Development of real-time visual feedback assistance in singing training: a review

D. Hoppe, M. Sadakata & P. Desain
Music, Mind, and Machine Group, NICI, Radboud University Nijmegen, HE Nijmegen, The Netherlands

Abstract

Four real-time visual feedback computer tools for singing lessons (SINGAD, ALBERT, SING & SEE, and WinSINGAD), and the research carried out to evaluate the usefulness of these systems are reviewed in this article. We report on the development of user-functions and the usability of these computer-assisted learning tools. Both quantitative and qualitative studies confirm the efficiency of real-time visual feedback in improving singing abilities. Having addressed these findings, we suggest further quantitative investigations of (1) the detailed effect of visual feedback on performance accuracy and on the learning process, and (2) the interactions between improvement of musical performance and the type of visual feedback and the amount of information it presents, the skill level of the user and the teacher’s role.

Keywords

real-time visual feedback, review, singing education.

Introduction

The classical method of teaching singing is typically based on a master–apprentice model, in which the teacher gives instructions and feedback on the performance of the student. This feedback is often given either on acoustic quality or a physiological aspect of singing performance, such as posture or use of the vocal apparatus. One of the common ways to do this is by providing the student with feedback using imagery, for example: ‘sing as if through the top of your head’. However, imagery remains susceptible to ambiguous interpretation. Additionally, the time lag between the student’s performance and the teacher’s feedback also contributes to the problem. Feedback from the teacher is hereby dissociated from the online proprioceptive and auditory sensations accompanying voice production (Welch 1985).

Welch (1985) developed a schema theory of the learning of the classical singing class as illustrated in Fig 1a. A model performance is presented by the teacher, which is followed by the student’s attempt to imitate the model as accurately as possible. After the imitation, the teacher gives feedback on the student’s performance. The student attempts to improve his/her performance on the basis of the teacher’s feedback. The effectiveness of this schema for instruction depends on the student’s ‘knowledge of results’ (KR), which is knowledge of the intended outcome of the sung response (e.g. how much the student knows about using the vocal apparatus to produce the requisite pitch). A drawback in this schematic is that the critical learning period (CP) is divided over two places in time. Learning occurs during the student’s response and during the teacher’s feedback. Welch has pointed out that the learning process should be optimal if these critical periods are reduced, particularly the time between receiving feedback and applying it. Real-time visual feedback (VFB) offers a solution by presenting the student with feedback on his/her performance during the response (see Fig 1b). Aside from reducing
CPs to be close to responses, presenting KR during the vocal response also offers the possibility of immediately modifying the response and simultaneously observing its effect.

In fact, the effectiveness of self-monitoring using real-time VFB for a better control of physical movements has been demonstrated in many fields. It has been applied for pedagogical use in different domains, such as pronunciation training (Arends & Povel 1991; Neri et al. 2002) and second-language acquisition (Dowd et al. 1998; Hirata 2004). Finding a way to present feedback in an immediate, understandable, and non-ambiguous way has been the primary drive to develop real-time VFB technology to assist in singing teaching. In this paper, we aim at clarifying the developments of the VFB systems for singing lessons.

Four real-time VFB computer tools for singing instruction will be described, including the experimental research carried out to evaluate the usefulness of these systems.

**Singing assessment and development (SINGAD)**

The pioneer project incorporating real-time VFB into educational software for singing was SINGAD (Howard & Welch 1989). The aim of the project was to develop a pedagogical tool for the assessment and development of the singing voices of UK Primary School children using real-time VFB technology. The software package was designed for a BBC microcomputer and made use of a peak-picker, originally designed for cochlear implants to make an accurate estimate of fundamental frequency (F0) in real time. To provide the singer with a musical reference, the system had libraries of note-sets in ones, threes, or fives. From these pre-set libraries, target notes were randomly picked and presented as a single sound with a fixed timbre through the audio speaker of the computer. The program was subdivided into two phases: an assessment phase and a development phase. In the assessment phase, a simple note was played to which the subject had to respond by singing into a microphone. The mean F0 of the whole sung response was then compared with the target note. In the development phase, a real-time trace of the F0 contour was plotted against time. Target notes could optionally be placed on the screen to guide the sung response.

