Published in Models and Analysis of Vocal Emissions for Biomedical Applications. 13th International Workshop, September, 12-13, 2023, which should be cited to refer to this work.

DOI:10.36253/979-12-215-0146-9

OPERATIC SINGING MULTI-SENSOR RECORDING PROTOTYPE: A PILOT EVALUATION STUDY

E. Angelakis[1], K. Bakogiannis[1], A. Andreopoulou[1], M. Habela[2], A. Georgaki[1]

[1] National and Kapodistrian University/Department of Music Studies, Laboratory of Music Acoustics and Technology, Athens, Greece

[2] Haute Ecole Spécialisée de Suisse Occidentale/Haute Ecole de Musique de Genève, Geneva, Switzerland

angelakisv@music.uoa.gr, kbakogiannis@music.uoa.gr, aandreo@music.uoa.gr, marcin.habela@hesge.ch, georgaki@music.uoa.gr

Abstract: **The Operatic singing music genre is characterized by the utilization of the vocal mechanism in some of its most demanding and complex kinetic manifestations. The study presented here serves as a) an introduction to a newly designed multi-sensor recording prototype for operatic singing, b) an account of its use in a research project, as well as c) a preliminary test for the prototype's evaluation. The proposed prototype employs sensors that record acoustic and electroglottographic data, breathing kinetic actions, and data regarding pertinent postural and body movement behavior. It was recently utilized for the recording of an experiment with 28 operatic singers. Captured data for three of the participants was used for the present pilot study, and short videos of these three singers were given to an expert vocal trainer for 'vocal technique problems' empiric evaluation via a questionnaire. A subsequent examination of the recorded multi-sensor data resulted to the successful detection of objective evidence to support the 'vocal technique problems' reported by the expert.**

Keywords: ***operatic singing,*** **sensors, EGG, posture, breathing**

I. INTRODUCTION

Operatic singing is an art form that spans more than 400 years. It is an empiric trait delivered from singer to singer through the ages, one that demands great precision, control, and artistry. Although scientific research on the subject has progressed much during the last few decades, and numerous studies have been conducted [1], there is still much to be uncovered regarding the details of the mystifying functions of the singing voice, and operatic singing in particular. Singing, in any genre, is essentially the result of neuromuscular processes which involve voluntary and involuntary control of both external and internal mechanisms of the human body [2].

It thus stands to reason that an examination and tracing of such biomechanical functions would be essential in order to delve deeper into the 'secrets' of this extremely rigorous art. Voice research "concern with functionality is increasing" since 2010 [1], and multi-sensor research [3,4] and software/hardware applications for the singing voice -such as VoceVista Video Pro (Sygyt Software, Bochum, Germany)- are expanding sectors. Practical implementations of the above, concerning the art form of singing, can be found in vocal pedagogy [5], but have also been used for the recording of rare vocal music genres [6].

This paper discusses the design of an operatic voice multi-sensor recording prototype and its pilot evaluation study. The prototype features the possibility to record with 6 sensors (using commercial software), a control interface programmed in Max/MSP for reading and recording data from the skeletal tracking camera, and MATLAB code for synchronizing and manipulating all the recorded data. The prototype comprises the following sensors:

- ✓ Condenser microphone – Behringer ECM8000
- ✓ Electroglottograph (EGG) - Glottal Enterprises EG2-PCX2 (recording both vocal fold degree of contact and Vertical Laryngeal Position -VLP)
- ✓ Two distinct Respiratory Effort Transducers – Biopac SS5LB
- ✓ Time-of-Flight (ToF) Skeletal tracking camera – Azure Kinect Microsoft
- ✓ HD video – iPhone 13 Pro.

The sensors were selected taking into consideration the requirements for high portability, low invasiveness, and representation of a relatively large number of pertinent-to-singing kinetic functions. The innovation of the proposed prototype lies in its ability to generate 12 sensor data streams, seven of which calculated by its gesture-following functionality. This enables the acquisition of synchronized, quantifiable data regarding singing-related biometrics. The present pilot study was conducted with the aim of investigating whether the combined use of these sensors yields data that is relevant for the comprehensive evaluation required for the final prototype assessment.

