-invasive instrumentation, J. Acoust. Soc. Am. 123 (2008) 3738.
[21] J. Vaissière, K. Honda, A. Amelot, S. Maeda, L. Crevier-Buchman, Multisensor platform for speech physiology research in a phonetics laboratory, J. Phon. Soc. Japan 14 (2010) 65–77.
[22] H. Kim, K. Honda, S. Maeda, ePGG, airflow and acoustic data on glottal opening in Korean plosives, in: Proc. 18th Int. Conf. of Phonetics Sciences (ICPhS), Glasgow, UK, 2015, p. 4.
[23] D. Delpy, M. Cope, Quantification in tissue near-infrared spectroscopy, Phil. Trans. R. Soc. Lond. B 352 (1997) 649–659.
[24] T. Lister, P. Wright, P. Chappell, Optical properties of human skin, J. Biomed. Opt. 17 (2012) 1–15.
[25] S. Jacques, Optical properties of biological tissues: a review, Phys. Med. Biol. 58 (2013) R37–R61.
[26] D. O'Shaughnessy, Speech Communication Human and Machine, Addison-Wesley Publishing Company, 1987.
[27] J. Cisonni, A. Van Hirtum, X. Pelorson, J. Willems, Theoretical simulation and experimental validation of inverse quasi one-dimensional steady and unsteady glottal flow models, J. Acoust. Soc. Am. 124 (2008) 535–545.
[28] J. Burnard West, Bio-engineering Aspects of the Lung, vol. 3, Marcel Dekker, Inc., 1977.

[29] T. Brancatisano, P. Collett, L. Engel, Respiratory movements of the vocal cords, J. Appl. Physiol. 54 (1983) 1269–1276.
[30] A. Scheinherr, L. Bailly, O. Boiron, A. Lagier, T. Legou, M. Pichelin, G. Caillebotte, A. Giovanni, Realistic glottal motion and airflow rate during human breathing, Med. Eng. Phys. 37 (2015) 829–839.
[31] A. Van Hirtum, X. Pelorson, High-speed imaging to study an auto-oscillating vocal fold replica for different initial conditions, Int. J. Appl. Mech. 9 (2017) 1–18.
[32] J. Haas, P. Luizard, X. Pelorson, J. Willems, Study of the effect of a moderate asymmetry on a replica of the vocal folds, Acta Acustica 102 (2016) 230–239.
[33] J. Nocedal, J. Wright, Numerical Optimization, Springer, 2006.
[34] P. Aichinger, Diplophonic Voice: Definitions, Models and Detection, Ph.D. thesis, Graz University of Technology, Austria, 2014.
[35] P. Aichinger, I. Roesner, B. Schneider-Stickler, M. Leonhard, D.M. Denk-Linnert, W. Bigenzahn, A.K. Fuchs, M. Hagmüller, G. Kubin, Towards objective voice assessment: the diplophonia diagram, J. Voice 31 (2017) 1–10.
[36] N.K. Stevens, The acoustic/articulatory interface, Acoust. Sci. Technol. 26 (2005) 410–417.