The system was experimentally tested on 32 primary school children aged 7 years (Welch et al. 1989). In a pre-test, the participants had to sing eight randomly ordered, relatively low pitches. The initial pitching abilities of the subjects were recorded using the SINGAD assessment program. Next, a training procedure took place in which three matched groups were assessed. Group A used the SINGAD system and received help from a teacher, group B used the SINGAD system without teacher intervention and group C only received input from conventional singing teaching. The training intervention was followed by a post-test in which the students again sung the same eight pitches as in the pre-test. The responses of the participants were recorded using the SINGAD assessment program. A significant improvement in pitching ability was found for groups A and B, but not for group C. Especially, group A showed a larger effect. This difference between groups A and B can be explained by the role of the teacher in the development phase. The teacher was instructed to monitor the students’ learning progress and select new tasks accordingly by switching to an alternative development program at appropriate times. Participants from group B, who did not receive teacher’s assistance, were free to use any development program whenever they liked. This contrast between groups A and B suggests that providing the proper information at the appropriate phase of the learning process may be crucial for optimal learning. Nevertheless, a positive effect of real-time VFB on the development of pitching ability of primary
school children, both with and without the guidance of a singing teacher when compared with the traditional singing lesson, was a significant finding of this study.

In the early 1990s, with the development of more powerful computers, the improved version of the SINGAD system was introduced (Howard & Welch 1993). This second version of the system was made for the ATARI range of machines because it came standard with MIDI hardware. In comparison with the BBC version, three major changes were made. First, the spectral output of the computer was improved. Unlike the BBC microcomputer, which had only a single sound with a fixed timbre, many different sounds could be selected from an external MIDI synthesizer using the new system (see Fig 2). This way, the sound of instruments of common use in the singing studio could serve as stimuli, such as the baroque flute, piano, or guitar. Second, the ATARI had a larger memory capacity than the BBC microcomputer, allowing for the storage of complete F0 contours. Owing to the lack of memory capacity, the BBC microcomputer could only store the mean F0, which was taken from the complete sung response. Averaging the F0 of the complete sung response appeared to be problematic as it is often subject to pre- and post-onset noise or random vocalization. With this system, the data could be analysed by strictly measuring the performer’s response, thereby omitting any irrelevant data from the analysis. The last problem in the former system was that the stimuli were not musically relevant, as the target notes were randomly picked. This was improved by arranging the stimuli in three tonal categories (major, minor, and pentatonic). Each category contained a fixed set of sequences using permutations of three or five notes within the specific tonality. By selecting a category, the system would randomly order these sequences per trial. Figure 3 shows an example of the ATARI SINGAD assessment display. Although not formally evaluated, the ATARI system was used successfully in schools as a singing assessment tool (G.F. Welch, personal communication).

ALBERT

ALBERT: Acoustic and Laryngeal Biofeedback Enhancement in Real Time is another real-time VFB computer tool, which aimed at providing the developing singer with a tool to enhance vocal production (Rossiter & Howard 1996). In addition to providing VFB on the acoustic output, ALBERT included monitoring of laryngeal action. The system provided a greater diversity of VFB displays and parameters: F0, CQ (larynx closed quotient), spectral ratio, SPL (amplitude), shimmer, and jitter. Importantly, the VFB should not present very detailed information that is at too low a level to be useful to the developing singer. For example: providing VFB on the physiological action of the vocal apparatus may distract beginner singing students from the goal of the singing task: namely, the desired acoustic response. Instead, the information presented by the VFB should be task relevant. To meet this criterion, the system was built to enter user-defined parameters by combining any of the primary parameters. Through a visualization-control window, the user could configure the displayed information. Parameters could either be visualized in one, two, or three dimensions. For pedagogical purposes, the labelling attributes of the axes could also be altered or even hidden from the screen. Additionally, the size and colour of the display could be changed.

ALBERT was tested on two participants (Rossiter et al. 1996). The aim of the study was to identify the quality of voice production during VFB implementation, measured as the pattern of change during training lessons. The parameters used were CQ and spectral ratio (both parameters positively correlate with energy in the singer’s formant). Each participant had six lessons in which both speech and singing were assessed. One participant received only classical voice
training, while the other participant received classical voice training and VFB with the following three parameter conditions: (1) CQ; (2) spectral ratio; and (3) a combination of CQ and spectral ratio. Each condition lasted for two lessons successively. For each condition, the effect of the presented parameter on the level of both parameters (CQ, spectral ratio) in responses was measured. An increase in the levels of these parameters indicates the presence of a pronounced singer’s formant. In the single parameter conditions (1 and 2), the level of the parameter presented as VFB increased, while the level of the other parameter, which was not presented as VFB, remained unchanged. For the third condition in which both CQ and spectral ratio were presented, the measured level of both these parameters increased. Clearly, the VFB was contributing to an improvement in voice quality.

