Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Musical Acoustics
Session 3aMU: Aeroacoustics of Wind Instruments and Human Voice II

3aMU6. Synchronous visualization of multimodal measurements on lips and glottis: Comparison between brass instruments and the human voice production system.

Thomas Hézard*, Vincent Fréour, René E. Caussé, Thomas Helie and Gary P. Scavone

***Corresponding author's address: IRCAM - CNRS UMR 9912 - UPMC, 1, place Igor Stravinsky, PARIS, 75004, France, France, thomas.hezard@ircam.fr**

Brass instruments and the human voice production system are both composed of a vibrating "human valve" (constriction in a pipe) coupled to an acoustic resonator: lips coupled to the brass instrument or vocal folds coupled to the vocal tract. In both cases, the aeroacoustic coupling is responsible for the self-oscillations and a large variety of regimes. Additionally, brass instruments and voice share difficulties for the in-vivo measurement of the exciter activity. Hence, the development of a common tool is relevant. It is also relevant to explore the effect of some known differences between these systems, namely, the strength of the coupling and the physiological characteristics. This paper introduces components for the development of such a tool. First, two corpuses of multi-modal measurements are presented: one for a singer's larynx during sustained vowels, one for a musician's lips during sustained notes. They include high-speed-video recordings, electrical impedance measurements and audio recordings. Then, we introduce two estimation algorithms: one of the opening area waveforms from videos, one of the LF-model parameters on these waveforms. Moreover, we build a video tool displaying, synchronuously, these signals. Finally, this tool is exploited to exhibit common behaviors and relevant differences between brass instruments and human voice.

Published by the Acoustical Society of America through the American Institute of Physics

©2013 Acoustical Society of America [DOI: 10.1121/1.4799951]
Received 22 Jan 2013; published 2 Jun 2013

MULTI-MODAL MEASUREMENTS ON LIPS AND VOCAL FOLDS

Brass instruments

Acquisition were performed using a transparent mouthpiece developed by Castellango et al. [1]. 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 open field for lip visualization. The mouthpiece was fixed on a stand and connected to a flexible plastic tube of same inside diameter and length than the trombone slide in closed position. The end of the tube was connected to the bell section of the instrument. This setup therefore enabled to easily maintain the mouthpiece in fixed position.

Two electrodes made of tin-plated copper foil shielding tape were glued on the rim of the mouthpiece and connected to a commercial electroglottograph signal conditioner (Voce Vista) as proposed by Freour and Scavone [2]. This enables monitoring of lip electrical impedance during measurement without impairing the lip filming. The output of the signal conditioner was 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 $180\,\mu s$ was quantified between actual variation of the electrical resistance at the electrodes and the signal conditioner output. This delay was then taken into account in the analysis.

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

The complete installation is presented in figure 1.

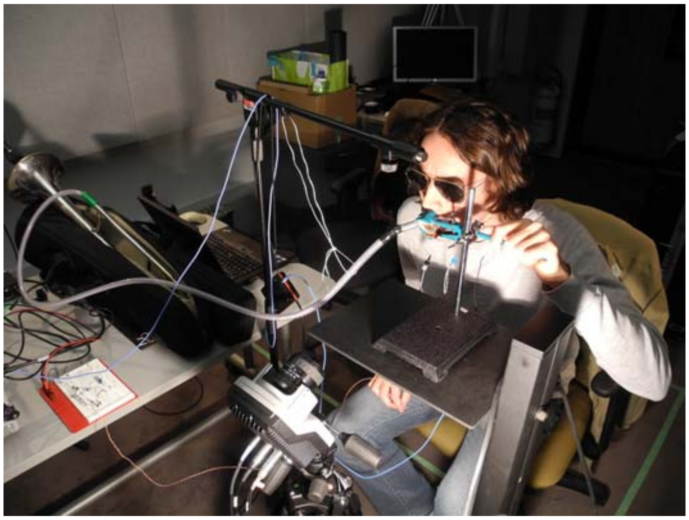

FIGURE 1: Data acquisition set-up for brass instruments

In this article, we report results from one player playing four tones (Bb2, F3, Bb3 and D4) corresponding to the second, third, fourth and fifth peaks of the trombone input impedance. Results for higher tones are not presented since the lip opening area was too small to be properly detected and extracted from video acquisitions.

