FONASKEIN: AN INTERACTIVE SOFTWARE APPLICATION FOR THE PRACTICE OF THE SINGING VOICE

ABSTRACT

A number of software applications for the practice of the singing voice have been introduced the last decades, but all of them are limited to equal tempered scales. In this work, we present the design and development of FONASKEIN, a novel modular interactive software application for the practice of singing voice in real time and with visual feedback for both equal and non-equal tempered scales. Details of the Graphical User Interface of FONASKEIN are given, along with its architecture. The evaluation results of FONASKEIN in a pilot experiment with eight participants and with four songs in various musical scales showed its positive effect in practice of the singing voice in all cases.

1. INTRODUCTION

Singing practices in Modern Greece have a long history and variability. Its roots go up to the interpretation of ancient Greek music which is considered as the theoretical fundament of Western music. The mathematical structure of Ancient Greek Music as referred in the works of Archytas, Philolaos, Didimos, Eratosthenis, Ptolemeos, and Aristoxenos still fascinate many researchers all over the world [1]. This written and oral tradition is being transferred with other types of music, through the centuries such as the written theory of Byzantine music [2], the oral tradition of Greek folk music, and even Rebetiko.

The unique characteristics that we found in each one of the diverse singing styles in Greece, along with their mathematical relationship can be described by a generative way using the well-tempered musical tonal system; thus there is a confusion between the oral tradition and the music notation. Many of these different singing practices are carried out in Greek schools through the traditional notation although the teaching approach does not take into account the different tuning systems [3]. The singing culture of children is still confused and depends on the cultural background of its family and the place of origin. Thus, in the same classroom we can listen the same song sung in 2-3 different tuning systems as the children's ears are tuned in a different way compared the well-tempered musical tonal system according to his/her particular influences which differ from region to region.

Although a number of visual feedback software applications for the practice of the singing voice have been introduced the last years, non–equal tempered music scales are not included.

In this paper, we present the design, development, and evaluation of FONASKEIN, a novel modular interactive software application for the practice of singing voice in non-equal tempered scales.

2. STATE OF THE ART ANALYSIS

2.1 A quick Overview

One of the first attempts for the design of software for the practice of the singing voice appeared at 1985 from G. Welch [4], who developed an innovative application for the BCC Microcomputer called SINGAD. SINGAD produced musical notes one-by-one and recorded the voice of the user who was asked to sing the notes. The application compared the fundamental frequencies (F0) from these two signals and displayed the results on the screen.

At the beginning of the 1990s, Welch and his team improved his software in three points, now running on an ATARI computer. First, instead of comparing only the fundamental frequencies of the two audio signals, SINGAD now could compare the whole audio contour which was more accurate. Second, as ATARI had a MIDI protocol, SINGAD could play audio via MIDI synthesizers or sound from General Midi like piano or flute. Finally, the graphical user interface became friendlier to musicians by including a viewer for the musical notes.

Another software application developed by Rossiter and his team in 1996 was ALBERT [5]. Except the voice training, ALBERT included monitoring of the laryngeal action. The system provided a greater variety of visual feedback by displaying the parameters F0, CQ (larynx closed quotient), spectral ratio, SPL (amplitude), shimmer and jitter. ALBERT was used in studies to identify the quality of voice production during visual feedback implementation, measured as the pattern of change during a training lesson.

Eight years later, in 2004, Callaghan and his team developed SING&SEE [6], one of the most popular application for the analysis of the singing voice with real time visual feedback (VFB). The main features of this research were the investigation of acoustic analysis technics, methods of displaying visual feedback in a meaningful way and the pedagogical approaches for implementing visual feedback technology into practice. Three parameters were distinguished as relevant for usage in the singing studio: pitch (F0 against time), vowel identity (R1, R2), and timbre (spectrogram). The major difference from former studies was that not only quantitative but also qualitative data were of interest in this development.

At the same year, 2004, Welch and his team come back with a new project called VOXed In this project Welch designed introduced the WinSINGAD [7]. The project also incorporated real-time VFB for singing education applications. While SING &SEE places emphasis on maximizing VFB technology itself, VOXed aimed at maximizing the collaboration between different fields. Psychologists, voice scientists, singing teachers, and singing students joined to form an interdisciplinary research team working for a better insight on the impact of VFB on the learning experience. Importantly, VOXed sought to work with participants as active agents rather than just passive recipients. The goal of the project was to investigate possible useful forms of VFB with the use of commercially available visual feedback software.