Interestingly, the changes in response accuracy showed the greatest effect when participants used the VFB for the first time (e.g. lesson 1, 3, and 5). Subsequent use only yielded a small effect. This effect was even more pronounced for the combined parameter condition, as subsequent use actually decreased response accuracy.

**SING & SEE**

Recently, the SING & SEE project was introduced at the Conference of Interdisciplinary Musicology (Callaghan et al. 2004). The project aimed at developing new VFB technology for the singing studio. The main features of research were the investigation of acoustic analysis techniques, methods of displaying VFB in a meaningful way and the pedagogical approaches for implementing VFB 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 project.

Four singing teachers and 21 of their students (seven beginning, 11 progressed, three advanced)
participated in the study. The study followed a within-subject pre–post-test design in which all students received the VFB during a 2-week intervention period. Because of different skill levels, the parameters used during the intervention differed between students. Moreover, the parameters used during the intervention also differed within groups because the teachers had their own teaching style. Each participant was recorded during the baseline, intervention, and follow-up periods. Both acoustic and perceptual measures were taken to quantify the changes in singing performance. After the follow-up period, all participants were interviewed on how the VFB had been integrated into the singing lessons, whether the software used had particular strengths or weaknesses and how to make possible improvements.

The paper reported an analysis of the interview data. All teachers were positive about the use of the pitch feedback. The spectrogram appeared to be more useful working with more experienced students than with beginners. More experienced and advanced students used the spectrogram for training timbre, dynamics, and the singer’s formant, whereas beginners only used it for note onset and offsets. The vowel quadrilateral appeared not to be appropriate for a high-pitched voice, being only suitable for an adult male voice. Importantly, the teachers suggested that the VFB displays should be more musically relevant. They also warned against the use of VFB without supervision, as students may not fully understand what they see. Also, teachers commented that the VFB is probably most efficient during specific stages in development.

As for the students, 90% were positive about the use of VFB in the singing lessons. They felt that the VFB improved their understanding of the desired outcome of the target model that they were required to imitate, as the feedback was immediate and unambiguous. Also, the use of the spectrogram received positive feedback. The vowel quadrilateral, on the other hand, met with the same criticism as from the teachers.

Recently, another study was conducted to examine a novel form of VFB and test its impact on singing using the pitch display from the SING & SEE software (Wilson et al. 2005). Besides testing the VFB’s impact on learning, this study also addressed the question of whether the amount of information presented by the visual display has an effect on singing performance. Fifty-six participants took part in this study, with skill level ranging from non-singer to trained singers. They were assigned to three groups. Groups A and B received VFB, while group C served as a control. Within each group participants, were divided into three subgroups according to their skill level. The distribution of skill level between the three groups was kept equal as far as was possible.

Group A trained with the pitch display from the SING & SEE program. This form of visual feedback presented the participants with a pitch trace, giving the

---

**Fig 4** (a) Display screen A. SING & SEE pitch grid, F0 against time. Target notes are indicated as dark bars. The pitch trace is indicated by the line; (b) display screen B. Target notes are indicated as light grey keys. Correct responses turn the target key into bright red. (The figure has adapted from Wilson et al. 2005).
user constant feedback on pitch history and target location (Fig 4a). Group B trained with a keyboard display (Fig 4b). Only when a note was correctly sung did the matching key light up. This form of VFB does not provide pitch history. It merely gives the user a specific right/wrong feedback. Group C was presented with the same display as group B, but without the pitch response feedback. In both a pre-test and a post-test, all participants were assessed on five test patterns of five interval sequences in upward semitone increments. Before the test recordings began, the pitch range of the test patterns was adjusted to the participant’s pitch range.

In addition to the effects of VFB on the learning process, the effect of VFB during the training period was analysed in this study. Comparisons were made between pre- and post-test as well as between pre-test and intervention. Surprisingly, participants in the experimental groups tended to worsen their performance when they received the VFB. Nevertheless, the comparison between the pre- and the post-test revealed that the experimental groups significantly improved their performance after the intervention compared with the control group (group C). Comparisons between the groups using VFB also indicated that the impact of the pitch display was remarkably larger for the non-trained singers, whereas the impact of the keyboard display was of greater use to the trained singers. Apparently, different skill levels require different forms of VFB.