Human voice production system

The data exhibited in this article are extracted from a corpus of multimodal measurements called "USC_2008_02" and built at IRCAM by Gilles Degottex and Erkki Bianco. Each recording is composed of three synchronous measurements. Electroglottograph signal came from a portable electroglottograph (EG-90 from F-J Electronics) and was recorded with a sampling frequency of 44150 Hz. Vocal folds motion was filmed using a high speed camera (HreS ENDOCAM 5562) running at a frame rate of 4000 frames per second for a 256×256 pixels resolution. Finally, the radiated sound outside the mouth was recorded with a standard microphone at a sampling frequency of 44150 Hz. All these measurements were synchronized using the Richard Wolf's HreS Endocam system. Unfortunately, the system specifies that the synchronization shows a temporal precision of one video frame, that is 250 μs. Nevertheless, this delay is relatively small regarding the usual speaking frequencies of male speakers.

All the measurements were performed during sustained vowels. In this article, we introduce five clean and interesting recordings corresponding to three different speakers. As we'll see in the following, these five measurements present an interesting variety of vocal qualities, although they correspond to the same French vowel /œ/. The three speakers are men speaking at relatively low frequencies so that the spectral content of the glottal activity is captured well enough regarding the camera frame rate. Unfortunately this implies that the recorded regimes were all similar.

A COMMON TOOL FOR MEASUREMENTS ANALYSIS: FEATURES EXTRACTION AND VIDEO TOOL

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

- **(HSV)** stands for High-Speed Video and refers to the video recording of lips or glottis,
- **(EI)** stands for Electrical Impedance and refers to the electroglottograph signal in both cases, (DEI) stands for its 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 the area between the lips or to the glottis, (DOA) stands for its derivative.

Opening area extraction and LF parameters estimation

The 2D opening areas are extracted from the (HSV) using adapted algorithms. In the case of vocal fold motion, an algorithm from G. Degottex has been used [3]. In the case of lips, we developed a simple, but yet perfectly effective, algorithm based on a frame by frame luminance thresholding. These two algorithms give us accurate estimation of the 2D opening areas over time. 1D (OA) signals are simply deduced from these 2D opening areas by computing the number of pixels inside the estimated area on each frame.

As one can see in figure 4 (Open Area signals), (OA) signals show a general pulse shape in both cases. As shown in [4], the Liljencrants-Fant (LF) model [5] can be relevant not only as a glottal pulse model but also as an opening area waveform model. Therefore, a parametric estimation of the (AO) signals should give us an interesting tool to compare lips and glottis activity during sustained sounds. The estimation of LF parameters is conducted as follows: (1) period detection and LF parameters initialization for each period by detection of maxima, minima and zero crossings on the signal, (2) from first to penultimate period: optimization of LF parameters for periods n and $n+1$ with the simplex method [6].

Figure 2 presents two results of LF parameters estimation on (DOA) signals, one for a lip measurement, one for a glottis measurement. As one can see, the estimation is much more accurate in the case of voice. This is confirmed by the results presented in table 1. The quadratic error $\epsilon = \frac{|\widetilde{DOA}-DOA|}{|DOA|}$, where DOA stands for the measured signal and $\widetilde{DOA}$ stands for the LF model, is significantly higher is the case of trombone. This means that the LF model is too restrictive for the (DOA) waveform we are studying. If LF model is overall relevant for the (DOA) signals in the voice, another model has to be considered in the case of the lips. Nevertheless, it is observed that, for high pitches, (DOA) signal tends to have a shape closer to the canonical LF shape. This result can be seen in table 1, considering that the second column is less relevant as the (DOA) signal is very noisy.

