An educational platform to capture, visualize and analyze rare singing

P Chawah, T Fux, Martine Adda-Decker, A Amelot, N Audibert, B Denby, G Dreyfus, A Jaumard-Hakoun, C Pillot-Loiseau, P Roussel, et al.

An educational platform to capture, visualize and analyze rare singing

P. Chawah[1], S. K. Al Kork [2,3], T. Fux[1], M. Adda-Decker[1], A. Amelot [1], N. Audibert[1], B. Denby[2,3], G. Dreyfus[3], A. Jaumard-Hakoun[2,3], C. Pillot-Loiseau[1], P. Roussel[3], M. Stone[4], K. Xu [2,3], L. Buchman[1]

[1]Phonetics and Phonology Laboratory, LPP -CNRS, UMR7018, University Paris3 Sorbonne Nouvelle
[2]Université Pierre Marie Curie, Paris, France,
[3]Signal Processing and Machine Learning Lab, ESPCI Paris -Tech, Paris, France,
[4]Vocal Tract Visualization Lab, University of Maryland Dental School, Baltimore, USA

`patrick.chawah@gmail.com, samer_alkork@hotmail.com, thibaut.fux@univ-paris3.fr`

Abstract

This paper presents an early version of an open extendable research and educational platform to support users in learning and mastering the different types of rare-singing. The platform is interfaced with a portable helmet to synchronously capture multiple signals during singing in a non-laboratory environment. Collected signals reflect articulatory movements and induced vibrations. The platform consists of four main modules: i) a capture and recording module, ii) a data replay (post processing) module, iii) an acoustic auto adaptation learning module, iv) and a 3D visualization sensory motor learning module. Our demo will focus on the first two modules. The system has been tested on two rare endangered singing musical styles, the Corsican "Cantu in Paghjella", and the Byzantine hymns from Mount Athos, Greece. The versatility of the approach is further demonstrated by capturing a contemporary singing style known as "Human Beat Box."

Index Terms: education platform, ultrasound sensor, singing analysis, synchronous data capturing, i-Treasures project.

1. Introduction

As the world becomes more interconnected, many different cultures come into contact and communities start losing important elements of their intangible cultural heritage (ICH) [1]. Many examples of singing traditions are in need of urgent safeguarding. The main objective of the "i-Treasures" project [2] is to develop an open and extendable platform to provide access to ICH resources, to enable knowledge exchange between researchers and to contribute to the transmission of rare know-how from Living Human Treasures to apprentices. Performing singing is usually learned under the supervision of a master. Therefore, such ICH features that require physical actions survive as they are transferred through a master-apprentice approach. In order to facilitate the transmission of such learning information, we are working on an educational platform that makes the link between the master and the apprentice by means of a variety of sensors and a developed software. To explore the complex and mainly hidden human vocal tract, a large number of sensing techniques have been used [3]. The system, based on vocal tract sensing methods developed for speech production and recognition, consists of a lightweight "hyper-helmet" (Figure *1*) containing an ultrasonic (US) transducer to capture tongue movements [4] , a video camera for lip movement detection, and a microphone. These three instruments are coupled to a further sensor suite including an electroglottograph (EGG), a nose-mounted accelerometer, and a respiration sensor.

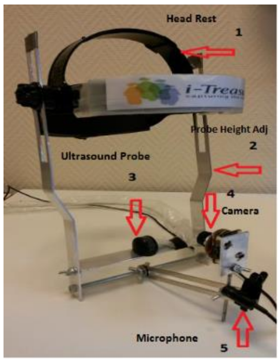

Figure 1 [3]: *Multi-sensor Hyper-Helmet: 1) Adjustable headband, 2) Probe height adjustment strut, 3) Adjustable US probe platform, 4) Lip camera with proximity and orientation adjustment, 5) Microphone*

2. The educational platform

2.1. Capturing, visualizing and recording

2.1.1. Motivation and objectives

Since configuring separated sensors and recording their outputs may be a complicated issue if managed individually, a common module is specifically designed. The proposed module named i-THRec (i-Treasures Helmet Recording software) contains multiple Graphical User Interfaces (GUI) forms, each of them aimed at one of the following objectives:

- Creating directories to organize and store the newly acquired data into corresponding sub-folders;
- Writing new .xml lyric files that contains the sub-session paragraphs to be pronounced;
- Handling the calibration session: Calibrating the sensors and supervising their performances;
- Operating the recording session.

2.1.2. Simulation environment

A first version of i-THRec is programmed and implemented in C++ and its graphical interface is constructed on Visual Studio® using Microsoft Developer Network (MSDN) libraries and the .Net Framework. Hence, the above mentioned programming techniques do not take into account the interaction between the sensors and the computer. The acquisition of the data communicated by the six sensors is insured by using the Real-Time Multi-sensor Advanced Prototyping software (RTMaps®, Intempora Inc. [5]). RTMaps has the ability to acquire, display and record data and could be sufficient by itself. However, we prefer to use RTMaps as a shadow toolkit in order to develop user-friendly software.

