



HAL
open science

Synchronous Multimodal Measurements on Lips and Glottis: Comparison Between Two Human-Valve Oscillating Systems

Thomas Hézard, Vincent Fréour, René Caussé, Thomas Hélie, Gary P. Scavone

► To cite this version:

Thomas Hézard, Vincent Fréour, René Caussé, Thomas Hélie, Gary P. Scavone. Synchronous Multimodal Measurements on Lips and Glottis: Comparison Between Two Human-Valve Oscillating Systems. *Acta Acustica united with Acustica*, 2014, 100 (6), pp.1172-1185(14). <10.3813/AAA.918796>. <hal-01134595>

HAL Id: hal-01134595

<https://hal.science/hal-01134595v1>

Submitted on 23 Mar 2015

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.



HAL Authorization

Synchronous multimodal measurements on lips and glottis: comparison between two human-valve oscillating systems

Thomas Hézard⁽¹⁾, Vincent Fréour⁽²⁾, René Caussé⁽¹⁾, Thomas Hélie⁽¹⁾, Gary P. Scavone⁽²⁾

⁽¹⁾ Institut de Recherche et Coordination Acoustique/Musique - CNRS UMR 9912 Sciences et Technologies de la Musique et du Son - Université Pierre et Marie Curie, 1 place Igor Stravinsky, 75004 Paris, France.

{thomas.hezard, rene.causse, thomas.helie}@ircam.fr

⁽²⁾ Computational Acoustic Modeling Laboratory, Centre for Interdisciplinary Research in Music Media and Technology, Schulich School of Music, McGill University, Montréal, Québec H3A 1E3, Canada.

{vincent.freour, gary.scavone}@mcgill.ca

Summary

The brass instrument-player and the human voice production systems are both composed of a vibrating “human valve” coupled to an acoustic resonator and can be modelled by very similar dynamical systems. Moreover, lips and glottis are both difficult to access during sound production without disturbing their mechanical behaviour and vibration characteristics. In this article, we introduce a common measurement and analysis framework in order to study and compare the vibration of lips and glottis during sound production. Based on previous studies conducted on vibrating vocal folds, our measurement system is composed of three synchronous measurements –electrical admittance (electroglottography and electrolabiography), high-speed video recording and sound recording– and allows relatively non-intrusive measurements to be performed on singers and trombone players. Analysis of the collected data highlights the interpretability of electrolabiographic signals. Furthermore, similarities and differences between the two valve systems are investigated with regard to high speed imaging, electrical admittance and basic characteristics of the radiated sound.

PACS no. 43.70.Jt, 43.75.Fg, 43.75.Rs, 43.75.Yy Last revised : Monday 23rd March, 2015

1. Introduction

1.1. Context

Brass instrument-player and the human voice production systems are both composed of a vibrating “human valve” –constriction in a pipe– coupled to an acoustic resonator (Fig. 1): lips coupled to the instrument air column or vocal folds coupled to the vocal tract. In both cases, the aeroacoustic coupling is responsible for the self-oscillations of the valve system. The various sounds and regimes are obtained by modifying the resonator geometry and the mechanical properties of the constriction. Hence, similar dynamical systems involving lumped-element physical models (mass, spring and dampers) have been developed for the simulation of brass instrument and voice sounds.

The dynamics of these two structures are commonly represented by a set of three coupled equations: a mechanical equation representing the dynamics of the non-linear oscillator, a flow equation relating

the acoustic pressure drop across the lips or glottis to the volume flow generated in the upstream and downstream resonators –vocal tract and instrument bore for brass, trachea and vocal tract for voice–, and the acoustic equations simulating the effect of the resonators. Physical representations of lip-reed valves commonly consider the lips as identical and placed symmetrically against the mouthpiece, then reducing the problem to the modelling of one lip. Common lip models include parametric one-mass one-dimensional models allowing the lips to strike according to either an outward or inward mechanism [1, 2, 3]. Refined versions of lip models have been proposed including sideway striking valve models [4, 5], as well as two-dimensional models combining the effect of the pressure on the internal, external and channel faces of the lips [6], or two-mass models involving two coupled oscillators [5, 7]. Similar approaches have been studied for the vocal folds, the most common physical representation being the n -mass model, also parametrized as one or two-dimensional

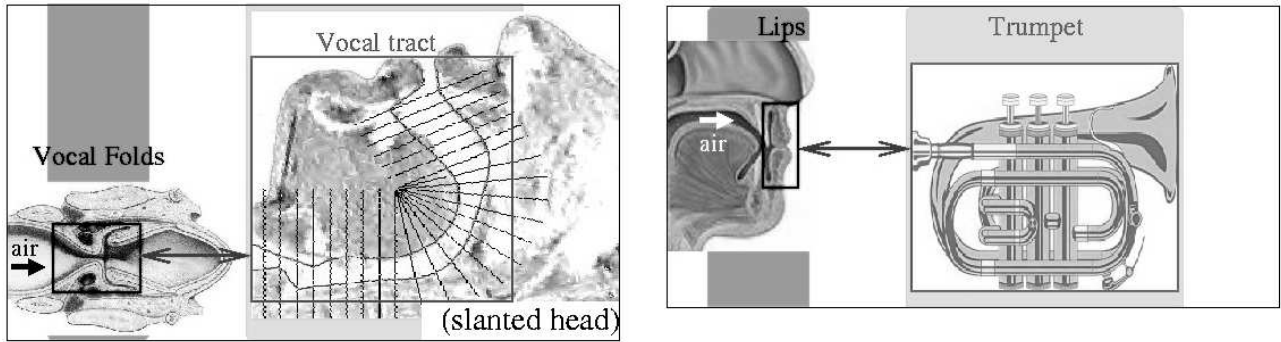


Figure 1. Schematic description of the human voice production system and the mechanism of sound production in brass instruments.

[8, 9, 10, 11, 12, 13, 14]. The pressure drop and flow equations between the lips or vocal folds can be of various complexity, from simplified Bernoulli to complete Navier-Stokes equations [15, 16, 17]. Finally, propagation inside the upstream and downstream resonators is commonly simulated using a linear model connected to radiation impedances at both ends [18, 19, 20, 21].

Brass instrument-player and the human voice production systems also share difficulties for the in-vivo exploration of the excitation mechanism. In both cases, it is difficult to monitor directly the vibration of valve oscillator without invasive methods likely to perturb the sound production mechanism.

Although they share strong similarities and can be modelled with the same class of systems, important differences exist between lips and glottis in terms of size and stiffness as it is revealed by estimations of elasticity modulus performed on both vibrating lips in brass performance [22] and vocal fold tissues [23]. Furthermore the nature of the acoustic interactions with the resonator is relatively different in voice production and brass instruments. In particular, the coupling between the lips and the air column can be very strong in brass instruments, primarily due to the geometry of the downstream resonator [5], whereas it is much more moderate in voice production. In other words, the strong impedance resonances of a brass instrument bore only allow in theory a finite number of frequencies to be produced for a given configuration of the air column.

1.2. Objectives

In light of this introductory comparison, it appears that a better understanding of similarities and differences between the two valve systems has not been the object of substantial research work, but may provide some interesting outcomes with regards to the vibroacoustic analysis and modeling of both systems. A number of vocal-fold and lip replicas, allowing in-vitro investigations on the behaviour of these artificial excitation systems with various conditions of acoustic loads (from a vocal-tract type resonator

to a brass instrument resonator) have been used by researchers for the study of excitation mechanisms in voice production and brass instrument playing [5, 24, 25, 26]. Although these systems provide good performance in reproducing the basic functioning of a human valve, and in sometimes producing some realistic sounds, they remain relatively rough approximations of the complex structure of human lips and vocal folds. In-vivo experimentation are therefore still necessary to better characterize and compare the complex characteristics of these systems. In-vivo investigations on the mechanical properties of human lips [27] or on vocal-fold vibrations [28] have been proposed using optical and LDV methods. Their implementation seems however relatively specific to one case or the other. Therefore, a measurement and analysis framework allowing in-vivo investigations of the behaviour of the lips and glottis is thus needed to be able to compare these two excitation mechanisms systematically. Several studies have been conducted on synchronous electroglottography and high-speed video endoscopy of the glottis [29, 30, 31, 32, 33, 34, 35, 36, 37, 28]. In the context of brass instruments, the concept of electroglottography has been adapted for the exploration of lip vibrations in recent works [38, 39] and high-speed video recording of vibrating lips has been made possible with transparent mouthpiece technologies [40]. In this paper a measurement framework and visualisation tool allowing synchronous display of high-speed video, electrical admittance and radiated sound for the lips and glottis is presented. A set of experimental data collected on singers and on one trombone player is analyzed, with the aim to further evaluate analogies and differences between these two systems in phonation and playing conditions.