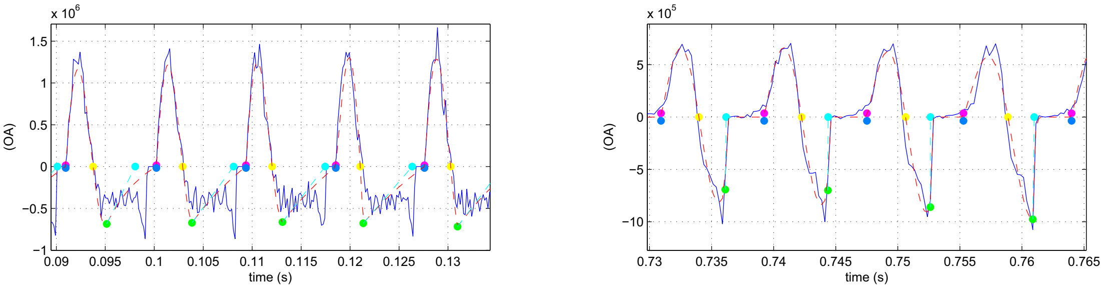

FIGURE 2: (DAO) measurement (blue) and LF estimation (dotted red). left: lips (AO), right: glottis (OA)

Delays between maximum aperture and closure on (AO) and (EI)

In order to provide a first comparison between multi-modal signals collected in singing and trombone playing, we propose to focus on the temporal relationship between (EI) and (OA) in both cases. For each period of each recording, the time shift between (OA) closing instant and (EI) maximum, as well as between (OA) maximum and (EI) minimum are calculated and normalized to the period T_0. We refer to these two parameters, respectively, as T_1 and T_2. Results are represented in figure 3.

Given the low temporal resolution of (OA) compared to (EI), the error introduced in T_1 and T_2 estimation for the highest tones makes it difficult to analyze the data at these frequencies. Therefore, only the results obtained for the lowest tones (fundamental frequency bellow $175Hz$) will be commented. The corresponding data points are represented in blue and cyan in figure 3.

We first notice that voice and trombone data are located in two distinct areas of the two-dimensional space. While voice shows negative values of T_1, trombone playing data displays values of T_1 centered around zero, suggesting that closing instants are well determined by (EI) maximum for the lips. In both cases, significant variation of (EI) occurs during the closed phase observed in the (OA) waveform. This supports the hypothesis of a variation in lip and vocal fold contact area along their thickness during the closed phase.

The time shift between (OA) maximum and (EI) minimum is found to be positive in both voice and trombone, the latter showing the higher values of T_2. This indicates that in both valve systems, the minimum in (EI) is delayed from the maximum is opening area. Again, this possibly results from a complex variation of focal-folds and lip contact area during the opening phase. In the case of the lips, this may be linked to the combination of a transverse and longitudinal motion of the lips, typically observed in brass performance [7].

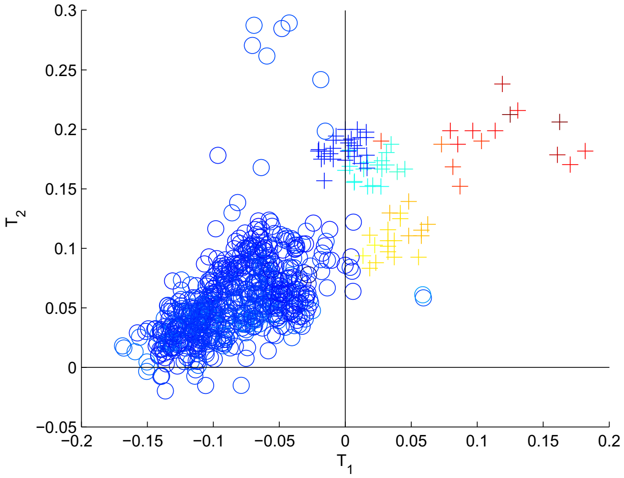

FIGURE 3: T_1: Time shift between opening area (OA) at closing instant and the maximum of electrical impedance (EI) signal, normalized to the period T_0 (horizontal axis); T_2: time shift between opening area (OA) maximum and electrical impedance (EI) minimum, normalized to the period T_0 (vertical axis). Each point corresponds to one period. Circles refer to the glottis and crosses refer to the lips. A color scale (ranging from $75Hz$ to $300Hz$) denotes the fundamental frequency.

Open quotient estimation on (EI) and (OA)

The open quotient (OQ) is defined for the glottis as the ratio between the open phase and the closed phase over one cycle. Efficient methods exist to estimate the open quotient from EGG signals (or its derivative) [8]. These algorithms are used to analyze (EI) signals in both cases.