The innovation of the MiruSinger software application developed by Nakako and his team [8] was the possibility to use a real audio CD as a sample for the comparison for its user. MiruSinger analyzes the voice of the user, but also analyze the voice from the song from the audio CD. Thus, it compares the audio signals from two human voices and not the human voice with a synthetic sound. Nakako aimed to develop a software for voice training with visual feedback with characteristics like tone accuracy, tempo, voice quality and singing technics (vibrato).

The commercially available Singing Coach[1] software has been used in a number of studies in order to investigate the children's voice profile in a real educational environment in Greek Public Elementary Schools by proposing computer-based approaches to music education [3].

2.2 Critical approach

We appreciate that the last 30 years there is a rapid evolution concnering the functionality and new parameters incorporation in the design of applications for the practice of singing voice. For exemple, SINGAD has used only one parameter which is the detection of the fundamental frequency. ALBERT advanced to the maximum the memory offered by the rapid development of computers in the 1990s. Furthermore, the advancement in combining different parameters for targeting different practices, such as singing and speech therapy has concretised the design of the software. The SING&SEE mainly focused on aspects related to the same singers: fundamental frequency, identity vowel, and spectrogram. Then, the VOXed project introduced the WinSINGAD, which essentially combined the research parameters with those required by the musicians, namely the waveform, the fundamental frequency, various types of spectrograms and all of these in real time. Moreover, information from a camera was introduced for immediate feedback on the user's posture [9]. MiruSinger was innovated for the fact that combined two real human voices, one given from commercial CD and not one human and one simple synthetic sound. The study with the Singing Coach pointed out its friendly graphical user interface for the children.

In general, the optical feedback parameters have become more versatile and interdisciplinary over the years. This allowed software to become more usable and friendly to a wider range of users. For example, SINGAD designed only for the development of children's voices, ALBERT designed for wider applications and not only for use in schools, while SING&SEE and WinSINGAD were specifically designed for singers of all ages and levels.

3. THE FONASKEIN APPLICATION

FONASKEIN is a software application for real-time analysis of the singing voice with visual feedback. While the existing applications are limited to only two western scales (major and minor scales), FONASKEIN for the first time introduces the possibility to study and practice with non-equal tempered scales, such as the Byzantine or the ancient Greek scales. It also offers the opportunity to the user to enter a scale that is not included in the above or even to "build" its own scale. This can be achieved thanks to an "alteration mechanism" of each of the 12 notes to three semitones by cent detail.

3.1 Design and Graphical User Interface

FONASKEIN was designed and implemented in Max/MSP. Thus, its GUI presented in Figure 1 was designed with the capabilities of Max/MSP and includes seven different windows.

The fist window is the main bar at the top of the screen. It can hide or unhide the other FONASKEIN’s windows.

The second window is the audio control window. It is located on the left side of the screen. On this, the user can control the audio input and output. Additionally, he can choose whether to record his voice or hear a prerecorded sample. In this window, the user can also control the audio signal level, both for the playback and recording.

The third one is the tuning window located on the right side of the screen. It includes an automatic tuner that indicates using a color scale the deviation of the note that the user sings in comparison

The fourth window is one of the most important windows, the score window. It is located in the middle of the screen and it has two functions: First, it includes the main control buttons reset, play and stop. Secondly, it presents three score windows where the user can read the music piece on piano roll view, normal view or see what he sang.