II. METHODS

Multi-sensor Recording Protocol: The proposed prototype was used to record an experimental part of a

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)
Claudia Manfredi (edited by), *Models and Analysis of Vocal Emissions for Biomedical Applications. 13th International Workshop, September, 12-13, 2023*,

larger project led by the Haute Ecole de Musique de Genève. The recordings took place in three venues. A total of 28 singers were recorded, all of them graduate or post-graduate students and some young professionals.

Prior to the measurements, participants were instructed to arrive for the study vocally warmed-up and in good vocal condition. They were also asked to memorize an Italian aria of their choice, a song in their native language, and the first two phrases of the Aria Antica 'Caro mio ben'. All participants were requested to sign informed consents, as well as to complete a demographics and vocal health questionnaire.

The measurement protocol for each participant began with a calibration process during which all sensors were manually adjusted. The recording phase commenced with data synchronization events, followed by various vocal exercises, the 'Caro mio ben' phrase, the Italian aria, the song, and concluded with ending synchronization events.

Data Collection: Microphone (voice) and Electroglottograph (EGG and Laryngeal Tracking signals) data were recorded on 3 mono channels using a Steinberg UR44-C external sound card, and Cubase 12 at a 48 kHz sampling rate. The thoracic and abdominal breath monitoring transducers were connected onto the specialized ‘Biopac MP35 Four Channel Data Acquisition System’ and their data were recorded at a 25 kHz frame rate, using BSL4 Pro Software (Biopac Systems, Inc., Santa Barbara, CA). Both devices (MP35 and UR44-C) were connected to the same laptop pc through USB-C and Hi-speed USB ports respectively. Breathing transducer data streams were exported as .wav sound files.

A second laptop was used to read and record the skeletal tracking data from the Kinect Azure DK camera. A Max/MSP patch was developed to record the coordinates and orientation of 15 body ‘joints’ (head, right eye, left eye, right ear, left ear, nose, head centre, neck, thorax, right shoulder, left shoulder, navel, pelvis, right hip, left hip). These data were automatically exported along with their corresponding timestamps and sound level values into a .txt file with a sample rate of 60 Hz. Video recordings of all experiment trials were made in 1080p 30 fps video using an iPhone 13 Pro.

Each measurement started and was concluded with a short synchronization sequence, which consisted of three hand claps, followed by three small ‘cough-like’ glottal attack sounds, produced simultaneously with an abdominal muscle inward activation, and a small, sharp downward head bend. This latter event was selected as it provides information recorded by all sensors (microphone, EGG, breathing transducers, skeletal tracking) and can thus be used for data synchronization verification. A large pause of about 10 seconds was introduced between the first and second clap, during which participants were asked to stand, in complete silence, in what they considered their personal optimal upright posture. Skeletal tracking data from this pause was used to set each user’s reference posture.

Data Processing: The collected data were resampled to 44.1 kHz, synchronized, and clipped automatically in MATLAB. Synchronization was achieved by automatic alignment of the audio streams recorded in all data packages (DAW, BSL4, Kinect Patcher, Video). MATLAB was also used to process the skeletal tracking data and output 7 distinct data streams of specific movements of the singer’s body. These movements were selected upon consultation with internationally acclaimed singing teachers as the most appropriate for this study. They were clearly visible movements that could either impact vocal production, or be good indicators of a technical, habitual, or physiological ‘issue’ that can impede the optimal function of the voice’s kinetic mechanisms. These selected movements were: 1) body posture, 2) up-down head bend, 3) left-right head turn, 4) parallel front-back head movement, 5a) right shoulder up-down, 5b) left shoulder up-down, 6) shoulder front-back (kyphosis-backward stretch).