Table 1 presents the mean estimated (OQ) for each signal. As one can see, (OQ) values are significantly higher for the brass case than for the voice case, as it was already observed [9]. This difference is clearly visible in the (HSV) and (AO) signals. The completely closed phase is very short in the trombone case but can be rather long in voice. This may be due to the fact that the lips are not closing and opening along their thickness, unlike the vocal folds.

Synchronous displaying of multi-modal measurements

A video tool has been built to synchronously display the measured signals (HSV), (EI) and (RS) in both time and frequency domains. This tool also displays a fundamental frequency estimation, the 2D open area estimation and the (OA) signal (in both time and frequency domains). Two screen shots of these videos are shown in figure 4.

These videos allow us to synchronously visualize the temporal evolution of the different measured signals and some extracted features. One can see the (OA) and its derivative (DAO) along with its local magnitude spectrum, the (EI) and its derivative (DEI) along with its local magnitude spectrum and the (RS) along with its local magnitude spectrum. It is a useful tool to visually explore the behavior of lips or glottis in different regimes. One can easily graphically estimate standard EGG features on the (EI) and (DEI) signals, such as the open quotient or the asymmetrical coefficient.

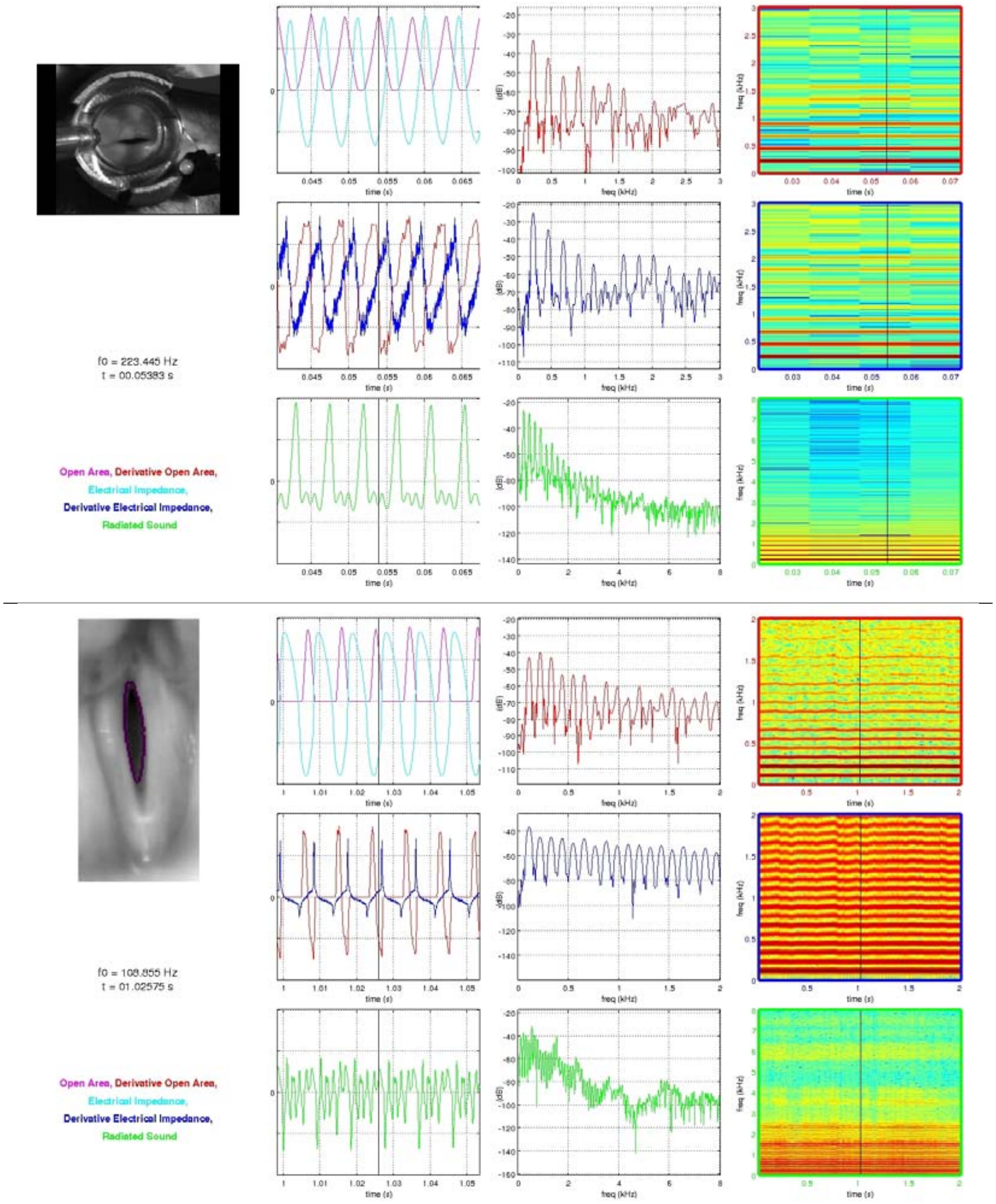

FIGURE 4: Synchronous display of multi-modal measurements of the lips and glottis

To our knowledge, there is no tool allowing simultaneous and synchronous visualization of all these measurements. This tool could therefore be the object of further development for medical and pedagogical applications.

TABLE 1: Numerical results: fondamental frquency f_0, quadratic error for LF-estimation. Left: lips, right: glottis ϵ, open quotient Oq

f_0 (Hz)	165	109	277	223	109	123	114	133	181
ϵ	0.499	0.454	0.469	0.479	0.327	0.356	0.294	0.377	0.400
Oq	0.662	0.831	0.557	0.626	0.567	0.571	0.488	0.554	0.623

Conclusion and perspectives

In this paper, we presented a common approach to explore the vibration behavior in the cases of two human valves: the lips of a trombone player and the vocal folds of a speaker. We highlighted a common structure for these two aero-acoustical systems and we spotted some interesting differences in their vibration characteristics, such as general shape of opening areas and open quotient.