This article is organized as follows: the method to measure simultaneously and synchronously high-speed video, electrical admittance and radiated sound for the lips and glottis is described in section 2. A visualisation tool allowing all measurements to be synchronously displayed is presented in section 3, leading to a first visual exploration

of the synchronised data. Section 4 is dedicated to the analysis of a set of measurements –the study presented here is a case study based on a small number of measurements– with respect to a set of descriptors. At first, electroglottographic and electrolabiographic signals are compared in sections 4.1 and 4.2, highlighting the interpretability of the electrolabiographic signals with regard to standard electroglottographic features such as the open quotient. Secondly, open area waveforms of the lips and glottis are compared in section 4.3 and 4.4, based on a parametric representation of glottis and lip open area waveforms introduced in [41]. Finally the geometrical characteristics of the open area, and the relationship between electrical admittance signals and open area waveforms are explored in section 4.5.

2. Experimental setup

In 1894, Manuel Garcia managed for the first time to visualize the vibrating glottis in a living human using his laryngoscope [42, 43]. The first laryngoscope inspired many other tools, including modern fibroscope and video endoscopy. In our work, we chose to record the motion of vocal folds using a high-speed video endoscope. This tool gives a direct access to the complete motion –unlike stroboscopic techniques– and, after some practice, allows the singers to sing as in natural conditions using a modern video endoscope recording the vocal-fold motion through the mouth or the nose. In section 2.2, a similar method using a standard high-speed camera and a transparent mouthpiece applied to the monitoring of lip motion is presented.

The second selected device for the non-invasive in-vivo exploration of lips and glottis is the ElectroGlottograph –EGG–, an electric device widely used by voice pathologists. The first electroglottograph was developed by Philippe Fabre in 1957 [44]. This technology is based on impedance analysis of the larynx; two electrodes are placed on each side of the throat, a low-intensity high-frequency modulated current (commonly around 2 MHz) is generated and applied to the laryngeal level, and the admittance between the two electrodes is measured. This admittance is approximately proportional to the vocal-fold contact area. A basic EGG device is schematically presented in Fig. 2-a. For a more detailed review about EGG technology, we refer the reader to [33, 34, 45]. Fréour and Scavone proposed a modified version of the electroglottograph adapted for the lips of a trombone player [38, 46], where two electrodes made of tin-plated copper foil shielding tape are glued on the rim of the mouthpiece, as shown in Fig. 2-b.

Unlike many imaging technologies (MRI, X-ray, echography), the frequency range of the EGG is sufficiently large to allow monitoring of vocal-fold

vibrations from 0.2 Hz (for swallowing) to F_{max} , where F_{max} is between 4 kHz (telephone bandwidth) and 20 kHz (audible bandwidth). The output of this measurement system is a time-domain one-dimensional signal, which does not reflect the whole spatial characteristics of vocal-fold vibratory movement, as it will be discussed in section 4.5. EGG is used in both medical and research fields. Typical applications of EGG include extraction of some vocal-fold vibration parameters and identification of laryngeal mechanisms [45, 47].

Finally, the radiated sound is recorded with a standard microphone. In the present study, it is only used for fundamental frequency estimation and qualitative assessment of the quality of the sound produced, in order to select the data to be analysed. As we’ll see in sections 2.1 and 2.2, a special attention has to be given to the synchronisation of the measurements, as high-speed video, electrical admittance and audio are not all recorded at the same sampling rate.

2.1. Detailed setup for the voice

The vocal data used in this article are extracted from the corpus “USC_2008_02” of multimodal measurements built at IRCAM by Gilles Degottex and Erkki Bianco [35]. Electroglottographic signals come from a portable electroglottograph (EG-90 from F-J Electronics) and are recorded with a sampling frequency of 44150 Hz. Vocal-fold motion is filmed in color using a rigid endoscope, which passes through the mouth, connected to a high speed camera (HreS ENDOCAM 5562) recording 4000 frames per second with a 256×256 pixel resolution. Finally, the radiated sound outside the mouth is recorded with a standard microphone at a sampling frequency of 44150 Hz. All these measurements are synchronized using the Richard Wolf’s HreS Endocam system [48]. The system specifies that the synchronization between high frame rate signals (electroglottograph and audio) and high-speed video shows a temporal precision of one video frame, that is $250 \mu s$. This delay is relatively small regarding the usual speaking frequencies of male speakers but must be considered in our study, especially when the fundamental frequency increases. Furthermore, the Shannon frequency corresponding to the high-speed video frame rate, 2000 Hz, is also relatively restrictive for the study of signals extracted from the video and must be taken into account in the choice of the data to be analyzed. Moreover, no precise information about the microphone position and the vocal tract length is available. Hence, we cannot properly synchronise the audio data to video and EGG recordings.

All the measurements considered in this paper were performed during sustained vowels. In this article, two sets of recordings are analyzed. A first set of five recordings –labeled G1 to G5– corresponds to three

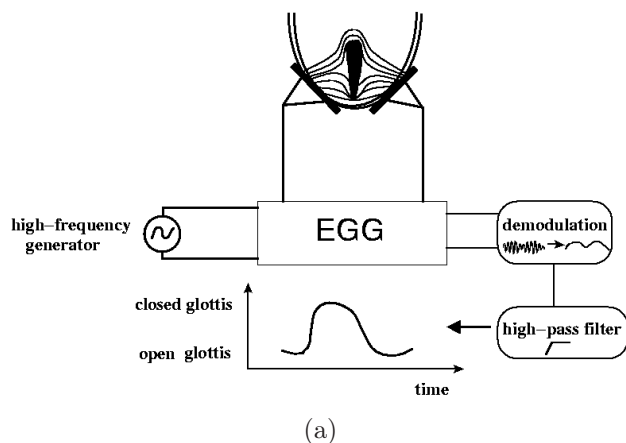


Figure 2. (a) Schematic description of an EGG device. (b) Transparent mouthpiece developed by Castellango et al. [40] equipped with copper foil shielding tape electrodes.

different speakers. These five measurements present an interesting variety of vocal qualities, although they correspond to the same French vowel / œ / in *mezzo-forte* dynamic. The three speakers are men singing at relatively low frequencies so that the spectral content of the glottal activity is captured with a relatively good accuracy regarding the camera frame rate. Unfortunately this implies that the recorded regimes were very similar (M1 mechanism). A second set –labeled G6 to G9– is also used for analysis. This data includes higher fundamental frequencies but without video recordings, as the ratios between fundamental frequencies and video frame rate is not high enough for the video analysis algorithms to be applied properly.

2.2. Detailed setup for the trombone

Video acquisitions are performed using a transparent mouthpiece developed by Castellango et al. [40]. This mouthpiece is made of a cylindrical cup closed by a flat surface perpendicular to the cylinder axis. The shank of the mouthpiece is oriented laterally allowing an open field for the camera. The mouthpiece is fixed on a stand and connected to a flexible plastic tube of the same inside diameter and length as the trombone slide in closed position. The end of the tube is connected to the bell section of the instrument. This set-up therefore allows to easily keep the mouthpiece in fixed position while maintaining the acoustical properties of the downstream resonator close to a normal instrument.

The two electrodes glued on the rim of the mouthpiece (Fig. 2-b) are connected to a commercial electroglottograph signal conditioner (Voce Vista) as proposed by Fréour and Scavone [38, 46]. This enables lip electrical admittance to be monitored during measurement without impairing the lip filming. The output of the signal conditioner is connected to an acquisition board (Qualysis USB-2533) running at 48 kHz. Prior to experiments, the latency of the EGG

signal conditioner was measured using a controllable variable resistance (FET optocoupler) mounted at the electrodes. A group delay of $180\ \mu\text{s}$ was quantified between actual variation of the electrical resistance at the electrodes and the signal conditioner output. This delay is then systematically compensated in the recorded data. In the following, the signal obtained with this method will be called ELG, standing for ElectroLabioGraph, as proposed in [38, 46].

Lip motion is captured using a high-speed camera (Qualisys Oqus 310) running at 6000 frames per second. A strong cool light source is used to compensate for the short exposure time required at this frame rate. The clock signal of the camera is shared with the USB acquisition board to guarantee synchronisation between both acquisition systems. In addition, an LED, attached to the mouthpiece and powered by a 1 Hz square wave from an external signal generator, is filmed during experiments. The voltage at the LED is simultaneously monitored by the USB-2533 acquisition board. This procedure allows for quantification of a possible small delay between the camera and acquisition board at the start of acquisition since the communication with both systems is performed with the same computer but through different ports: Ethernet for the camera and USB for the acquisition board. The radiated sound is measured using a standard studio microphone located 30 cm from the bell and connected to the USB interface.

The complete installation is presented in Fig. 3.

In this article, we report results from one player playing four tones (Bb2, Bb3, F3, and D4) labelled L1 to L4 in the following, and corresponding to the second, third, fourth and fifth resonances of the trombone input impedance. The player was instructed to play at a stable and comfortable dynamic. Results for higher tones are not presented since the lip opening area was too small to be detected and properly extracted from the video acquisitions.

