In Margaret S. Barrett, Gary E. McPherson, & Rosalynd Smith (Eds.) *Children and Music: Developmental Perspectives*, pp105-111. Proceedings of IMERS-99 International Music Education Research Symposium, Launceston, February 4-7, 1999.

COMPUTER-ASSISTED VISUAL FEEDBACK IN THE TEACHING OF SINGING

Jean Callaghan[1], William Thorpe[2] & Jan van Doorn[3]

INTRODUCTION

Learning to sing commonly uses a master-apprentice approach relying on modelling, coaching, exploration, reflection and feedback. As Collins (1991) has pointed out, computer technology is well adapted to this kind of apprenticeship training. Recent studies suggest that musical performance skills depend largely on practice and self-regulated learning, activities greatly assisted by feedback (Butler & Winne, 1995; Weidenbach, 1996). The use of visual feedback of voice and articulation parameters to assist in the teaching of singing has been advocated for at least 30 years (Scott, 1968) and is now being used in some tertiary education institutions (Miller & Franco, 1991; Nisbet, 1995; Miller & Doing, 1998).

In speech pathology and second-language learning visual feedback is better established as an instructional tool. For instance, commercial systems such as IBM Speech Viewer and the Kay CSL have been specifically developed for clinical applications. Visual feedback in these areas allows learners to be actively involved in making conscious judgments about their own speech, it assists with explaining abstract concepts of speech articulation, it provides more consistent feedback than a listener with more opportunities to practise speech production skills (Maki, 1980), and it provides enhancement of the speech perception process (Alvarez et al, 1998).

It has been surmised that for singing visual feedback would provide information to supplement the student's own aural and kinaesthetic feedback and the verbal feedback provided by the teacher. Indeed, a group from the Music Department of Drew University, New Jersey advocating the use of technology in singing teaching has claimed that 'the differential between the student's perception and acoustic reality is a massive roadblock to the learning process' (Nair, 1997). This group is highly enthusiastic about the introduction of visual feedback of spectrographic features to assist students of singing to

[1] School of Contemporary Arts, University of Western Sydney, Nepean, PO Box 10, Kingswood NSW 2747, Australia.
[2] National Voice Centre, University of Sydney, Australia
[3]School of Communication Sciences and Disorders, University of Sydney, Australia

overcome that block, but to date there are few evaluative studies of visual feedback in singing teaching. One study by Rossiter et al. (1996) compared the effectiveness of teaching using visual feedback of vocal parameters with conventional teaching methods. That study concluded that instruction using visual feedback was able to produce more consistent and sustained improvements in parameters related to vocal performance than a conventional singing instruction method.

Evaluative studies of visual feedback of voice and speech parameters in any area of speech pathology, second language learning or singing teaching will need to address several questions: Do methods that use visual feedback provide better results (e.g. improved pronunciation, or more consistent production of the singer's formant) than conventional methods? Or can they obtain the same results more quickly? Will users readily accept the invasion of computer technology into the clinic, classroom or studio? Despite the importance of this last question to the successful introduction of visual feedback there have been few studies of this aspect in either speech pathology or second language learning. Most evaluative studies have concentrated on the effectiveness of the device from an outcomes perspective (e.g. Pratt, Heintzelman & Deming, 1993; Nouza & Madlikova, 1998). However, one study that has looked at learners' perceptions of using visual feedback in second language learning reported that they found it a positive experience (Oster, 1996).

The current study looks at the introduction of technology into the singing studio environment from the important perspective of user experience, asking the following questions:

> Is it feasible and productive to utilise computer technology for the purpose of assisting the learning of singing ?
>
> Can a simple visual feedback of voice parameters, using existing speech analysis technology, provide significant benefits to students who are learning to sing ?

METHOD

Participants

Four singing teachers and eight students participated in the research. The teachers - two female and two male - ranged in age from 46 to 63 years and were all well qualified and highly experienced. All taught in private studio and three, in addition, taught in other settings (university, conservatorium, fee-paying secondary school). Their experience with computing ranged from none at all to usage for word-processing, spread-sheets and e-mail. Characteristics of the teachers are summarised in Table 1. The teachers worked with their own students: three girls (aged 14, 15 and 16), one boy (aged 12), one adult female (aged 28) and three adult males (aged 31, 28 and mid-40s). The students exhibited a range of computer literacy, from having no computer experience at all to being highly-sophisticated computer users. Characteristics of the students are summarised in Table 2.

Procedure

The teachers were provided with an initial training session of approximately two hours in duration where they were given instructions and practice in using the computer equipment. Prior to the training, all the teachers had some familiarity with phonetic symbols and with the concept of formants; they had little detailed knowledge of acoustic analysis. They were therefore provided with some background information on acoustic analysis techniques, to assist them in interpreting displays such as spectrograms and formant charts. During this session, the teachers were asked to decide how they could use the system within their lessons.