[1] http://singingcoach.com/

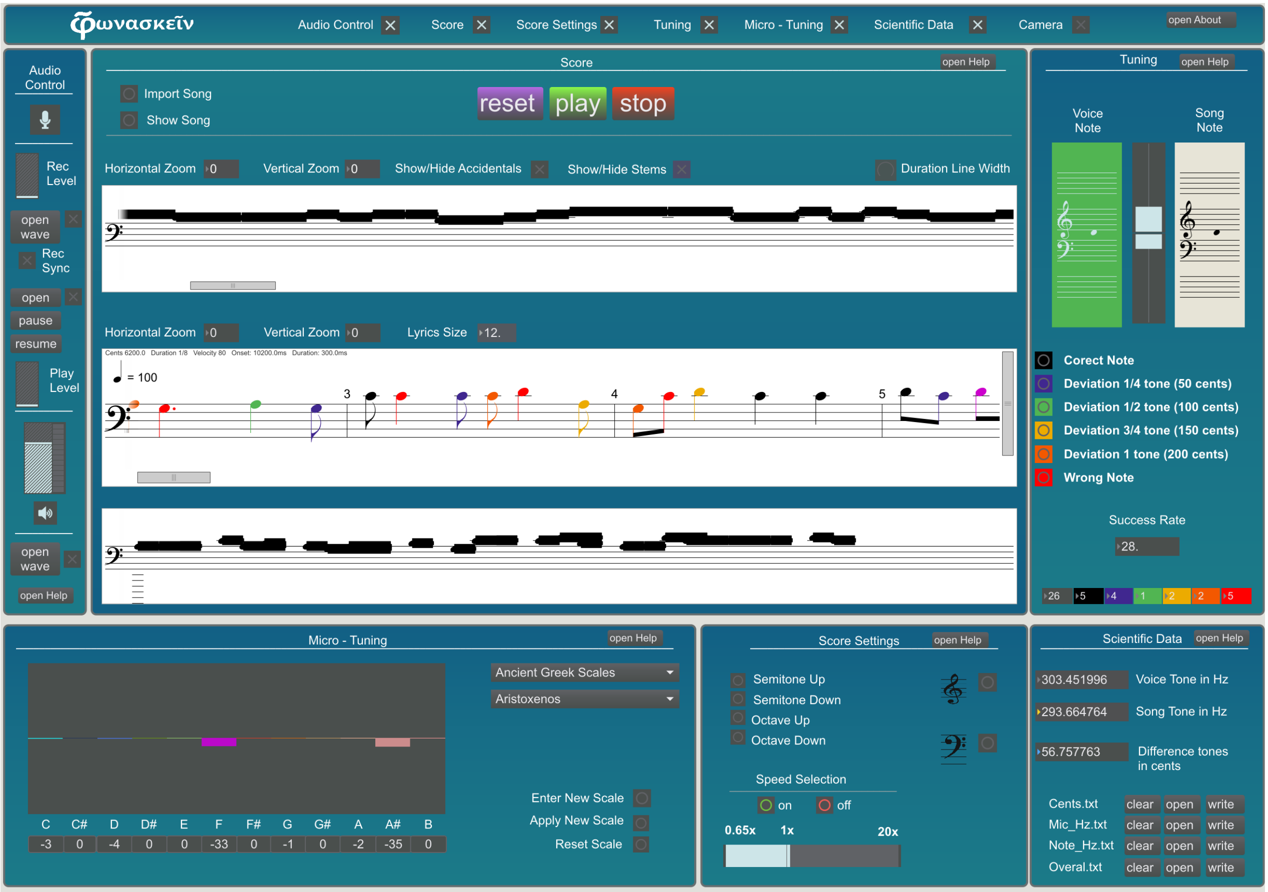

Figure 1. The Graphical User Interface of FONASKEIN.

The next three windows are in the bottom of the screen and their main function is the settings of FONASKEIN.

The window on the left side of the screen is a micro-tuning window. In this, the user can select one of the default scales. There are three categories of scales, Western, Byzantine and Ancient Greek. Each of these has its own subcategories. The user can also import his/her own musical scales by writing the deviation of each note in cent under the multi-slider. FONASKEIN has the capability to play the song with these micro-tuning scales allowing the user to hear the correct musical intervals.

The next window is the score settings window, it is located in the middle of the screen and it has three functions. The first one is the possibility to transpose the song semitone lower, a semitone higher, an octave lower and octave higher without affecting the micro tuning. The second possibility is to change the song's clef depending on the user's voice (bass clef for basses and tenors and treble clef for altos and sopranos). The last function is the speed selection, where the user can choose the speed of playback.

The last window is the data window which shows the current frequency that user is singing, the current frequency of the correct note and the deviation in cents. The user has the possibility of viewing and saving these data in *.txt files.

3.2 Architecture

The core of FONASKEIN includes two parts. The first is related to the analysis and transformation of the sound from the microphone signal and the second is dedicated to convert the MIDI file to score as well as the import, playback, and control of microtonal scales.