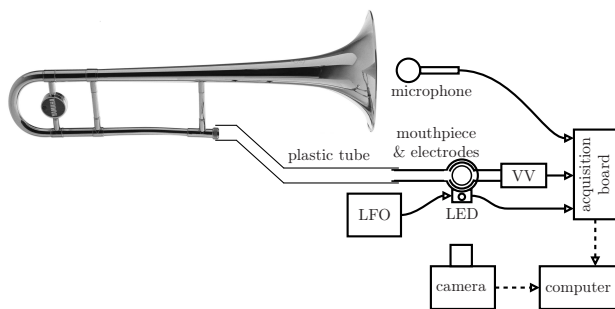


Figure 3. Data acquisition set-up for brass instruments. VV stands for Voce Vista signal conditioner.

3. Visualisation tool

In order to run some analysis algorithms indifferently on lips or glottis data, a general data organisation is needed. The following abbreviations will be used:

- **HSV** stands for High-Speed Video and refers to the video recording of lips or glottis,
- **EA** stands for Electrical Admittance and refers to the **EGG** or **ELG** signals, **DEA** stands for its temporal derivative,
- **RS** stands for Radiated Sound and refers to the audio recording of the radiated sound in both cases,
- **OA** stands for Opening Area and refers to the signal representing open surfaces (in pixels or in mm^2) between the lips or the vocal folds, **DOA** stands for its temporal derivative. **2D-OA** stands for the 2D signal representing the shape of the opening area over time. In the voice case, 2D-OA is a direct representation of the *rima glottidis* shape seen from above.

3.1. Opening area extraction

The 2D-OA signals are extracted from the HSVs using adapted algorithms. For the glottis, an algorithm from G. Degottex [49, 50] is used and give satisfying results with our measurements. In the case of lips, we developed a simple, but yet perfectly effective, algorithm based on luminance thresholding. The extracted opening areas have been visually validated for each frame of the measurements studied in this article, as in the example presented in Fig. 10. The OA waveforms are simply deduced from the 2D-OA signals by computing the number of pixels inside the estimated area on each frame. Note that, for the voice case, the distance between the camera and the glottis may vary due to the vertical motion of the larynx. Hence, neither scaling information nor the exact position of the larynx is available, and the vertical motion of the larynx is likely to modulate the size of the estimated 2D-OA. Some methods for the dynamic estimation of glottis actual size on HSVs exist (see for example [51, 52]) but none has been implemented for the USC_2008_02 corpus. For the trombone, the distance between the camera and the

lips is fixed during one measurement and the scaling factor is retrieved by detecting the number of pixels corresponding to the radius of the mouthpiece.

3.2. Synchronous display of multimodal measurements

A video tool was designed to synchronously display HSV, EA, DEA, OA, DOA and RS signals in both time and frequency domains. This tool also displays an instantaneous fundamental frequency estimation (computed with the YIN algorithm [53] on RS), the 2D-OA estimation and the temporal position. Two screen shots of these videos are shown in Fig. 4. Video samples (in color) are available on the corresponding author's website [54].

These videos allow to synchronously visualise the temporal evolution of the different signals. One can see OA and its derivative DOA along with its local magnitude spectrum, EA and its derivative DEA along with its local magnitude spectrum and RS along with its local magnitude spectrum. Signals are identified with a captioned color code. It is a useful tool to visually explore the behaviour of the lips or glottis in different regimes. One can graphically estimate standard EGG features on the EA and DEA signals, such as the open quotient or the asymmetrical coefficient. Due to a labeling error during video captures, Electrical Admittance signals have been labeled "Electrical Impedance" in the videos. To our knowledge, there is no tool allowing simultaneous and synchronous visualisation of all these measurements.

4. Analysis and comparison of the two valve systems

4.1. EA waveforms

Four representative excerpts of ELG and EGG signals are presented in Fig. 5. The four ELG examples are extracted from sustained tones with instantaneous fundamental frequency of 109.5 Hz, 165.2 Hz, 223.5 Hz and 276.6 Hz. The four EGG examples are extracted from sustained vowels with instantaneous fundamental frequency of 143.5 Hz, 191.4 Hz, 271.0 Hz and 422.3 Hz. EGG patterns can be associated with different vibration mechanisms of the vocal folds [45]. It has been shown these vocal fold vibration mechanisms can be classified into four main mechanisms, each of which referring to a given laryngeal configuration characterized by the shape of the vocal folds and a degree of muscular tension [47]. Going from mechanism M0 to M3 involves a decrease of the vibrating mass, increase of the vocal-fold tension and decrease of vocal-fold thickness. According to the literature, these changes in the valve configuration result in significant differences in the vibration patterns; while mechanisms M0 and M1 induce shorter open phases, mechanisms M2 and

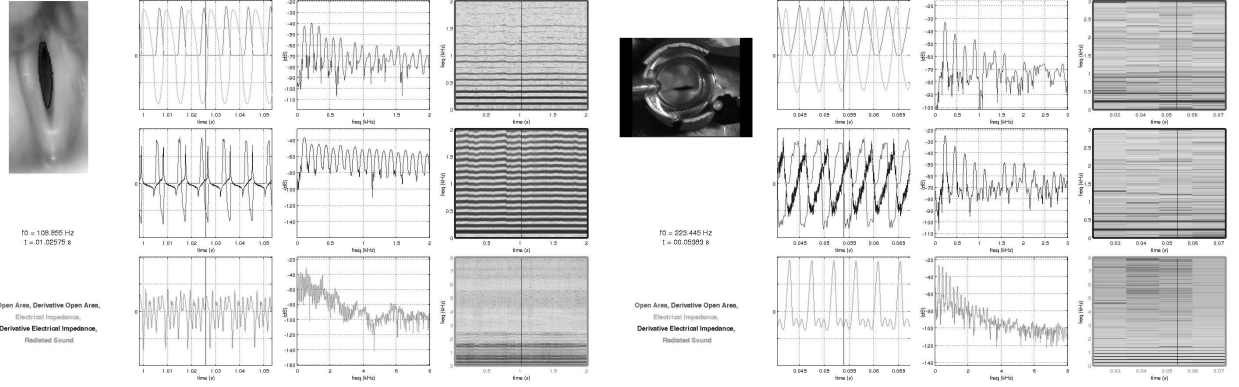


Figure 4. Synchronous display of multimodal measurements of lips and glottis. Video samples in color are available on the corresponding author’s website [54].

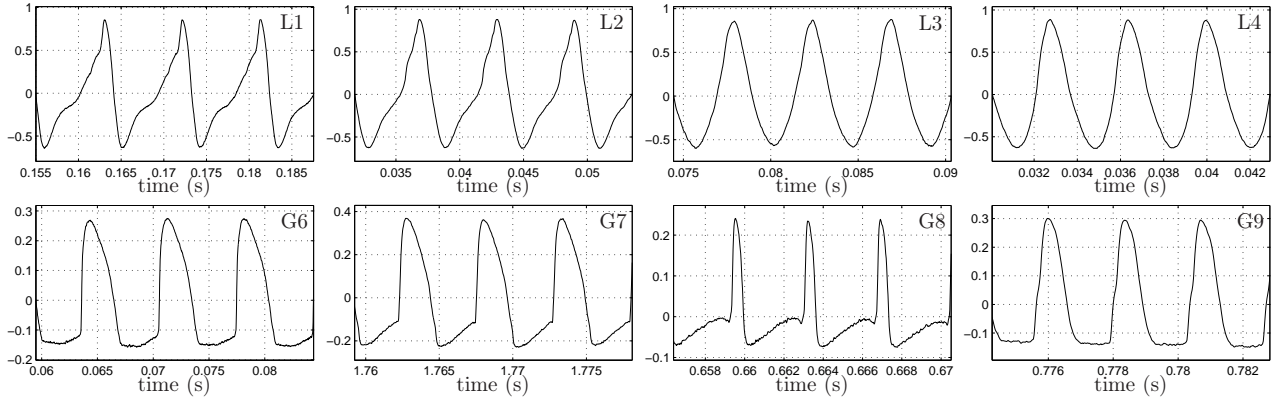


Figure 5. Extracts of EA signals of the lips (top line, instantaneous fundamental frequency of 109.5 Hz, 165.2 Hz, 223.5 Hz and 276.6 Hz from left to right) and the vocal folds (bottom line, instantaneous fundamental frequency of 143.5 Hz, 191.4 Hz, 271.0 Hz and 422.3 Hz from left to right). Vocal-fold EA signals correspond to, from left to right, a male voice in mechanism M1, a female voice in M1, a female voice in M2 and a male voice in M2. According to the definition of EA, higher values correspond to higher degree of contact between the lips or vocal folds.

M3 involve shorter closed phases. According to the convention presented in [45], the two bottom left excerpts correspond to an M1 laryngeal mechanism while the two bottom right correspond to an M2 mechanism.

Overall, ELG and EGG signals show different waveforms. The open phases in ELGs (low values of EA signals) are marked by significant variations of the signal, while EGGs remain almost constant in both M1 and M2. Moreover, the four ELG excerpts fall in two categories according to the fundamental frequency f_0 ; the two top left signals show a relatively triangular waveform, particularly asymmetrical and closer to a sawtooth for the lower pitch. The increasing phase of the signal, corresponding to an increase in contact between the lips, is characterized by an irregular evolution, while the decreasing phase is short and linear. On the contrary, the two top right signals show a rather sinusoidal and regular waveform.