Each teacher gave one lesson to each of two of their students. They used one or more of the feedback modalities described below during at least part of the lesson. Lessons were conducted in a studio at the University containing the computer system and an electronic piano (Clavinova, Yamaha). One of the investigators was present at the beginning of each lesson to ensure that the equipment was operational, and thereafter was available on call in an adjacent room to provide assistance with the computer equipment if required. At the completion of their two lessons, each teacher was interviewed by one of the investigators.

Equipment

The equipment consisted of the Computer Speech Lab (Kay Elemetrics) with the microphone mounted on a stand. For this study four of Kay's programs were invoked: CSL (for spectrographic analysis), Sona-Match (formant and vowel analysis), Visi-Pitch (pitch display), and Visi-Pitch Games (visual feedback through a games-oriented model).

The software provided by Kay provides both menu and keyboard control, with the ability to assign particular operations to individual keys. A set of key commands was defined to encapsulate the range of actions which had been identified in the training session as being of use. A template describing the action of each key was printed and provided to the teachers. As much as possible, each action was programmed into a single key-press, and dialog boxes were disabled or bypassed to try to avoid any distractions to the teacher.

For spectrographic display, a delayed-time display was employed (making use of the CSL standard software). The frequency range was set to 5 kHz to focus mainly on the voiced regions of the spectrum, with the option of either wide-band (600 Hz) or narrow-band (30 Hz) analyses. This program was used to provide feedback about types of voice onset, sustained voice quality (singer's formant intensity), and vibrato patterns.

The Sona-Match program provides two real-time modes: a chart of vowel identity with respect to a graph of Formant 2 versus Formant 1 frequencies; and a display of vowel resonance peaks, showing a plot of sound intensity versus frequency at each instant in time. The spectral intensity is computed via the LPC algorithm. The frequency display therefore represents mainly the formant resonances of the sound, with the first two formants being automatically located for use in the vowel chart. The spectral resonance display was used to give feedback about the voice quality, especially with regard to maintaining consistency across the pitch range. A set of Italian vowels was installed (Ferrero et al., 1978) in addition to the standard (American) English vowel charts. Alternatively, the teacher could demonstrate a desired vowel and the student try to match the resulting target area on the display.

The Visi-Pitch program gives real-time display of pitch and sound intensity. Two displays were provided so that the teacher could display and replay two sets of data. The pitch range was initially set at 0-600 Hz, but could be adjusted by the teacher if necessary. The Visi-Pitch games program provides a range of modules that give feedback of pitch and vocal intensity through manipulation of pictures on the screen. Both these programs were used to give visual feedback about the magnitude of pitch excursions. Some use was also made of the comparative features of the dual displays in the Visi-Pitch program to give feedback on the vibrato in the teacher's and in the student's voice.

RESULTS

Usefulness

After their initial training in the use of the technology, the participant teachers made their own decisions about what aspects to employ and how to incorporate them into their two trial lessons. Teacher A taught her two teenage students together using all modules, while the other three taught individual lessons using slightly different aspects of the technology to meet the differing needs of their students. The formant and vowel analysis module was used in all eight lessons, the pitch display in three, and the games in four. The aims of the lessons varied from correcting specific vowels, through increasing awareness of the singer's formant, teaching vibrato, and helping to eradicate the pitch break in a male voice at peak of mutation, to improving poor-pitch singing (see Table 3).

All four teachers found the computer-assisted visual feedback useful. Reactions varied from: 'Yes, it is useful, but it wouldn't work with all students' (Teacher A) to 'Enormously useful, extremely exciting!' (Teacher C).

Teacher C, who worked to correct mispronunciation of specific vowels, used the analysis on speech-quality singing and was delighted with the results, observing that Student 5 managed to achieve in that one session what they had together been working on all year. With Student 6, he modelled correct vowels, recorded them, and got the student to match those, achieving a marked improvement there also. Teacher D commented, 'I like the way you can relate what appears on the screen to the front/back of the mouth and to the position of the tongue. Students can understand that that has a direct correlation with what they are doing with their muscles.'

All teachers found the formant analysis extremely useful. The spectrographic analysis was adapted to teaching different types of vocal onset; to teaching production of the singer's formant for different vowels and on changing pitch; to teaching vibrato; and to showing dynamics. Teacher B was delighted with Student 3's achievement in respect to vibrato: 'Basically, she learnt the difference between an amplitude vibrato and a pitch vibrato. It helped her a lot: she'll remember that.'

Difficulties

While all teachers found the vowel analysis useful, they also found it frustrating in some respects. Teachers A and B, who worked with female adolescents, found it inadequate for the high female voice. Teacher B commented that the American vowels were not ideal for classical singing and Teachers A and D that the Italian vowels provided were perhaps not correct for singing. Also, the recognition of sung vowels on the vowel chart was poor for some voices, particularly at high pitches. The fact that the red dots indicating the location of vowels ('blood splatters' as they were commonly termed) remained on the screen made for confusion after a number of attempts.

Teacher B, who hoped that using the spectrograph to compare the upper and lower range of the changing voice would assist in blending the range, found the display inadequate for that purpose.