Evaluation: The preliminary evaluation of the multi-sensor prototype relied on the objective and subjective assessment of a subset of the collected data (three participants), selected using the following criteria: a) same gender, b) recorded at the same venue c) different level of expertise (advanced: singer01, novice: singer02, and young professional: singer03). The videos of these three singers were exported synced with the audio from the measurement microphone, and the following excerpts were selected for subsequent evaluation: two scale exercises, the ‘Caro mio ben’ phrases, and the most demanding part of their selected aria. Their approximate duration ranged between 2’20’’ and 2’40’’. These excerpts were used for the subjective assessment section of the preliminary evaluation, which aimed towards a two-fold objective. First and foremost, to demonstrate as to whether ‘vocal technique problems’ (VTPs) and their indications (as reported by an expert) had been recorded and were discernable within the research data. The second objective was to establish a methodology for a large-scale study with many expert judges. In the above scope, the data was sent to an expert assessor for subsequent evaluation along with a questionnaire and detailed instructions. The selected expert judge is an internationally acclaimed operatic singer with a 23-year singing career, 18 years of teaching experience, and a comprehensive understanding of the physiological mechanisms pertinent to vocal production. The questionnaire which they were asked to follow consisted of 9 questions regarding mainly a) the VTPs they perceived, b) the indications that led to the detection of each VTP, and c) suggestions on muscular systems each participant should work on.

III. RESULTS

A. Expert judge report and examination of multi-sensor data: Questionnaire answers from the expert who judged the three participants' videos were examined and the reported VTPs are listed below, sorted by sensor data stream where they have been recorded, or where evidence for these 'problems' can be detected.

- ✓*EGG signal*: insufficient glottal adduction, arytenoid cartilages strain, increased air pressure,
- ✓*VLP signal*: low position of larynx, downwards pressure of larynx, lack in laryngeal control of movement,
- ✓*Kinect spinal posture in conjunction with abdominal breathing sensor*: Insufficient body support, air control,
- ✓*Kinect shoulder forward/backward bend/stretch in conjunction with thoracic breathing sensor*: thoracic spinal region tension,
- ✓*Uncategorized (no pertinent recorded data)*: tongue lower part stiffness.

The expert was also requested to provide indications that led to the report of VTPs. These indications are listed here and can mostly be detected in an audio spectrographic analysis, EGG signal analysis, and supported by data from the breathing sensors and postural data.

Indications: instability in vibrato & intonation and voice quality change, unstable intonation and vibrato, breathy sound, unstable dynamics, sound distortion, growling sound, short duration of high note, face observation, posture observation.

B. Data analysis:

In order to provide a few characteristic examples of quantifiable indications for reported 'vocal technique problems', a selection of recorded data analysis for participant singer01 is presented in this section. Reported issues for this singer that can be observed here are: "Larynx position without adequate control", "insufficient air pressure control through body support", "intense subglottal air pressure", "vibrato frequency decrease", "breathy sound", "muscular tension" in the thoracic spinal region.

Example #1: Evidence of a reported VTP is illustrated in Fig. 1 from an analysis of the VLP and audio data streams. It concerns a sustained note of singer01 on a G3 note (196 Hz) and the word "ben" (first phrase of 'Caro mio ben' Aria), which was indicated as problematic by the expert judge. The expert's report for lowering of the larynx is apparent on the onset of this and the following syllable, while lack of its control could be attested by the constant variation of the VLP during the sustained vowel/tone, which is connected also to the reported vibrato instability.

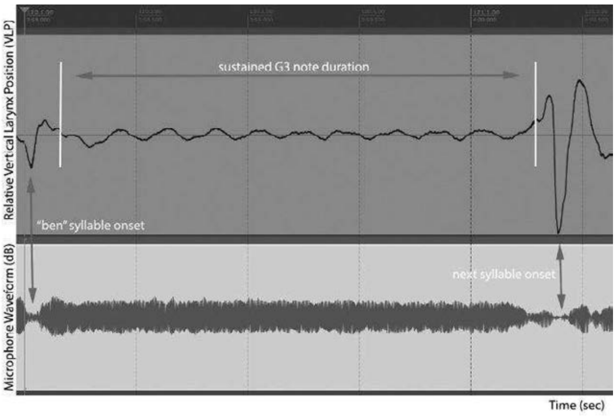

Figure 1. Example #1 depiction of VLP (top) and microphone (bottom) waveforms for reported G3 note.