Although the two lowest-frequency ELG signals do not share particular features with M1 EGG signals, the two highest-frequency ELGs show common characteristics with M2 EGGs. The pulse-like shape

characteristic of EGG in M2 is apparent in ELG signals, with a similar asymmetrical pattern particularly visible in L4 and G9.

4.2. DEA waveforms: Open Quotient and number of closing peaks

The derivative of the ELG and EGG signals presented in Fig. 5 are displayed in Fig. 6. Similarly to DEGG, DELG signals display peaks at closing and opening instants. Although these peaks are clearly visible for the two lowest tones, the change in ELG waveform for the two highest tones produce a sawtooth-like DELG waveform making the identification of closing and opening peaks in the DELG less accurate. Again, DELG signals can be categorized according to f_0 ; the two lowest tones show a similar rectangular waveform including a plateau during the open phase as observed in DEGG signals, but with non-negligible positive value, while DEGGs are approximately null during the open phase. On the contrary, the two highest tones are closer to a triangle or sawtooth waveform and the open phase is marked by important variations in DELG values. In all DELG signals, the distance

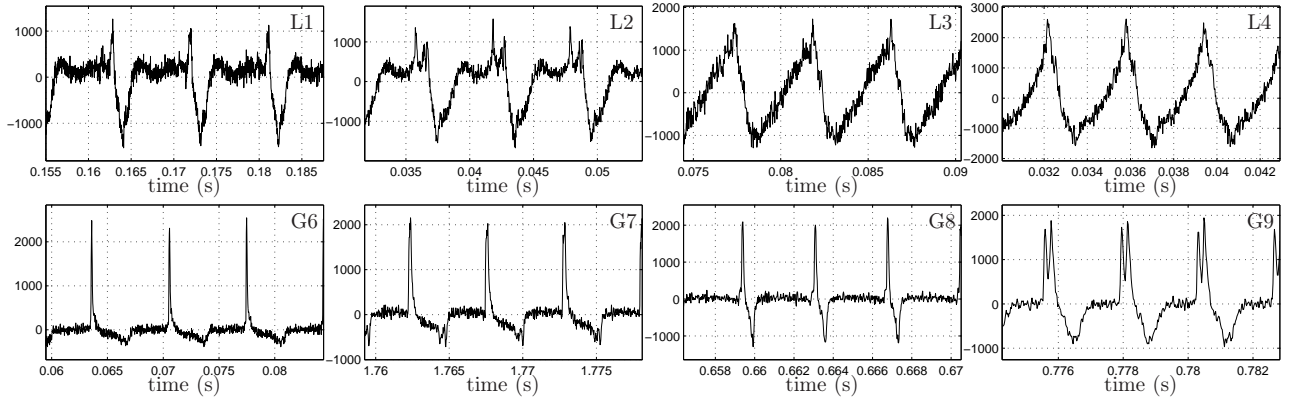


Figure 6. Electrical admittance derivative DEA signals of the lips (top line) and the vocal folds (bottom line), corresponding to EA waveforms presented in Fig. 5.

between closing and opening peaks is relatively small compared to DEGGs, suggesting shorter closed phases for the trombone player’s lips than for the singer’s glottis, as confirmed by further examination of AO signals (see section 4.3).

The double-peak feature observed in voice (clearly visible in G9, see [45]) is also present in some DELG signals, especially in L2. Instances of a triple-peak are even visible in this signal. Analogously to the voice, this attribute observed at the closing peaks suggests an irregular or sequential closing phase along the length of the lip opening area. This feature is indeed visible in the high speed video captures for L2.

The open quotient (Oq) is defined for the glottis as the ratio of the duration of the open phase to the wave period. Efficient methods have been developed to estimate the open quotient from EGG signals or from its time derivative [34, 55, 56, 57, 58, 59, 45]. In this study, the state-of-the-art DECOM algorithm [45] is used to extract Oq values from lip and glottis DEA signals. In addition, DECOM allows the number of opening and closing peaks to be estimated for each period of the DEA signals. The average Oq , the average number of closing peaks n_{cl} , as well as the average number of opening peaks n_{op} are presented in Table III for all the measurements. The use of DECOM on lip DEA signals has been validated by manually locating the peaks on DEA signals. The number of peaks and the Oq values obtained with DECOM and from manual annotation are less than 0.5% different.

Firstly, Oq values obtained for the lips are higher than for the glottis, and tend to decrease with f_0 . This may be partly explained by the fact that, unlike the vocal folds, the lips do not close and open along their thickness. These differences between both systems are in agreement with a qualitative observation of HSV recordings and Fig. 6, which thus lends support to the use of DEA for the extraction of lip Oq values.

Secondly, observations of DEA signals in Fig. 6 show that G8 (female M2) and G9 (male M2) present similar features to L1 and L2. Particularly, G9 (male

M2) show similar characteristics to L2 with regards to the double closing peaks and single opening peak. This qualitative observation is confirmed by the calculated Oq values. Indeed, Oq values in G8 and G9, corresponding to a M2 laryngeal mechanism for a female and male singer respectively, are close to the Oq values calculated for the lips, as opposed to the Oq values calculated for G1 to G7 corresponding to M1 mechanisms. This observation, together with the higher average number of closing peaks in G9, suggest a possible analogy between the vibration mechanism of the lips and a M2 laryngeal mechanism.

4.3. OA and DOA waveforms: LF parametric representation

Fig. 7 presents eight representative excerpts of OA signals from L1 to L4, G1 and G3 to G5 measurements.

As previously mentioned, these plots clearly confirm that Oq values are significantly higher for the lips than for the glottal folds. The categorisation of the lip vibration according to the fundamental frequency is confirmed by the observation of OA waveforms. Low frequency and high frequency lip OA signals tend to have on opposite asymmetry, as already seen in the EAs. Moreover, the difference in the roughness of the closure appears very clearly in the OA signals. The closing phase (transition between decreasing and null OA values) is abrupt and angular in the voice case while it is much smoother in the lips case, with the possible exception of L3. Finally, the irregular closure of lips is visible in the decreasing phase of OA signals, especially in L1 and L2. These observations should be considered with care though, because of the smaller sampling rate of the high speed cameras.

Despite the huge variability of the extracted OA waveforms, we observed that most of the signals show a rather similar periodic pulse shape. A very common pulse model used in voice production models is the Liljencrants-Fant model (LF model) [60, 61]. One

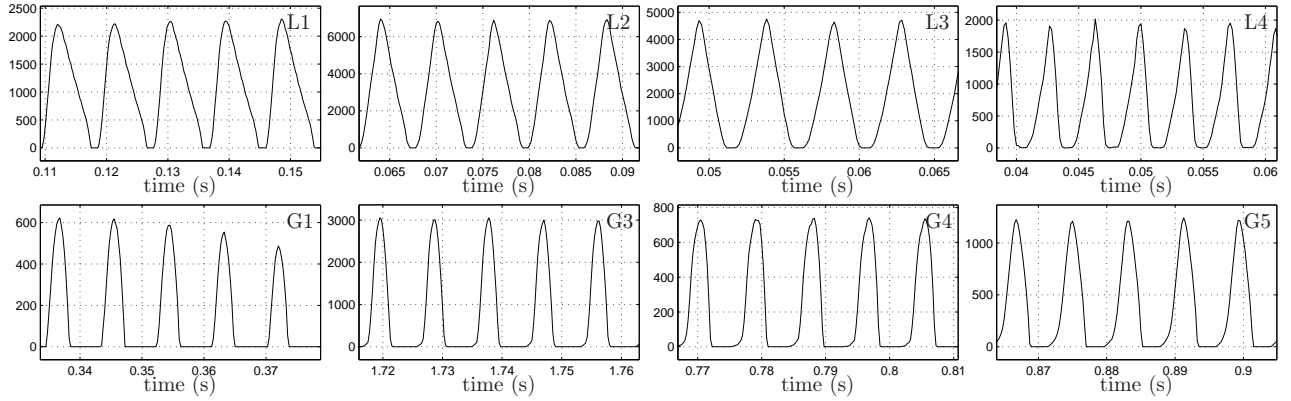


Figure 7. Open area (OA) signals of the lips (top line) and the vocal-folds (bottom line).

period of this parametric model is described in its derivative form by:

$$s'(t) = \begin{cases} s'_1(t) & 0 \leq t \leq t_e \\ s'_2(t) & t_e < t \leq t_c \\ s'_3(t) & t_c \leq t < T \end{cases}, \quad (1)$$

where

$$s'_1(t) = -E e^{\alpha(t-t_e)} \frac{\sin(\pi \frac{t}{t_p})}{\sin(\pi \frac{t_e}{t_p})}, \quad (2)$$

$$s'_2(t) = -\frac{E}{\varepsilon t_a} \left(e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c-t_e)} \right), \quad (3)$$

$$s'_3(t) = 0. \quad (4)$$

It is driven by one temporal parameter T (the period), one amplitude parameter E and four shape parameters $\{t_p, t_e, t_a, t_c\}$. Two parameters α and ε are determined by the implicit equations

$$\lim_{t \rightarrow t_e} s'_2(t) = s'_1(t_e) \quad \text{and} \quad (5)$$

$$\int_0^T s'(t) dt = 0. \quad (6)$$

Its general shape is presented on Fig. 8.