For the first part, we used a Max Object called fiddle~. The operation of the algorithm of fiddle~ is based on the number of peaks of the audio signal where each one finds the tone of height and intensity. Specifically, the incoming signal is broken into segments of N samples with N a power of two typically between 256 and 2048. A new analysis is made every N=2 samples. For each analysis, the N samples are zero-padded to 2N samples and a rectangular window Discrete Fourier Transform (DFT) is taken using a rectangular window [10].

The next step is to calculate the frequency F0. Fundamental frequencies are guessed using a scheme somewhat suggestive of the maximum-likelihood estimator. The "likelihood function" is a non-negative function L(*f*), where *f* is frequency. The presence of peaks at or near multiples of *f* increases L(*f*) in a way which depends on the peak's amplitude and frequency as shown:

$$L(f) = \sum_{i=0}^{k} a_i \, t_i n_i$$

where k is the number of peaks in the spectrum, a_i is a factor depending on the amplitude of the ith peak, t_i depends on how closely the ith peak is tuned to a multiple of *f*, and n_i depends on whether the peak is closest to a low or a high multiple of *f* [10].

The next step to build the FONASKEIN was the GUI in the score level. The Max/MSP does not support embedded objects with the creation pentagram, of notes and general notation. For this reason, we used not only an object designed by an external programmer, but a whole library comprising a large number of objects, the *bach library.*

The *bach library* is a cross-platform set of patches and externals for Max, aimed to bring the richness of computer-aided composition into the real-time world. In addition to that, it includes a large collection of tools for operating upon these new types and a number of advanced facilities and graphical interfaces for musical notation, with support for microtonal accidentals of arbitrary resolution, measured and non-measured notation, rhythmic trees and grace notes, polymetric notation, MusicXML and MIDI files [11].

As already stated, *bach* is a library of objects and patches for the software Max/MSP. At the forefront of the system are the bach.score and bach.roll objects. They both provide graphical interfaces for the representation of musical notation: bach.score expresses time in terms of traditional musical units and includes notions such as rests, measures, time signature, and tempo; bach.roll expresses time in terms of absolute temporal units (namely milliseconds), and as a consequence has no notion of traditional temporal concepts: this is useful for representing non-measured music, and also provides a simple way to deal with pitch material whose temporal information is unknown or irrelevant [12].

3.3 Non-equal tempered scales

One of the important novel features of FONASKEIN is its ability to import micro-tunings for singing in the Greek language. For the first time, the user is able to hear a song that is written on a different scale from that of western music while can exercise his voice on these interstices.

In FONASKEIN, as mentioned above, includes a field with twelve sliders, one for each note. The sliders are able to move ± 300cents that can vary each note by three semitones. When the user presses the *Apply New Scale* button, a simple yet lengthy process allows the introduction of interstices of the two graphical objects of the *bach* library.

When the user changes the slider of a note by x cents, then the program will have to move all those notes in all octaves at the same distance. To do this it needs to follow a series of steps. The first step should be to choose the notes. After that, a second instruction enters the change of the note. This command is Cents = Cents + X. In this way, all selected notes have changed by the same pitch with cent accuracy. The time it takes FONASKEIN to do this is just 94 milliseconds, which is less than 1/10 of a second.

4. EVALUATION METHODOLOGY

The goal of the evaluation is to measure the change of the tonal errors in singing voice by a number of participants after they practice with FONASKEIN in four songs with different music styles. The first song selected was "Ta paidia kato sto kampo" of Manos Hatzidakis (S1), a song written in the Western scale. The second was "Thalassaki", a song in Greek tradition scale Dorios (S2). The third song was "Apolitikion tou Staurou", a Byzantine hymn written in the First Mode (S3) and the last song was "Epitaph of Seikilos", an ancient Greek hymn written in 2nd century B.C. (S4). Eight postgraduate students of the University of Athens participated in the evaluation experiments. Among them, four were male and four female. Half of them are musicians.

The applied procedure follows the educational/training scenarios approach which is appropriate in testing computer-based tools in learning [13].The educational scenario takes place through a series of educational activities. The structure and flow of each activity, the role of the learners in it and their interaction with the interactive software are described in the context of the scenario [14].