2.1.3. *Architecture and frames*

Both the calibration and the recording sessions frames contain picture boxes displaying the acquired images of the lips-camera and the tongue ultrasound sensor in addition to the RTMaps oscilloscopes. The oscilloscopic displays include the speech signals from the microphone, the nasality signals piezo-electrical instrument, the respiration belt measurements and the EGG sensor of vocal fold motion. The first derivative of the EGG signal is calculated and displayed by its turn on the oscilloscopes. The calibration session (see *Figure 2*) window also contains the settings of the sensors such as the camera frame rate, the ultrasound luminance, the audio line sample rate and the oscilloscope offsets. These settings are then transmitted to the recording session window.

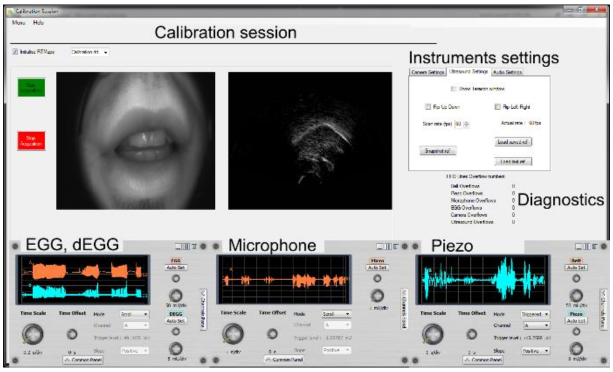

Figure 2: An example of a calibration session window. The chosen settings are used in the subsequent recording session.

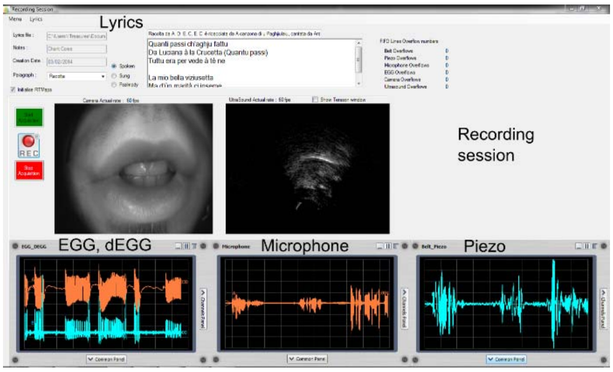

Figure 3: An example of the recording session window. The settings section disappears to reveal the lyrics section.

In addition, the recording session window (*Figure 3*) contains code that analyzes a chosen lyric .xml file and fragments it into separated paragraphs. Each of these paragraphs which contains either vowels, syllables, phonemes or song texts, is displayed in a text-box to be easily read by the performing singer. The camera, the ultrasound images, and the audio signals are time-aligned and saved. These data are henceforth ready to be post-processed using a MATLAB® based graphical user interface (GUI) named i-THAn.

2.2. Replay and Post processing

The post processing module referred by i-THAn (i-Treasures Helmet Analyzer) is a MATLAB tool that displays and analyzes the data streams captured by the multi-sensor hyper-helmet. i-THAn also provides a comprehensive set of capabilities to create measurement reports, figures, images and documentation. The current version of i-THAn includes tools dealing with the speech and the EGG signals. The pitch information, the open quotient and the spectrogram can be computed and viewed synchronously with the signals.

3. Future work

Two major modules are still under development and will be added to the platform in the near future. The first will support new apprentices who aim to better perform rare songs. This module will rely on the already recorded data of the professional singers. The module will be called up by the acoustic auto-adaptation learning module.

The second module, 3D sensorimotor visualization and learning, will feature a 3D singing avatar that consists of the tongue, lips and face. The domain knowledge of real masters will be transferred to the 3D avatar. This sensorimotor visualization learning module is expected to increase the awareness and dissemination of the above mentioned rare singing styles.

4. Conclusions

Keeping alive intangible cultural heritage, especially rare singing may be sustained using new technologies in view of an educational approach. In this paper, we described the progress of a learning educational platform that aims to link the two parts. This approach is currently in the phase of the professional singer data collection. Future work is being planned to create two modules that help the apprentice to auto-adapt singing techniques in addition to 3D vocal tract visualization to correctly perform such rare singing

5. Acknowledgements

This work is funded by the European Commission via the i-Treasures project (Intangible Treasures - Capturing the Intangible Cultural Heritage and Learning the Rare Know-How of Living Human Treasures FP7-ICT-2011-9-600676-i-Treasures).

6. References

[1] UNESCO, "http://www.unesco.org/culture/ich/en/convention"
[2] Intangible treasures - capturing the intangible cultural heritage and learning the rare know-how of living human treasures," http://i-treasures.eu/"
[3] S. K. Al Kork, A. Jaumard-Hakoun, M. Adda-Decker, A. Amelot, L. Buchman, P. Chawah, G. Dreyfus, T. Fux, C. Pillot, P. Roussel, M. Stone, K. Xu, B. Denby, "A Multi-Sensor Helmet to Capture Rare Singing, An Intangible Cultural Heritage Study", 10th International Seminar on Speech Production, Cologne, Germany, 5 - 8 May 2014
[4] B. Denby, M. Stone, "Speech synthesis from real time ultrasound images of the tongue", in Acoustics, Speech and Signal Processing; ICASSP, 1, pp.685, 2004.
[5] INTEMPORA S.A. "http://www.intempora.com/"