Initially conceived as a glottal flow model, this model appears to be relevant as an OA model for the glottis case, as shown in [41]. We have developed an estimation method of the LF parameters for DOA signals. The goal of this algorithm is to estimate the LF parameters which provide the best temporal estimation of a DOA target signal.

Assuming that the target signal contains N periods, the estimation algorithm of the $6N$ parameters $\{T(i), E(i), t_p(i), t_e(i), t_a(i), t_c(i))\}_{i \in [1, N]}$ consists of N steps:

- rough estimation of all the parameters,
- $1 \leq n \leq N - 1$: optimisation of the LF parameters of periods n and $n + 1$.

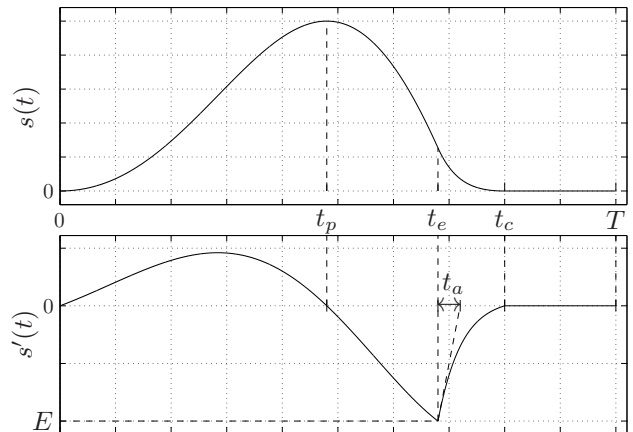


Figure 8. One period of LF model. This model has one temporal parameter T (the period), one amplitude parameter E and four shape parameters $\{t_p, t_e, t_a, t_c\}$.

The first step is achieved thanks to the event detection method described in [62]. After event detection (which determines N) has been performed, the global maximum and the global minimum are localised between each pair of adjacent events. Then, the zero crossing points are detected between adjacent minima and maxima. The set of detected minima, maxima and zero crossing points give access to a rough estimation of every LF parameter for each period. The N following optimisations are performed using the simplex method [63], the criterion being the quadratic distance between the temporal model $s'(t)$ and the DOA target within the two periods considered.

Two versions of the algorithm are proposed. The first one estimates all the LF parameters for each period of the target, the second one removes the parameter t_c , i.e. fixes $t_c = T$ or removes the exactly null phase at the end of each period. The two versions of the algorithm have been tested on a set of 10 synthetic signals of 200 periods generated by LF model with random parameters. Table I presents the results obtained with both versions.

	T_{comp}	ε	ε_E	ε_{T_0}
LF with t_c	8.37 sec	0.1569	0.1838	0.1342
LF without t_c	6.83 sec	0.1645	0.2233	0.1475

	ε_{t_p}	ε_{t_e}	ε_{d_a}	ε_{t_c}
LF with t_c	0.2793	0.2606	3.567	0.1979
LF without t_c	0.1749	0.1558	2.063	X

Table I. Performance of two versions of the LF algorithm (with and without t_c parameter) applied to synthetic signals.

	LF with t_c	LF without t_c
ε	0.2070	0.2097
T_{comp}	9.40 sec	7.37 sec

Table II. Performances of the two versions of the LF parameters estimation algorithm (complete LF model and LF model without parameter t_c). T_{comp} is the normalised computation time for one period.

- T_{comp} is the mean computation time normalised by the periods number.
- ε is the mean quadratic error on the whole set of signals.
- ε_X is the mean relative estimation error for the parameter X .

Although the mean quadratic error is slightly smaller with the t_c parameter, it appears that the estimation error on each parameter is globally better without the taking t_c into account. This may be due to a better regularisation of the estimation problem induced by removing a parameter. Moreover, the algorithm without the t_c parameter is faster. Hence, this latest version of the algorithm seems to be more appropriate for our estimation problem. Table II presents the mean quadratic error ε and computation time T_{comp} for LF parameters estimation on a set of DOA signals extracted from the corpus USC_2008_02, containing approximately 1500 periods with various voice qualities. As it can be seen in Table II, removing the t_c parameter offers significant gain in computation time for a negligible performance loss, as observed for synthetic signals. Moreover, we observed in practice that the parameter t_c is often mis-estimated. Consequently, the LF model without the parameter t_c will be considered in our study.

Figure 9 presents some DOA signals along with the results of the estimation method, and Table III indicates the quadratic error ε between the LF model obtained by the algorithm and the target DOA signals for each measurement of the two corpora. Although the LF model seems relevant to model glottis DOA signals, it does not seem appropriate when applied to lip DOAs. These results on voice data hence show promising benefits for parametric sound analysis and synthesis, as presented in [41]. However, the direct application of this type of LF pulse representation to lip DOAs remains limited in the current state of the model.

	L1	L2	L3	L4
f_0	109.1	165.3	223.5	276.6
ε	0.454	0.499	0.479	0.469
Oq	0.8521	0.7278	0.6050	0.6209
n_{cl}	1.000	3.020	1.918	1.289
n_{op}	1.167	1.102	1.721	1.087

	G1	G2	G3	G4	G5
f_0	108.9	109.5	110.3	114.1	123.4
ε	0.400	0.327	0.337	0.294	0.356
Oq	0.3967	0.4301	0.4473	0.5029	0.4319
n_{cl}	1.549	1.000	1.000	1.000	1.003
n_{op}	1.235	1.000	1.028	1.752	1.088

	G6	G7	G8	G9
f_0	143.5	191.4	271.0	422.3
Oq	0.5725	0.5976	0.8662	0.6932
n_{cl}	1.115	1.000	1.000	1.987
n_{op}	1.816	2.041	1.000	1.852

Table III. Numerical results for lips and glottis: fundamental frequency f_0 (Hz), quadratic error for LF-estimation ε , open quotient Oq , mean number of closing peaks during one closing phase n_{cl} , mean number of opening peaks during one opening phase n_{op} .

Observation of DOA signals from the database, as well as of the short representative segments in Fig. 9, highlights significant differences between lip and glottis DOA waveforms. The minima of the DOA signals, corresponding to maxima of the closing velocity, is very precise and well temporally defined in the glottis case, whereas in all lip measurements but L4, a noisy relatively flat phase appears around the minimum, corresponding to the smoother closure of the lips. This property of the lip motion may contribute to explain the poor performance of the LF model in representing lip DOA signals. This may also be emphasized by ripples in the positive phase of lip DOA signals, which cannot be represented by the proposed LF model.

Because our database does not present M2 DOA signals for the glottis due to the limitation of the camera rate, low correlation is observed between the lip and the vocal-fold DOA signals presented in Fig. 9. This is in agreement with our previous hypothesis of the lips possibly behaving differently from a M1 laryngeal mechanism and rather analogously to a M2 mechanism.

4.4. OA geometry: D and d dimensions

In order to study the characteristics of the closing phase of lips and glottis, we define the ‘‘OA corners’’ of lip and glottal openings as the position of the corner extremities of the opening area. Looking at the high speed video capture of the lips, we particularly observe that, on the contrary to the vocal folds, the position of the OA corners at the extremity of the lip opening varies along the length of the lip contact during one cycle. This attribute is not

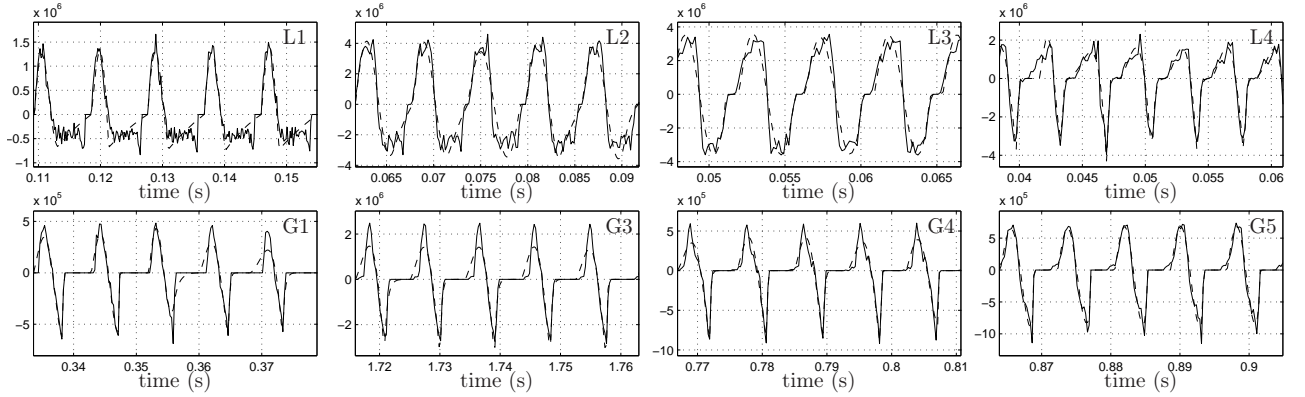


Figure 9. LF estimation examples. (—) measured DOA signal, (---) LF model (without t_c).

verified in the vocal folds whose OA corners are characterized by an almost constant position located at the extremity of the folds. In other words, the lips seem to close and open progressively along the lip length (zipper-like closure), while the whole length of the vocal folds tends to close and open within a much shorter time frame. This contrast in the opening mechanism between the two valve systems may contribute to explain the differences observed between ELG and EGG waveforms, particularly with respect to the characteristics of ELG signals during the open phase.