**VOXed: WinSINGAD**

Besides the SING & SEE project, the VOXed project (Welch et al. 2004) was presented during the same Conference of Interdisciplinary Musicology. The project also incorporated real-time VFB for singing education, suggesting that there are growing interests in computer assistance in musical education. 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 searching 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 windows-compatible software tool WinSINGAD was developed, which is the successor of the early SINGAD systems (Howard et al. 2003, 2004). Because changes in vocal output over time are of primary interest to singers, the majority of displays are plotted against time. The parameters available were as follows: input waveform; F0 against time; short-term spectrum; narrow-band spectrogram; spectral ratio against time; vocal tract (VT) area; and mean/min VT area against time. Any parameter-window could be displayed on the screen in combination with another.
Also, a side-view Web cam could be selected for direct postural feedback (see Fig 5).

Two teachers agreed on using the VFB in their teaching lessons. Each teacher had four students participating (eight in total), two of whom were taught with the VFB and the other two serving as controls. The teachers were given full control over the way to use these parameters during their teaching lessons. Recordings were made during two sample lessons. In the first lesson, VFB was not implemented and served as a baseline. The second lesson fully established the use of the VOXed technology. Of each lesson, real-time observations as well as video recordings were made. Outside the lessons, both teachers and students gave semi-structured interviews. Thus far, this study has reported on quantitative analyses of sample lesson observational data, supplemented by qualitative commentary (Welch et al. 2005).

The results showed an overall positive appreciation of the use of VFB in the singing studio. Teachers were so keen on using the VFB that they even started to use it with other students as well, in one case, even with those who were initially assigned as controls. Concerning the implementation of the technology in the singing lessons, the teachers reported that they found the system to be user-friendly and non-obstructive to the normal course of teaching events. Of all displays, the spectrogram was most fully exploited by both teachers. Also, the facility to play back the sung response of the student had great advantages, as both teacher and student could now attend to the display and discuss its contents. Furthermore, the time-course analysis of the lessons confirmed the different teaching strategies between the two teachers. The amount of time that each teacher made reference to the VFB also differed. Overall, the introduction of VFB into the singing lesson was met with great enthusiasm.

Discussion

Four real-time VFB computer tools for professional voice development were reviewed in this paper. Over time, the original designs have been developed to provide the user with more information. SINGAD made use of a single parameter: fundamental frequency (F0). ALBERT maximally exploited memory capacity offered by the rapid development of computer hardware in the mid 1990s. A unique function of this system was that any up to three of the six primary parameters could be combined to form new parameters. In this way, the VFB could be altered so that it was most appropriate to the context of the task, be it singing training, pronunciation training, or voice therapy. SING & SEE also focused specifically on singer-related parameters: pitch (F0 against time), vowel identity (R1, R2), and timbre (spectrogram). The last project we reviewed, VOXed, introduced the successor of the early SINGAD systems, WinSINGAD, and included the widest range of singer-oriented parameters: input waveform; F0 against time; short-term spectrum; narrow-band spectrogram; spectral ratio against time; VT area; and mean/min VT area against time. A side-view Web cam could be selected for direct information about the student’s posture. In general, VFB features have become more multifaceted over time. This has allowed the programs to be accessible and useful to a wide range of singers. Although SINGAD was designed only for the child’s voice development, ALBERT aimed at a broader application that was not restricted to just singing training. Both SING & SEE and VOXed were specifically designed for singers of all ages and skill levels. We can conclude that the usability of these systems has improved over time by the addition of new functions.

Interestingly, the effectiveness of VFB on the learning process seems to depend on the amount of the student’s musical experience (as initially postulated by Welch 1985). In a controlled experiment, Wilson et al. (2005) investigated the effect of the amount of information presented by the visual display on the singing performance of both trained and non-trained singers. The keyboard display, which presented specific right/wrong feedback, was of greater use to the trained singers. The pitch display, which presented more detailed and contextualized information, was found to have a remarkably larger impact on singing in tune for the non-trained singers. However, Callaghan et al. (2004) found that the greatest use was made of the spectrogram, especially with progressed students: the spectrogram appears to be a hard visual display to interpret because it presents much information at once. Although these two studies showed that the amount of information displayed in the VFB interacts with the skill level of the student, exactly how these two factors relate is not yet clear, and therefore requires further investigation.
In designing ALBERT, careful consideration was made in constraining the amount of information that was shown as VFB. To make sure that the VFB showed information relevant to the task, users were given the possibility to enter new parameters by combining any of the six primary parameters. Rossiter et al. (1996) investigated the impact of this feature on the singing process during training lessons. Single and combined parameter display conditions were evaluated. The combined parameter condition showed the greatest improvement in performance accuracy during initial use. However, during subsequent use, the same condition yielded a negative effect. Also, greater appreciation of the feedback was found during initial use of all feedback conditions, in contrast to a general decline in performance accuracy during subsequent use. Apparently, quality of performance does not always increase during the use of VFB. In fact, Wilson et al. (2005) reported that participants tend to worsen their performance during the use of VFB, although an increase in singing accuracy from pre-test to post-test was observed. These findings are counter intuitive, as one would expect a linear improvement in performance accuracy during the use of VFB. Thus, the actual impact of VFB on the singing process and its relation to the learning process still have to be understood.