Example #2: Vibrato rate for singer01 has been measured in VoceVista, as seen in Fig. 2, to occasionally decrease from an already relatively low 6 Hz [7] to values between 4.5-5.0 Hz in sustained notes, thus confirming the expert's report for "vibrato frequency decrease". Analysis also interestingly revealed a high vibrato extent. While vibrato extent in operatic singers has been found to range "between ±34 and ±123 cent" and "the mean across tones and singers amounted to ±71 cent." [8], singer01 was measured to have an extent range of about ±49 to ±200 cent, with an astonishing extent rise during the note in question (among others) up to a range of ±159 to ±282 cent, which should greatly add to the audible vibrato fluctuation effect.

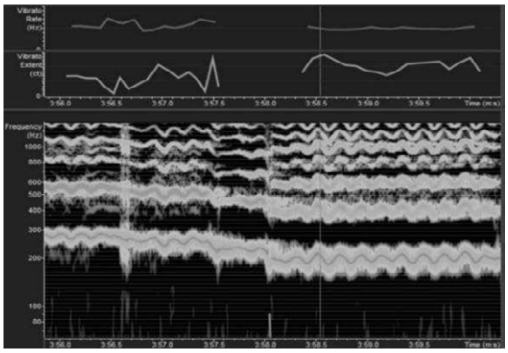

Figure 2. Example #2 vibrato rate in Hz (top), vibrato extent in cents (middle), and spectrogram (bottom) for the first 4-note except of 'Caro mio ben' aria, with the last note being the note used for Example #1.

Example #3: Reported muscular tension in the thoracic part of the spine cannot easily be measured with non-invasive techniques, such as the ones deployed in the present prototype. However, recorded data (as shown in Fig. 3 for the whole first phrase of 'Caro mio ben') demonstrate that singer01 tended to employ a combination of abdominal and thoracic breathing that seem to commence simultaneously but follow distinct recession slopes.

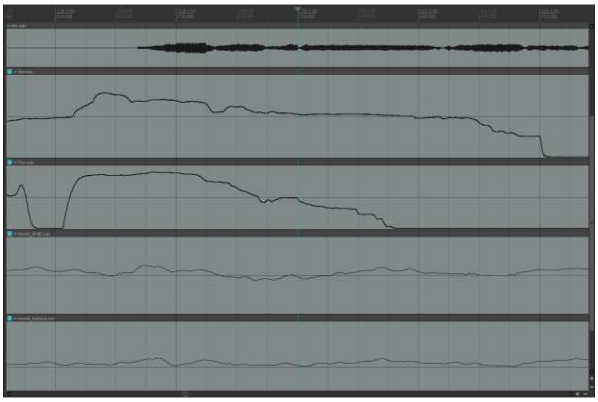

Figure 3. Example #3 selected waveforms (top to bottom): 1) microphone in Hz, 2) abdominal circumference relative variation, 3) thoracic circumference relative variation, 4) spinal posture relative variation from lordotic (negative values) to kyphotic (positive values), 5) shoulders bend relative variation from backward (negative values) to forward (positive values). Depiction pertains to the first 8-note except of 'Caro mio ben' aria.

More specifically, thoracic volume appears to collapse rapidly, while abdominal muscle activation seems to persist a while longer. Consequently, the second part of the singer's phrase is sung with what the expert described as "insufficient air pressure control through body support". This lack of proper muscular activation, and especially the chest cavity collapse, is coupled with a tendency for a kyphosis-like spinal and shoulder forward bend (which can be seen in lines 4 and 5 of Fig. 3). This conjunction of elements that are apparent in the data streams, illustrate a condition in which the singers' breathing muscles are contracted and tension starts to build up, as the body has no resources with which to control the subglottal pressure. The above effect could be what the expert judge noted as "muscular tension" in the thoracic spinal region.