To further investigate the differences of opening and closing mechanisms between the two valve systems, two geometrical parameters derived from the 2D-OA signals are introduced. These quantities allow the main characteristics of the 2D-OA to be quantified. For each frame, the smallest rectangular box containing the detected opening area is computed and the two dimensions of the box are extracted. The largest dimension of the box is denoted D and the smallest dimension is named d as illustrated in Fig. 10.

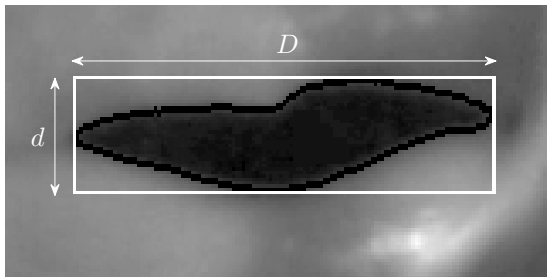


Figure 10. Geometrical parameters D and d on a lip frame. Detected 2D-AO is represented in black.

Parameters D and d are represented as a function of OA for two lip signals (L1 and L3) and two glottis signals (G2 and G5) in Fig. 11. This representation allows the influence of the two main dimensions of the opening area to be evaluated with respect to OA.

Particularly, these figures are useful tools in order to detect zipper-like closures, as well as abrupt closures occurring on the whole length of the lips or vocal folds.

The relationship $OA \mapsto d$ is relatively linear for both systems. On the contrary, the $OA \mapsto D$ shows a two-linear-segment shape, a first short segment of high slope and a second segment of much smaller slope. This feature is particularly emphasized in the glottis signals (G2 and G5) and may be explained by an abrupt opening and closing of the vocal folds with no gradual opening nor closing along the length of the vocal folds. Indeed, during the opening and closing phases, the two vocal folds move rapidly toward or away from each other, as depicted by the first short segment of high slope in $OA \mapsto D$. During the open phase, the vocal folds are not in contact and the variation of OA is mainly caused by variations in d , as shown by the second almost constant segment in $OA \mapsto D$.

This behaviour is not as striking for the lips (L1 and L4). In lowest tones (L1), the relationship $OA \mapsto D$ tends to show a two-segment structure similar to the one from glottis measurements. However, for the highest tones (L4), the $OA \mapsto D$ and $OA \mapsto d$ relationships are closer, indicating a balanced contribution of D and d to the opening area OA, possibly correlated to the zipper-like opening and closure of the lips (moving lip corners). Note that this logarithmic shape of the $OA \mapsto d$ relation is in agreement with previous quantitative analysis of the relationship between both variables observed in artificial and real trombone players' lips [64].

4.5. Relationship between OA and EA

In this last section, we aim to further concentrate on the relationship between OA and EA signals. The values of OA are represented as a function of EA along several cycles of vibrations for the lips and glottis in Fig. 12.

As predicted by $OA \mapsto D$ and $OA \mapsto d$ relationships, $EA \mapsto OA$ relations are marked by strong hysteresis behaviours, apparently uncorrelated

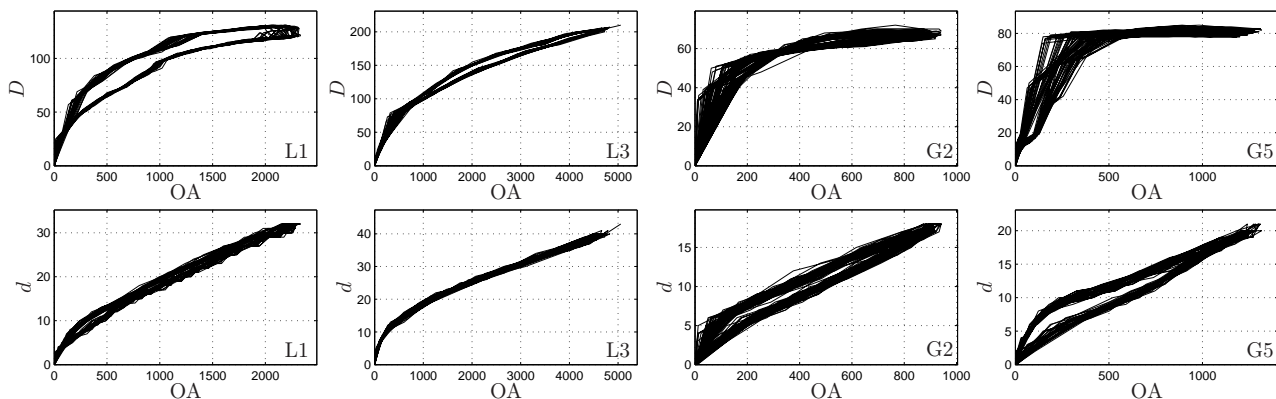


Figure 11. Opening Area OA versus geometrical parameters D (top line) and d (bottom line) for lips (2 left columns) and glottis (2 right columns). OA, D and d are in pixels.

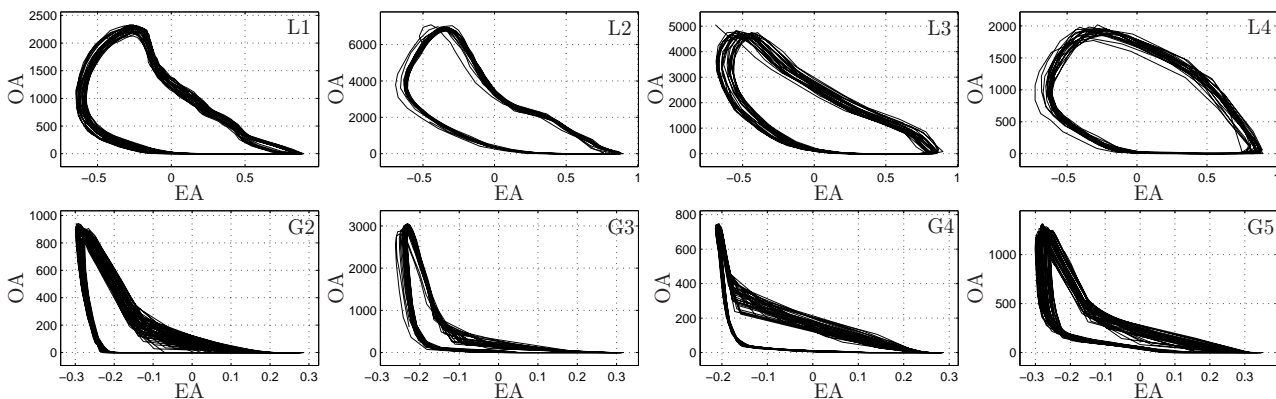


Figure 12. Electrical admittance EA versus opening area OA. Top line: lips, bottom line: glottis. EA is normalised, OA is in pixels.

with the playing frequency and particularly emphasized in the case of the lips. The two-step sequential relationship $OA \mapsto D$ observed for the vocal folds in Fig. 11 is clearly visible in the glottis (EA,OA) representations. It indicates a predominant influence of the dimension D on EA for the glottis. Unlike the vocal-fold results, this predominant influence of one geometrical dimension on the EA signal is not apparent for the lips. Nevertheless, this observation is in agreement with previous evidence of a zipper-like lip closure and opening of the lips (moving corners).

This figure also illustrates that, in the case of the lips, the instant of maximum in OA does not correspond exactly with a minimum in EA, which suggests that despite a maximum in opening area, the degree of contact between of the lips is not necessarily minimum. It also highlights the lack of information provided by the EA signal during the open phase. On the contrary, the steadiness of the OA corners for the vocal folds tends to reduce the hysteresis and to establish a more systematic synchronicity between maximum of the OA and minimum of the EA. This observation raises the question of the relevance of the DECOM algorithm for the Oq estimation in the case of the lips. Although closing and opening peaks signal

are properly identified by the DECOM algorithm from the lip DEA signal, Fig. 12 reveals that OA and EA signals are correlated which each other to a lesser extent in case of the lips. The lip Oq values obtained with the DECOM algorithm case should therefore be confronted to other Oq estimation methods based on OA signals.

Furthermore, Fig. 12 confirms that the EA signal brings information during the completely closed phase that cannot be interpreted with the present tools. We strongly believe that this information is precisely related to the zipper-like opening of the vocal folds along their thickness and possibly to an outward motion of the lips typically observed in brass instruments [6, 7, 65, 22].