These first observations led us to the development of a new video tool that can be used either for research, medical or pedagogical purposes. Nevertheless, we highlighted in our study that the motion over the third dimension (the thickness), which can't be seen on the (HSV), is far from being negligible. For now, no tools allow to explore this motion *in vivo*, but recent work on multi-channel EGG [10] should lead to a more effective device for 3D exploration of vocal fold motion.

References

[1] M. Castellengo and a. n. S. B. Caussé, R., "Étude acoustique de l'émission multiphonique aux cuivres", in *Proc. 11th International Congress on Acoustics*, 355–357 (Paris, France) (1983).

[2] V. Freour and G. Scavone, "Development of an electrolabiograph embedded in a trombone mouthpiece for the study of lip oscillation mechanisms in brass instrument performance", in *Canadian Acoustics*, volume 39, 130–131 (2011).

[3] G. Degottex, E. Bianco, and X. Rodet, "Estimation of glottal area with high-speed videoendoscopy", in *Speech Production Workshop: Instrumentation-based approach* (ParisIII/ILPGA, Paris, France) (2008).

[4] T. Hézard, T. Hélie, R. Caussé, and B. Doval, "Analysis-synthesis of vocal sounds based on a voice production model driven by the glottal area", in *Proceedings of Acoustics 2012* (2012).

[5] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow", STL-QPSR **26**, 1–13 (1985), URL `http://2.inarchive.com/1206/14/213/GpHUDG.pdf`.

[6] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder-mead simplex method in low dimensions", SIAM Journal of Optimization **9**, 112–147 (1998).

[7] S. Yoshikawa and Y. Muto, "Lip-wave generation in horn players and the estimation of lip-tissue elasticity", Acust. Acta Acust. **89**, 145162 (2003).

[8] N. Henrich, C. d'Alessandro, B. Doval, and M. Castellengo, "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation", The Journal of the Acoustical Society of America **115**, 1321–1332 (2004).

[9] S. Bromage, M. Campbell, and J. Gilbert, "Open areas of vibrating lips in trombone playing", Acust. Acta Acust. **96**, 603613 (2010).

[10] T. Hézard, T. Hélie, D. B., H. N., and K. M., "Non-invasive vocal-folds monitoring using electrical imaging methods", in *Proc. 100 Years of Electrical Imaging*, 145–148 (Paris, France) (2012).