Two activities were included in our evaluation scenario, each with two tasks. In the first one each participant receives four audio files made by FONASKEIN that correspond to the first seconds of the songs S1, S2, S3 and S4. The participants have to study themselves for a period of one week how to sing these songs, without any help and relying only on their ear. During the next task of this activity each participant sings the four songs he/she studied and the researcher digitally recorded their voices in a studio. Then the recordings were analyzed by FONASKEIN and the measured tonal errors constitute the comparison basis before the participants use FONASKEIN for training.

In the second activity first the participants were asked to practice the four songs using FONASKEIN for the same period of one week. They fully exploited both its features of micro-tuning and the capability of visual feedback in real time. During the second task participants sing the four songs using FONASKEIN.

Finally, the participants completed a questionnaire with their demographic details, included both their cultural background and their relationship with the music and the four songs.

5. RESULTS

The analysis of the measurements in both activities was based on the following number of notes for each of the four songs: S1=61, S2=49, S3=66 and S4=37. We used MS-Excel 2010 for all the statistical analysis of the measurements.

Figure 2 presents for each one of the four songs S1-S4 the average of the positive and the negative errors in cents for all the participants and for all the notes for the two activities, i.e. before (b) and after (a) using FONASKEIN for the training their singing voices, along with their standard error of the mean.

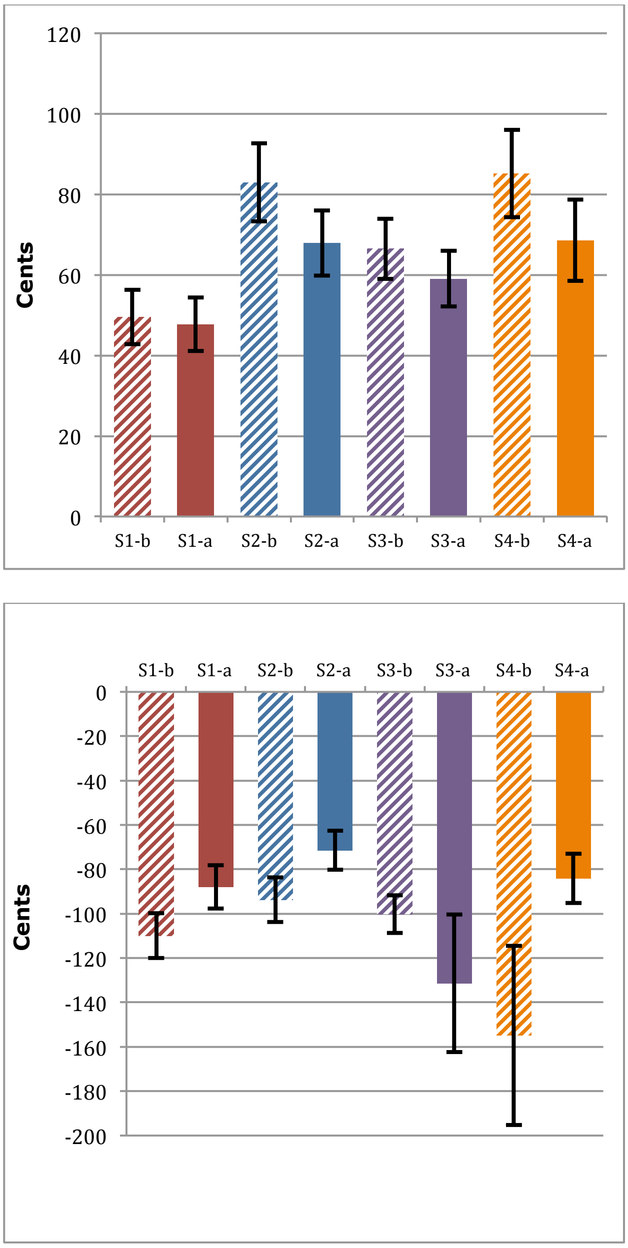

Figure 2. Average positive (above) and negative (bellow) errors in cents, for all the participants and for all the notes, before (b) and after (a) using FONASKEIN.

The number of negative errors was larger for all the songs. We observe a positive effect on using FONASKEIN as the number of errors was reduced in all the cases of the songs S1-S4. The largest improvement was for S4 (71 cents for the negative errors and 17 cents for the positive errors). The smallest improvement was for S1 (22 cents for the negative errors and 2 cents for the positive errors).