5. Conclusion

An in-vivo and synchronous measurement and analysis framework was presented for the comparison of the vibratory mechanisms in oscillating vocal folds and lips. A visualisation tool was developed to allow the synchronous examination of the multimodal data. Analyses of the experimental data was conducted with respect to : 1) EI and DEI waveforms and Oq values,

2) parametric LF modelling of the DOA signals, 3) relative influence of the geometrical parameters d and D on OA waveforms, 4) relationship between OA and EI signals. This analysis framework leads to the following conclusions:

(C1) The transposition of the EGG device to the ELG device for monitoring the lips of a brass instrument player by Fréour and Scavone [38, 46] proved to be efficient and able to deliver interpretable signals with respect to standard EGG descriptors such as the open quotient Oq .

(C2) Analogously to the vocal folds, EI signals from the lips can be categorised according to their fundamental frequency. They show more complex waveforms for the lowest tones, while displaying more sinusoidal shapes in the highest ones (such as a M2 glottal mechanism), corroborating previous evidence of a more sinusoidal motion of the lips with increasing pitch [66, 65, 22, 64]. Moreover, lip EI signals show much activity during the open phase, while EGG signals are rather constant. This could be related to emphasized variations in OA corner positions during opening and closing for the lips.

(C3) The differentiation observed in lip EI signals according to the playing frequency is well visible in DEI waveforms; a transition from rectangular to triangular waves is observed with increase in pitch. Analogously to the glottis, lip DEI signals show closing and opening peaks, especially visible at lower frequencies and including occurrences of multiple closing and opening peaks. Observations of DELG waveforms also show more similarities with DEGG in a M2 laryngeal mechanism. Analogously to classical analysis of DEGG signals, we propose to compute the open quotient Oq of the vibrating lips based on recorded DELG signals. Due to the frequent occurrence of multiple opening and closing peaks in lip DEI signals, classical Oq estimators [67, 68, 69] are not able to provide a proper Oq estimations for the lips. Nevertheless, the DECOM algorithm [45] is able to detect and enumerate the multiple opening and closing peaks in DEI signals and take them into account for the estimation of Oq . Thus, this latter algorithm allows to properly estimate the open quotient for either EGG or ELG signals. The Oq values obtained for the lips fall within the same range of values than Oqs observed for the vocal folds in a M2 laryngeal mechanism. This further suggest a possible analogy between the vibration mechanisms of these two systems.

(C4) A major difference which appears between lips and glottis vibration behaviours in all the signals, is the difference of roughness in the closure. This can be seen especially in DEA and DOA signals where the long null phases of glottis signals have no equivalent in lips signals. This observation is confirmed by the analysis of the 2D-OA signals, where the two-segment structure of the glottis (D,OA)

curves comes from the abrupt closure of the vocal folds, while the smooth closure of the lips may induce smooth (D,AO) and (d,OA) curves. As a consequence, the LF pulse model show poor agreement with lip OA signals, whereas it appears as a satisfactory parametric model for glottis AO signal, which brings interesting perspectives for the analysis and synthesis of voice sounds [41]. Regarding brass instruments, the ability of a LF model to efficiently represent the flow at the lips (derived from OA and the quasi-static pressure difference across the lips) could be the object of further investigations.

(C5) Geometrical parameters extracted from OA signals reveal an overall linear correlation between the large opening dimension D and OA in both lips and glottis. On the contrary, the influence of the small opening dimension d on OA is more emphasized in the lips than in the glottis. This is in agreement with variations in the position of OA corners occurring for the lips and in a much lesser extent for the glottis. This may contribute to explain the relatively low performance of the LF model applied to the lip DOA signals.

(C6) The representation of OA against EI makes appear a strong hysteresis behaviour and a particular asynchrony between the maximum in OA and minimum in EI for the lips. Although the reasons of this asynchrony remain unclear, we may hypothesize that this is partly related to the variation of lip corner position during the opening and closing phases: while the lip opening area is still increasing the nature and surface of the contact between the lips is varying in way that makes the conductance between the lips decrease.

(C7) Despite the interest of multimodal measurements in the context of this study, the main limitation in the exploration of the lips and glottis is the impossibility to monitor the motion within the thickness dimension, perpendicular to the lip or glottis plane. Although EI signals provide some information regarding the evolution of the contact during the closed phase, more information is needed to fully characterized the motion of the lips and glottis when they are in contact with each other. A new multichannel EGG/EIT device [70] is currently in development and we have good hope that this equipment will give access to the characteristics this motion.

Finally, the visualization tool presented in this paper presents some obvious interests for medical uses, as well as for vocal and brass-instrument pedagogy. Some developments in the context of a medical and/or pedagogical platform based on some elements of the proposed analysis framework would certainly open interesting perspectives for practitioners in the voice and musical community.

Acknowledgements

The authors would like to thank Gilles Degottex and Erkki Bianco for their kind authorisation to access and study the corpus USC_2008_02. They wish to thank Nathalie Henrich for her precious comments and advice on EA signal analysis and for her numerous bibliographic references, which led to a significant improvement of this work.

References

- [1] H. L. F. Helmholtz: On the sensations of tone. Dover, New York, 1954.
- [2] S. Adachi, M.-A. Sato: Time-domain simulation of sound production in the brass instrument. *The Journal of the Acoustical Society of America* **97** (1995) 3850—3861.
- [3] M. Campbell: Brass instruments as we know them today. *Acta Acustica united with Acustica* **90** (2004) 600–610.
- [4] N. H. Fletcher: Autonomous vibration of simple pressure-controlled valve in gas flows. *The Journal of the Acoustical Society of America* **93** (1993) 2172–2180.
- [5] J. Cullen, J. Gilbert, D. Campbell: Brass instruments: Linear stability analysis and experiments with an artificial mouth. *Acta Acustica united with Acustica* **86** (2000) 704–724.
- [6] S. Adachi, M.-A. Sato: Trumpet sound simulation using a two-dimensional lip vibration model. *The Journal of the Acoustical Society of America* **99** (1996) 1200–1209.
- [7] O. Richards: Investigation of the lip reed using computational modelling and experimental studies with an artificial mouth. Dissertation. University of Edinburgh, Edinburgh, United Kingdom, 2003.
- [8] J. L. Flanagan, L. Landgraf: Self-oscillating source for vocal-tract synthesizers. *Audio and Electroacoustics, IEEE Transactions on* **16** (1968) 57–64.
- [9] K. Ishizaka, J. L. Flanagan: Synthesis of voiced sounds from a two-mass model of the vocal chords. *Bell System Technical Journal* **51** (July-August 1972) 1233–1368.
- [10] I. R. Titze: The human vocal cords: A mathematical model. i. *Phonetica* **28** (1973) 129–170.
- [11] I. R. Titze: The human vocal cords: A mathematical model. ii. *Phonetica* **29** (1974) 1–21.
- [12] X. Pelorson, A. Hirschberg, R. R. van Hassel, A. P. J. Wijnands, Y. Auregan: Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. application to a modified two-mass model. *The Journal of the Acoustical Society of America* **96** (1994) 3416–3431.
- [13] J. Horáček, P. Šidlof, J. G. Švec: Numerical simulation of self-oscillations of human vocal folds with hertz model of impact forces. *Journal of Fluids and Structures* **20** (2005) 853–869.
- [14] L. Cveticanin: Review on mathematical and mechanical models of the vocal cord. *Journal of Applied Mathematics* **2012** (2012) 18.
- [15] J. D. Markel, A. H. Gray: Linear prediction of speech. 1976, 288.
- [16] C. Vilain, X. Pelorson, C. Fraysse, M. Deverge, A. Hirschberg, J. Willems: Experimental validation of a quasi-steady theory for the flow through the glottis. *Journal of Sound and Vibration* **276** (2004) 475–490.
- [17] P.-Y. Lagrée, E. Berger, M. Deverge, C. Vilain, A. Hirschberg: Characterization of the pressure drop in a 2d symmetrical pipe: Some asymptotical, numerical, and experimental comparisons. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik* **85** (2005) 141–146.
- [18] J. L. Kelly, C. C. Lochbaum: Speech Synthesis. Proceedings of the 4th International Congress on Acoustics, 1962, 1–4.
- [19] S. Maeda: A digital simulation method of the vocal-tract system. *Speech Communication* **1** (1982) 199–229.
- [20] B. H. Story: A parametric model of the vocal tract area function for vowel and consonant simulation. *The Journal of the Acoustical Society of America* **117** (2005) 3231–3254.
- [21] P. Eveno, J.-P. Dalmont, R. Caussé, J. Gilbert: Wave propagation and radiation in a horn: Comparisons between models and measurements. *Acta Acustica united with Acustica* **98** (2012) 158–165.
- [22] S. Yoshikawa, Y. Muto: Lip-wave generation in horn players and the estimation of lip-tissue elasticity. *Acta Acustica united with Acustica* **89** (2003) 145–162.
- [23] S. Kazemirad, H. Bakhshae, L. Mongeau, K. Kost: Non-invasive in vivo measurement of the shear modulus of human vocal fold tissue. *Journal of Biomechanics* **47** (2014) 1173–1179.
- [24] T. Hélie, N. Lopes, R. Caussé: Open-loop control of a robotized artificial mouth for brass instruments. *The Journal of the Acoustical Society of America* **131** (2012) 3470.
- [25] N. Ruty: Modèles d’interactions fluide/parois dans le conduit vocal. Applications aux voix et aux pathologies. Dissertation. Insitut National Polytechnique de Grenoble, 2007.
- [26] L. J. Chen, L. Mongeau: Verification of two minimally invasive methods for the estimation of the contact pressure in human vocal folds during phonation. *The Journal of the Acoustical Society of America* **130** (2011) 1618–1627.
- [27] M. J. Newton, M. Campbell, J. Gilbert: Mechanical response measurements of real and artificial brass players lips. *The Journal of the Acoustical Society of America* **123** (2008).
- [28] A. Chan, L. Mongeau, K. Kost: Vocal fold vibration measurements using laser doppler vibrometry. *The Journal of the Acoustical Society of America* **133** (2013) 1667–1676.
- [29] T. Baer, A. Löfquist, N. S. McGarr: Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques. *The Journal of the Acoustical Society of America* **73** (Apr. 1983) 1304–1308.
- [30] D. G. Childers, J. M. Naik, J. N. Larar, A. K. Krishnamurthy, G. P. Moore: Electroglottography, speech and ultra-high speed cinematography. *Vocal fold physiology and biophysics of voice* (1983) 202–220.

