Audio Engineering Society

Convention e-Brief 372

Presented at the 143rd Convention
2017 October 18–21, New York, NY, USA

"Match Your Own Voice!": An Educational tool for Vocal Training

Angelakis, E.[1], Velianitis, P.[1], Andreopoulou, A.[1], and Georgaki, A.[1]

[1]*Laboratory of Music Acoustics and Technology (LabMAT), National and Kapodistrian University of Athens, Greece*

Correspondence should be addressed to Angelakis, E. (angelakisv@gmail.com)

ABSTRACT

In this paper, we discuss the development and preliminary evaluation of a new educational tool, intended for novice and advanced vocal students. The software, written in Max / MSP, aims to assist singing practice by providing users with a visual substitute to their subjective auditory feedback. Under the guidance of their professional vocal instructor, students can store in the software spectral representations of accurately produced sounds, creating personalized Reference Sound Banks (RSBs). When students practice on their own, the software can be put into practice, assisting them to match their current Voice Spectrum Harmonic Content to the stored RSBs one note at a time. Results of a preliminary evaluation showed that, when using this software, students achieve a larger number of accurately produced sounds in a smaller amount of time.

1 Introduction

When dealing with novice music students, it is often the case that they are rather reluctant, or even unable, to achieve the same quality of sound when practicing their instruments at home, as when they study in class with their professional instructors. Similarly, novice vocal students, oftentimes do not feel confident enough to repeat on their own what they have learned in a vocal lesson. This setback is often due to the students' lack of experience and understanding of how to achieve a good quality performance. In addition, vocalists are unable to hear their sound without being affected by their internal hearing. This is a big limitation, especially during the early learning stages. "Match Your Own Voice!" is a novel educational tool, developed for vocal training purposes. It acts mainly as a facilitator during unsupervised vocal practice sessions, but can also serve as a reference to the users' vocal progress over time, as a means of boosting the students' vocal confidence, or even as a practice motivator.

Vocal pedagogy today relies often on methods that were developed more than 150 - 200 years ago. One of the most important differences is that in the past, students had daily supervised lessons, while today lessons are given once a week and students have to be their own coach in-between classes [1]. This tool aims to bridge the gap between the guided, by the vocal instructor, somatosensory perception of accurate vocal production, and the unassisted individual vocal practice of vocal students of all levels. It can assist users to start relying less on their vocal auditory feedback, early in their course of studies. This is achieved by providing them with a) real-time visual feedback, by means of a voice spectrum and b) a measure of comparison, between their own produced sound and the indicated by their instructors' own ideal sound.

Recent cognitive fMRI studies have indicated that professional opera singers tend to:

1. Have increased grey matter volume in right hemispheric regions, in the ventral primary somatosensory cortex (larynx aerea) and the adjacent rostral supramarginal gyrus, as well as in the secondary somatosensory cortex, in comparison to the general population [2].

2. Employ increased functional activation of bilateral primary somatosensory cortex representing articulators and larynx, when singing, in comparison to non-singers [3]. This seems to be linked to their ability to partly disregard their auditory feedback and control their vocal production through kinesthetic functions [3, 4, 5, 6, 7]. Same studies have also noted that switching rapidly between feedback modalities (auditory and somatosensory) is possible for professional singers, an ability which seems to be dormant in non-singers, but is brought out through repetition of vocal practice.

Furthermore, according to Kleber et al.,"expertise in classical singing correlated most notably with increased functional activation of bilateral primary somatosensory cortex, representing proprioceptive feedback from the articulators and the larynx, in concert with increased involvement of the cerebellum and implicit motor memory areas at the subcortical level" [3].

In accordance with these findings, "Match Your Own Voice!" was designed in a manner that promotes somatosensory-oriented singing practices through the use of visual feedback, which has been shown to be an objective alternative to the users' own subjective auditory feedback [8].

2 State of the art

Since the middle of the previous decade, in parallel with the fast growth of software for personal computers and mobile devices, there has been an abundance of applications on vocal training utilizing visual feedback. Most of these application focus on vocal aspects such as, pitch accuracy, phoneme quality, formants, harmonics / partials and voice quality. Selected examples of such software, which have a related approach to vocal training as "Match Your Own Voice!", follow:

1. *MiruSinger* [9] compares a user's voice to a certain standard and returns information in the form of visual feedback. This design resembles closely the "Match Your Own Voice!" software. The main difference is that MiruSinger compares users to a commercial recording, while our design compares them to their own best sounds (adaptive learning) and helps them repeat them until new muscle memory is built.