Figure 3 presents for each one of the four songs S1-S4 the average of the positive and the negative errors in cents for the participants who are musicians, for all the notes for the two activities, i.e. before (b) and after (a) using FONASKEIN for the training their singing voices, along with their standard error of the mean.

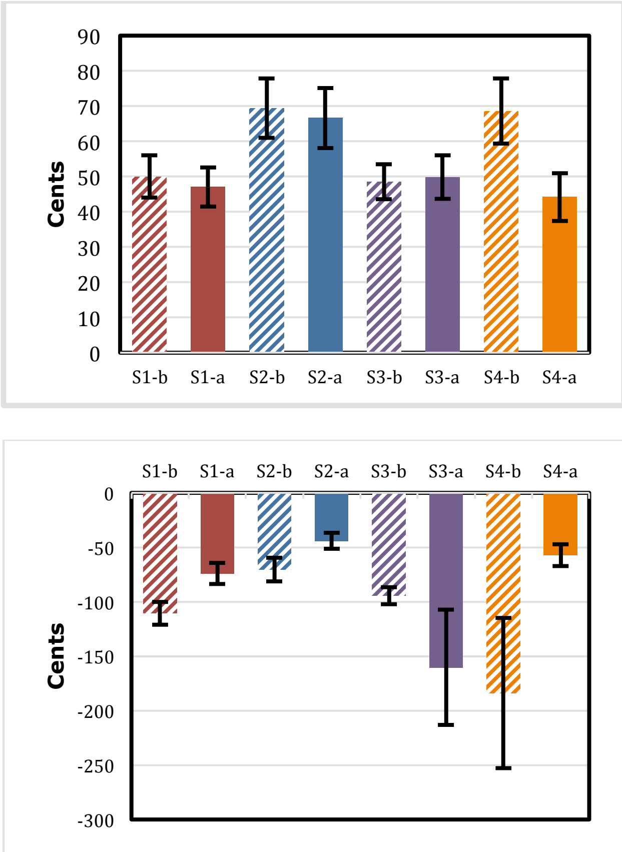

Figure 3. Average positive (above) and negative (bellow) errors in cents, for the participants who are musicians, for all the notes, before (b) and after (a) using FONASKEIN.

The number of negative errors was larger in almost for all the songs. We observe a positive effect on using FONASKEIN as the number of errors was reduced in all the cases of the songs S1-S4. The largest improvement was for S4 (126 cents for the negative errors and 24 cents for the positive errors). The smallest improvement was for S2 (3 cents for the negative errors and 27 cents for the positive errors).

Figure 4 presents for each one of the four songs S1-S4 the average of the positive and the negative errors in cents for the participants who are not musicians, for all the notes for the two activities, i.e. before (b) and after (a) using FONASKEIN for the training their singing voices, along with their standard error of the mean. The number of negative errors was larger for all the songs. We observe a positive effect on using FONASKEIN as the number of errors was reduced in all the cases of the songs S1-S4, but much smaller compared to the relative for musicians. The largest improvement was for S2 (15 cents for the negative errors and 18 cents for the positive errors). The smallest improvement was for S1 (3 cents for the negative errors and 8 cents for the positive errors) and equally for S3 (7 cents for the negative errors and 4 cents for the positive errors).