Example #4: Contact Quotient (CQ) appeared to be over 0,60 for the most part of the trial, when computed with VoceVista using a hybrid method, where contact instant computed by the derivative of EGG signal, opening instant computed using a threshold set at 0,43 as indicated by Herbst [9]. For comparison with a previous study on CQ in pressed phonation [10], CQ was also calculated using a criterion method with a threshold set at 0,35 and was found to range between 0.61 and 0.73 for the specified note, values higher than previously reported even for pressed phonation [10]. This seems to be in accordance with the expert judge's mention of increased subglottal pressure.

Example #5: Vocal sound 'breathiness' characteristic in singing has been shown to be predictable from the audio and the EGG data, using computations, such as the Multi-Dimensional Voice Profile [3], or the (currently expanded/revised) multiple regression model CDH [11].

IV. Discussion

A comprehensive inspection of the recorded data revealed evidence of expert-reported VTPs in most cases. Such information could be obtained either from singular sensor output, or through combinatory analysis of two or more synchronized data-streams. The sole case of reported VTP that was not recordable in the prototype data was "tongue lower part stiffness". Similarly, the sole 'indication' not quantified was "facial muscle activation", which was nevertheless recorded in video. Moreover, in response to a questionnaire question, the expert judge provided a list of suggested vocal exercises targeting various muscular systems, most of which can be monitored by the proposed prototype. This positions the prototype as a promising candidate to evolve into an assistive tool for vocal pedagogy.

Limitations and Future work: The present study served as a pilot study and, therefore, was conducted using a limited number of participants and judges. A large-scale study is already in progress, using a revisited analysis methodology, more participants and experts, and participant questionnaire analysis. Finally, there is the possibility of replacing the breathing sensors with more readily available options, as well as the potential evolution of the synchronization method and code.

Acknowledgements: The authors would like to thank Biopac Systems Inc. and David Fedorko, as well as the expert judge and participants to the study.

References

[1] P.M. Pestana, S. Vaz-Freitas, M.C. Manso, Trends in Singing Voice Research: An Innovative Approach, J. Voice. 33 (2019) 263–268. https://doi.org/10.1016/j.jvoice.2017.12.003.

[2] P.J. Davis, S.P. Zhang, A. Winkworth, R. Bandler, Neural control of vocalization: Respiratory and emotional influences, J. Voice. 10 (1996) 23–38.

[3] M. Aaen, J. McGlashan, K.T. Thu, C. Sadolin, Assessing and Quantifying Air Added to the Voice by Means of Laryngostroboscopic Imaging, EGG, and Acoustics in Vocally Trained Subjects, J. Voice. 35 (2021) 326.e1-326.e11.

[4] S. Ternström, S. D'Amario, A. Selamtzis, Effects of the Lung Volume on the Electroglottographic Waveform in Trained Female Singers, J. Voice. 34 (2020) 485-e1.

[5] F.M.B. Lã, M.B. Fiuza, Real-Time Visual Feedback in Singing Pedagogy: Current Trends and Future Directions, Appl. Sci. 12 (2022).

[6] P. Chawah, S.K. Al Kork, et al., An educational platform to capture, visualize and analyze rare singing, Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH. (2014) 2128–2129.

[7] J. Sundberg, Acoustic and psychoacoustic aspects of vocal vibrato, 1994.

[8] E. Prame, Vibrato extent and intonation in professional Western lyric singing, J. Acoust. Soc. Am. 102 (1997) 616–621.

[9] C.T. Herbst, Electroglottography – An Update, J. Voice. 34 (2020) 503–526.

[10] K.G. Ong Tan, Contact Quotient of Female Singers Singing Four Pitches for Five Vowels in Normal and Pressed Phonations, J. Voice. 31 (2017) 645.e15-645.e22.

[11] E. Angelakis, N. Kotsani, A. Georgaki, Towards a singing voice multi-sensor analysis tool: System design, and assessment based on vocal breathiness, Sensors. 21 (2021) 1–25.