2. *Singing Coach Pro*[1] is interesting in the fact that it addresses users of all ages and music education levels. It also stands out in the way it engages users, through an inviting graphic interface and a challenge-based / game-like design.

3. *Phonaskein* [10] is a real-time voice analysis software with visual feedback, which focuses on pitch accuracy. It stands out for its novel functionality of letting users practice in non-standard, non-equal temperament scales , such as the Byzantine, Pythagorean, and Aristoxenian scales.

3 Software Overview and Technical Analysis

"Match your own voice" runs on Windows and Mac OS platforms, and will soon be available for mobile devices. It can either run as a standalone application, or work as a Max / MSP patcher file, provided the computer has a working licence. The standalone version of the program does not require installation and can run from any portable device, such a USB flash drive.

The software uses the Max / MSP external object *Sigmund*[2], designed by Miller Puckette and Ted Apel, to decompose the users' vocal sounds and analyze them into partials. Spectral analysis is limited to the 20 spectral peaks with the highest magnitude within the range of 0 Hz to 5 kHz. The software can be used during two different stages: a) Vocal Training Sessions (VTSs) with a professional instructor and b) Unassisted Practice Sessions (UPSs) of the vocal students.

3.1 Vocal Training Sessions

During a VTS the vocal coach can guide a student towards producing sustained vocal sounds with the desired timbral and volume qualities, one tone and vowel

[1] http://www.singingcoach.com/sc5-pro.html
[2] http://vud.org/max/

at a time. When the coach's criteria are met, students can take a partial "snapshot" of their sound with the press of a button, using a device which runs the "Match Your Own Voice!" software. The sound's spectral representation is then stored, building a personalized Reference Sound Bank (RSB). This bank can hold up to 5 vowel-sounds for 36 distinct tones per user, and can be updated or revised at will, as the student progresses.

3.2 Unassisted Practice Sessions

While students practice on their own, they can resort to the software to attain sounds approximating those indicated by their professional instructors. They, first, need to load their RSB and select the tone and vowel-sound they wish to practice on. In doing that, they gain visual access to the spectrum of the stored sample up to 5 kHz, and a synthesized audio representation of it using harmonics within the same frequency range. The synthesized sound is preferred over the original recording, in order to provide enough information to the students to achieve the correct sound production, while preventing them from trying to imitate, even unwillingly, the distinct human voice color, using auditory feedback. Every time a new tone is produced by the user, a panel on the interface shifts color to indicate the tone's proximity (in Hertz) to the selected RSB reference. When the produced sound is within a predefined tolerance range to the reference, a second, semi-transparent spectrum graph, appears layered over the first, in order to assist users to match or exceed the amplitudes of the RSB spectrum harmonics, achieving, in this way, the tone quality of the reference.

In order to achieve optimum comparison conditions between the two stages and extract accurate results, two factors should be taken into consideration:

1. Equipment: To avoid differences between microphone spectral responses and Digital to Analog Convertors (DAC), it is best to use a single setup (microphone, and external sound card), at identical input levels for the VTS and UPS stages.

2. Acoustic Adjustments: To achieve comparable acoustic results between VTS and UPS, users should maintain a constant distance from the microphone. Ideally, rooms with similar acoustic responses should be used, but if this is not possible, a smaller microphone distance paired with a lower gain sensitivity could partly compensate for the effect of room acoustics on the system. Microphone placement in the room should also be taken into consideration. Assuming the VTSs and UPSs take place in different rooms, an acoustic evaluation of the rooms and microphone placements could be of great benefit to the accuracy of the measurements. This can be done with the help of various external software tools, such as the freeware TimDec[3].

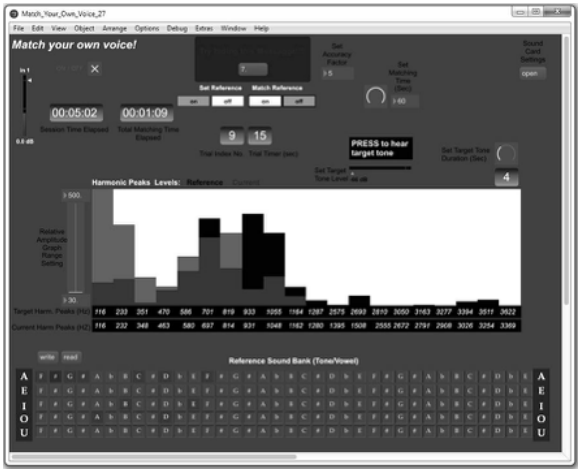

Fig. 1: View of the "Match Your Own Voice!" interface when comparing vocal sounds in real-time.