- [31] D. G. Childers, J. N. Larar: Electroglottography for laryngeal function assessment and speech analysis. *Biomedical Engineering*, IEEE Transactions on **BME-31** (1984) 807–817.
- [32] D. G. Childers, A. M. Smith, G. P. Moore: Relationships between electroglottograph, speech, and vocal cord contact. *Folia Phoniatica et Logopaedica* **36** (1984) 105–118.
- [33] D. G. Childers, A. K. Krishnamurthy: A critical review of electroglottography. *Critical Reviews in Biomedical Engineering* **12** (1985) 131–161.
- [34] D. G. Childers, D. M. Hicks, G. P. Moore, L. Eskenazi, A. L. Lalwani: Electroglottography and vocal fold physiology. *Journal of Speech and Hearing Research* **33** (1990) 245–254.
- [35] G. Degottex, E. Bianco, X. Rodet: Usual to particular phonatory situations studied with high-speed videendoscopy. The 6th International Conference on Voice Physiology and Biomechanics, ICVPB, Tampere, Finland, August 2008, 19–26.
- [36] M. Golla, D. Deliyski, R. Orlikoff, H. Moukalled: Objective comparison of the electroglottogram to synchronous high-speed images of vocal-fold contact during vibration. *Proceedings of the 6th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, 2009, M. C. (ed.), 141–144.
- [37] S.-Z. Karakozoglou: Glottal source analysis: a combinatory study using high-speed videendoscopy and electroglottography. Diploma Thesis. University of Paris-Sud XI, Computer Science Department - University of Crete, Computer Science Department - LIMSI-CNRS, Audio & Acoustics Group, June 2010.
- [38] V. Fréour, G. P. Scavone: Development of an electrolabiograph embedded in a trombone mouthpiece for the study of lip oscillation mechanisms in brass instrument performance. *Journal of the Canadian Acoustical Association* **39** (2011) 130–131.
- [39] V. Fréour, G. P. Scavone: Investigation of the effect of upstream airways impedance on regeneration of lip oscillations in trombone performance. *Proceedings of the Acoustics 2012 conference*, Nantes, France, 2012, 2225–2230.
- [40] M. Castellengo, R. Caussé, B. Sluchin: Étude acoustique de l'émission multiphonique aux cuivres. *Proceeding of the 11th International Congress on Acoustics*, Paris, France, 1983, 355–357.
- [41] T. Hézard, T. Hélie, R. Caussé, B. Doval: Analysis-synthesis of vocal sounds based on a voice production model driven by the glottal area. *Proceedings of Acoustics 2012*, April 2012, 1–5.
- [42] M. Garcia: Observations on the Human Voice. *Proceedings of the Royal Society of London*, Jan. 1854, 399–410.
- [43] M. Garcia: Hints on singing. London: E. Ascherberg, New York: E. Schuberth, 1894.
- [44] P. Fabre: Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation: Glottographie de haute fréquence. *Bulletin de l'Académie Nationale de Médecine* (1957) 66–69.
- [45] N. Henrich, C. D'alessandro, B. Doval, M. Castellengo: On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *Journal of the Acoustical Society of America* **115** (March 2004) 1321–32.
- [46] V. Fréour, G. P. Scavone: Acoustical interaction between vibrating lips, downstream air column, and upstream airways in trombone performance. *The Journal of the Acoustical Society of America* **134** (2013) 3887–3898.
- [47] B. Roubeau, N. Henrich, M. Castellengo: Laryngeal vibratory mechanisms: The notion of vocal register revisited. *Journal of Voice* **23** (2009) 425–438.
- [48] www.richard-wolf.com/en/human-medicine/ent-ork/products/phoniatrics.html.
- [49] G. Degottex, E. Bianco, X. Rodet: Estimation of glottal area with high-speed videendoscopy. *Speech Production Workshop: Instrumentation-based approach*, ParisIII/ILPGA, Paris, France, July 2008.
- [50] G. Degottex: Glottal source and vocal-tract separation. Dissertation. Université Pierre et Marie Curie (UPMC, Paris VI), 2010.
- [51] S. Kazemirad, H. Bakhshae, L. Mongeau, K. Kost: Non-invasive in vivo measurement of the mechanical properties of human vocal fold tissue. *Proceedings of Meetings on Acoustics*, 2013, A. S. of America (ed.), 1–8.
- [52] M. Kob, T. Frauenrath: A system for parallel measurement of glottis opening and larynx position. *Biomedical Signal Processing and Control* **4** (July 2009) 221–228.
- [53] A. de Cheveigne, H. Kawahara: Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America* **111** (2002) 1917–1930.
- [54] <http://anasynth.ircam.fr/home/people/thomashezard>.
- [55] D. G. Hanson, B. R. Gerratt, G. S. Berke: Frequency, intensity, and target matching effects on photoglottographic measures of open quotient and speed quotient. *J Speech Hear Res* **33** (1990) 45–50.
- [56] F. L. E. Lecluse: Elektrogloottografie - Een experimenteel onderzoek betreffende de elektrische impedantie van het mannelijkstrottehoofd. Dissertation. Universiteit te Rotterdam, 1977.
- [57] F. L. E. Lecluse, M. Brocaar: Quantitative measurements in the electroglottogram. *17th International Congress of Logopedics and Phoniatrics*, 1977.
- [58] D. G. Miller, J. G. Švec, H. K. Schutte: Measurement of characteristic leap interval between chest and falsetto registers. *Journal of Voice* **16** (2002) 8–19.
- [59] R. F. Orlikoff: Assessment of the dynamics of vocal fold contact from the electroglottogram: Data from normal male subjects. *Journal of Speech and Hearing Research* **34** (1991) 1066–1072.
- [60] G. Fant, J. Liljencrants, Q. Lin: A four-parameter model of glottal flow. *STL-QPSR* **26** (1985) 1–13.
- [61] G. Fant: The LF-model revisited. *Transformations and frequency domain analysis*. *STL-QPSR* **36** (1995) 119–156.
- [62] H. Kawahara, Y. A., P. Z.: Accurate vocal event detection method based on a fixed-point analysis of mapping from time to weighted average group delay. *Proceedings of the 6th International Conference on Spoken Language Processing, ICSLP 2000 / INTERSPEECH 2000*, Beijing, China, 2000, 664–667.

-
- [63] J. Lagarias, J. Reeds, M. Wright, P. Wright: Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on Optimization* **9** (1998) 112–147.
- [64] S. Bromage, M. Campbell, J. Gilbert: Open areas of vibrating lips in trombone playing. *Acta Acustica united with Acustica* **96** (2010) 603–613.
- [65] D. C. Copley, W. J. Strong: A stroboscopic study of lip vibrations in a trombone. *The Journal of the Acoustical Society of America* **99** (1996) 1219–1226.
- [66] D. W. Martin: Lip vibrations in a cornet mouthpiece. *The Journal of the Acoustical Society of America* **13** (1942) 305–308.
- [67] M. Rothenberg, J. J. Mahshie: Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech and Hearing Research* **31** (1988) 338–351.
- [68] D. M. Howard, G. A. Lindsey, B. Allen: Toward the quantification of vocal efficiency. *Journal of Voice* **4** (1990) 205–212.
- [69] D. M. Howard: Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. *Journal of Voice* **9** (1995) 163–172.
- [70] T. Hézard, T. Hélie, B. Doval, N. Henrich, M. Kob: Non-invasive vocal-folds monitoring using electrical imaging methods. *100 years of electrical imaging, Paris, 2012*, 145–148.