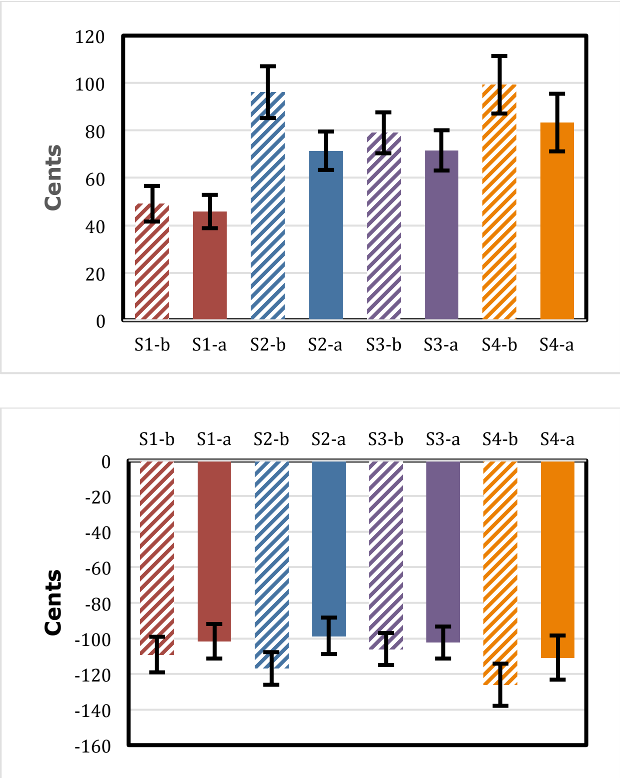

Figure 4. Average positive (above) and negative (bellow) errors in cents, for the participants who are not musicians, for all the notes, before (b) and after (a) using FONASKEIN.

6. CONCLUSIONS

We have presented the design and development of FONASKEIN, a novel modular interactive software application for the practice of singing voice in real time and with visual feedback for both equal and non-equal tempered scales.

The evaluation results of FONASKEIN in a pilot experiment with eight participants and with four songs in various musical scales showed its positive effect in practice of the singing voice in all cases.

In our future work we will study larger numbers of participants in various types of songs with non-equal tempered scales.

7. AKNOWLEDMENTS

We would like to aknowledge Panagiotis Velianitis, Professor Stelios Psaroudakes and George Chrysochoidisfor the precious assistance.

8. REFERENCES

[1] M. L. West, and M. Litchfield, "Ancient greek music," Clarendon Press, 1992.

[2] M. Chrysanthos, "Great Theory of Music" (1832) (translated by Katy Romanou), New York, 2010, pp.180-181.

[3] S. Stavropoulou, A. Georgaki, and F. Moschos, "The Effectiveness of visual feedback singing vocal technology in Greek Elementary School," in Proc. Int. Conf. Joint ICMC|SMC|2014, Athens, 2014, pp.1768-1792.

[4] G. Welch, C. Rush, and D. Howard, "Realtime visual feedback in the development of vocal pitch accuracy in singing," in Psychology of Music vol 17, 1989, pp.146-157.

[5] D. Rossiter, and D. Howard, "ALBERT: real-time visual feedback computer tool for professional vocal development," in Journal of Voice vol 10, 1996, pp.321-336.

[6] J. Callaghan, W. Thorpe, and J. Van Doorm, "The science of singing and seeing," in Proc. Int. Conference on Interdisciplinary Musicology (CIM04). Graz, 2004.

[7] G. Welch, E. Himonides, D. Howard, and J. Brereton, "VOXed: Technology as a meaningful teaching aid in the singing studio," in Proc. Int. Conference on Interdisciplinary Musicology (CIM04). Graz, 2004.

[8] T. Nakano, M. Goto, and H. Yuzuru, "MiruSinger: A Singing Skill Visualization Interface Using Real-Time Feedback and Music CD Recordings as Referential Data," in Proc. Int. Conf. Ninth IEEE International Symposium on Multimedia, 2007, pp.75-76.

[9] D. Hoppe, M. Sadakata, and P. Desain, "Development of real-time visual feedback assistance in singing training: a review," in Journal of Computer Assisted Learning, 2006, pp.308-316.

[10] S. Puckette, T. Apel, and D. Zicarelli, "Real-time audio analysis tools for Pd and MSP," in Proc. Int. Conf. ICMC, Cologne, 1988.

[11] A. Agostini, and D. Ghisi, "bachproject," http://www.bachproject.net/, Retrieved 24 April 2016.

[12] A. Agostini, and D. Ghisi, "Bach: an environment for computer-aided composition," in Proc. Int. Conf. ICMC 2012, Ljubliana, 2012, pp.373-378.

[13] C. Kynigos, and E. Kalogeria "Boundary crossing through in-service online mathematics teacher education: the case of scenarios and half-baked microworlds," in Proc. Int. Conf. ZDM Mathematics Education, Springer, 2012.

[14] C. Kynigos, M. Daskolia, and Z. Smyrnaiou "Empowering teachers in challenging times for science and environmental education: Uses for scenarios and microworlds as boundary objects," in Proc. Int. Conf. Contemporary Issues in Education, 2013.