A difficulty in evaluating educational tools is often to balance the fundamental research and its application to real situations. In most studies that we reviewed, preserving natural singing class conditions seems to overrule controlling experimental factors. For example, in their study, Welch et al. (1989) trained the experimental groups to sing single notes, whereas controls had to sing songs, just as in a common primary school singing lesson. As the pre- and post-assessment procedures used a task similar to the experimental training procedure, the experimental groups had a clear advantage over the controls during the final assessment. It might be the case that the nature of the training task can explain the observed results of this study. Another point is that the pattern of social interaction differed between the experimental groups and the controls. While the experimental groups worked in pairs or threes, the controls had to sing together as a group. In such group-singing activity, it is very likely that a phenomenon such as ‘social loafing’ might occur: the tendency to exert less effort on a task when working as a part of a co-operative group than when working on one’s own (Latane et al. 1979). The observed results of this study might therefore be influenced by the pattern of social interaction. In contrast, for example, Wilson et al. (2005) conducted a well-controlled experiment with many participants. Based on the large number of observations, they have shown that VFB does, in fact, significantly enhance learning to sing in tune. However, the number of such controlled experiments is rather small. Many of the valuable qualitative findings should therefore still be evaluated quantitatively.

Another finding from the SINGAD experiment was that improvement takes place in pitching ability when VFB is used even without supervision. The program could therefore serve as a replacement of singing teachers instead of just being used as a helpful tool to singing teachers. However, in observational studies, greater improvement was found when teachers assisted with VFB. Furthermore, Callaghan et al. (2004) reported that their participant teachers were actually against the use of VFB without supervision, as students may not understand fully what they see. In their study, the spectrogram appeared to work very well and was very informative, but only with the teacher’s assistance. This real-life application in the singing studio suggests that VFB may be more effective when it is properly understood, and for this, teacher’s assistance is helpful.

As in other voice development research domains, such as pronunciation training (Neri et al. 2002) and second-language acquisition (Dowd et al. 1998; Hirata 2004), we addressed several studies showing that VFB helps in learning to sing. It is noteworthy that VFB serves well as a tool for assistance, rather than a replacement of the singing teacher. Accordingly, VFB technology has been met with great enthusiasm from professional singing teachers. For future research, further quantitative research on the detailed effect of VFB on performance accuracy and on the learning process is necessary, as well as a closer investigation of its interactions with the type of VFB and the amount of information it presents, the skill level of the user and the presence/absence of a teacher.

Of valuable insight to the field might be the research on feedback and motor skill learning with respect to ‘focus of attention’. The effect of the learner’s ‘focus of attention’ on the learning process was reviewed by Wulf and Prinz (2001). They showed that an internal focus of attention, which is directed to ‘one’s own movements’,
appears to be less beneficial to the learning process than an external focus of attention, which is directed to ‘the effects of one’s movements’. Accordingly, VFB on singing performance that is directed to one’s own movements (e.g. the vocal tract) may be less effective than VFB on the acoustical output (e.g. real-time spectral information). Indeed, taking the phenomena of attentional focus into consideration accounts for some previous findings, such as the interview data from Welch et al. (2005), where singing teachers preferred the spectrogram to all other feedback options. Furthermore, the option for the ALBERT user to define new parameters that are most relevant to the task can also be interpreted in terms of an ‘external focus of attention’, as the goal of a singing task is to reach ‘a desired effect from the movement of the vocal apparatus (e.g. the desired acoustic output)’. A further understanding of internal and external focus of attention in relation to real-time VFB learning would therefore enrich the field.

Acknowledgements

This research was supported by the Technology Foundation STW, applied science division of NWO and the technology program of the Ministry of Economic Affairs (NNN6301). We would like to thank Alex Brandmeyer for English editing.

References