Fig. 2: Demonstration of the software to some of the study participants.

4 Preliminary Evaluation

11 participants (10 female) from a vocal training summer-camp, with singing proficiency levels ranging

[3] http://www.tolvan.com/index.php?page=/timdec/timdec.php

Fig. 3: Using the "Match Your Own Voice!" tool during the study.

from novice to advanced, participated in the preliminary evaluation of "Match Your Own Voice!". After 9 days of full-day vocal training, students, guided by their instructor, created partial personal RSBs in the software, consisting of 4 tones. The evaluation of the software began the following day, when participants attempted to match each of the tones in their RSBs in 30 sec. trials, under 2 conditions: a) without and b) with the use of the software. Under the first condition students had to rely on their memory of the optimal sound quality they had produced the day before, while under the second they had to use the software for feedback.

5 Preliminary Observations

The study has shown that students had much greater success matching their respective RSB sounds when using the software (condition 2), than when relying on their memory and auditory feedback (condition 1). In a total of 44 tones tried under condition 1 there were only 5 complete matches (11%) -achieved by 2 of the most advanced students in the group- and 9 partial ones. On the other hand, under condition 2, all students managed to achieve the target sounds for the majority of the test-cases. More specifically, there were 37 complete matches (84%) and 2 partial ones. In addition, under this condition, students managed to match their RSB sounds 50% faster than under condition 1.

6 Discussion and Future Work

"Match Your Own Voice!" is an educational tool, under development. It is not intended as a substitute, but rather as a complementary addition to vocal lessons, guided by professional instructors. In its current state, the software is expected to have a stronger impact on the progress of novice and medium-level students, as it is restricted to the practice of single sustained tones, rather than melodic phrases or full songs. The conducted preliminary evaluation showed encouraging short-term results on students of all levels.

Further development of the software will be based on user comments and the results of a full-length evaluation study, currently underway. The software interface is being, currently, redesigned to facilitate interaction, based on the informal comments of the preliminary study participants. Planned additions to the tool include user challenges and trials, aiming to motivate and promote long-term use of the software, separate teacher and student modes, as well as customization options.

References

[1] Chapman, J. L., *Singing and teaching singing: A holistic approach to classical voice*, Plural Publishing, 2011.

[2] Kleber, B., Veit, R., Moll, C. V., Gaser, C., Birbaumer, N., and Lotze, M., "Voxel-based morphometry in opera singers: Increased gray-matter volume in right somatosensory and auditory cortices," *Neuroimage*, 133, pp. 477–483, 2016.

[3] Kleber, B., Veit, R., Birbaumer, N., Gruzelier, J., and Lotze, M., "The brain of opera singers: experience-dependent changes in functional activation," *Cerebral Cortex*, 20(5), pp. 1144–1152, 2009.

[4] Zarate, J. M. and Zatorre, R. J., "Experience-dependent neural substrates involved in vocal pitch regulation during singing," *Neuroimage*, 40(4), pp. 1871–1887, 2008.

[5] Zarate, J. M., Wood, S., and Zatorre, R. J., "Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers," *Neuropsychologia*, 48(2), pp. 607–618, 2010.

[6] Zarate, J. M., "The neural control of singing," *Frontiers in human neuroscience*, 7, 2013.

[7] Kleber, B., Friberg, A., Zeitouni, A., and Zatorre, R., "Experience-dependent modulation of right anterior insula and sensorimotor regions as a function of noise-masked auditory feedback in singers

and nonsingers," *NeuroImage*, 147, pp. 97–110, 2017.

[8] Hoppe, D., Sadakata, M., and Desain, P., "Development of real-time visual feedback assistance in singing training: a review," *Journal of computer assisted learning*, 22(4), pp. 308–316, 2006.

[9] Nakano, T., Goto, M., and Hiraga, Y., "MiruSinger: A singing skill visualization interface using real-time feedback and music CD recordings as referential data," in *Multimedia Workshops, 2007. ISMW'07. Ninth IEEE International Symposium on*, pp. 75–76, IEEE, 2007.

[10] Moschos, F., Georgaki, A., and Kouroupetroglou, G., "Fonaksein: an Interactive Software Application for the Practice of the Singing Voice," in *Proceedings of the Sound and Music Computing (SMC) Conference*, September, pp. 326–331, Hamburg, Germany, 2016.