Teachers expressed frustration that none of the modules was adequate for specific work on poor-pitch singing. Teacher D, who worked with Student 8 specifically on pitch, felt that aural prompting of the target pitch was required for the visual feedback to be really useful and that the pitch targets presented in the 'Hummingbird' game should have been assigned specific pitches.

DISCUSSION

Results of this initial study certainly suggest that it is both feasible and productive to utilise computer technology for the purpose of assisting the learning of singing. As Teacher C said:

> The students found that being able to plot things visually was an enormous bonus - having that immediate feedback. Normally in the studio, they do something, I listen, they stop, I give feedback, they try again. That's a laborious process. There's also a communication difficulty - I can't guarantee that what I feed back they interpret exactly as I mean it. Also there's the frustration and feeling of being judged that comes from the human interaction. The machine is non-judgmental and working that way is perceived as fun. Children of this generation are very comfortable with machines and trust them. It's mutual exploration rather than my telling them what to do.

All the participant teachers found the technology useful and made constructive suggestions both on possible improvements and on its use in the teaching/learning process.

Teacher D pointed out that our use of a fixed microphone on a stand was very inflexible and could be confronting to a student. Getting the correct angle for using the microphone while also being able to see the monitor caused some problems and affected head position. A microphone attached to a head fitting was suggested. Teachers C and D also suggested incorporating aural feedback to the right ear.

Teacher B pointed out that using the technology meant losing eye contact with the student, since both were relating to the computer monitor rather to each other. All the participant teachers were conscious that although in a lesson it was possible to achieve quick results using the computer, time and practice would be needed for the student to acquire the 'muscle memory' necessary for motor learning. They therefore suggested that the technology would be of most use for student practice sessions: 'It is terrific to get a student to make an immediate change through the visual feedback, but in achieving muscle memory it would be great to be able to put a disk in and say "this is the model of your best effort - go and practise matching it".' (Teacher B). None of the teachers felt they would use the technology in all lessons or for all students.

While the results suggest that computer literacy did not impact on the usefulness of the technology, Teacher B cautioned that 'it's only as good as the teacher is good at analysing the information'.

All the students participating in this initial investigation derived some benefit from the simple visual feedback of voice parameters using existing speech analysis technology. It is clear, however, that further modifications of that technology are desirable. The difference between spoken vowels and sung vowels has been long acknowledged (Titze, 1995). The need in singing to produce specific pitches, a range of vowels in different languages and appropriate to different styles, and the modification of vowels to meet aesthetic demands all require modification of the technology designed for speech.

REFERENCES

Alvarez, A. Martinez, R. Gomez, P. & Dominguez. (1998). A signal processing technique for speech visualisation, *Proceedings of ESCA Workshop on Speech Technology in Language Learning, Marholmen, Sweden*, pp. 33-36.

Butler, D. & Winne, P. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245-281.

Collins, A. (1991). Cognitive apprenticeship and instructional technology. In L. Idol & B.F. Jones (Eds.), *Educational values and cognitive instruction* (pp. 121-138). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ferrero, F.E., Magno-Caldognetto, E., Vagges, K. & Lavagnoli, C. (1978). Some acoustic characteristics of the Italian vowels, *Journal of Italian Linguistics*, 3(1), 87-96.

Maki, E.M. (1980). Visual feedback as an aid for speech therapy. In J. D. Subtelny (Ed.), *Speech Assessment and Speech Improvement for Hearing Impaired* (pp 167-176). Washington, DC: The Alexander Graham Bell Association for the Deaf.

Miller, D. & Doing, J. (1998). Male *passaggio* and the upper extension in the light of visual feedback. *Journal of Singing*, 54(1), 3-14.

Miller, R. & Franco, J.C. (1991). Spectrographic analysis of the singing voice. *The NATS Journal*, 48(1), 4-5, 36.

Nair, G. (1997) www.users.drew.edu/~gnair/philosophy.htm

Nisbet, A. (1995). Spectrographic analysis of the singing voice applied to the teaching of singing. *Australian Voice*, 1, 65-68.

Nouza, J & Madlikova, J. (1998). Evaluation tests on visual feedback in speech and language learning, *Proceedings of ESCA Workshop on Speech technology in Language Learning*, pp. 151-154. Marholmen, Sweden.

Oster, A.-M. (1997) Auditory and visual feedback in spoken L2 teaching, *Phonum* 4, 145-148.

Pratt, S. , Heintzelman, A. & Deming, S. (1993). The efficacy of using the IBM SpeechViewer vowel accuracy module to treat young children with hearing impairment, *Journal of Speech and Hearing Research*, 36, 1063-1074.

Rossiter, D., Howard, M., & DeCosta, (1996). Voice development under training with and without the influence of real-time visually presented feedback. *Journal of the Acoustical Society of America*, 99(5), 3253-3256.

Scott, A. (1968). A study of the components of the singing tone utilizing the audio spectrum analyzer, *The NATS Bulletin*, May, 40-41.

Titze, I. (1995). Voice research: Speaking vowels versus singing vowels, *Journal of Singing*, 52(1), 41-42.

Weidenbach, V. (1996). *The influence of self-regulation on instrumental practice.* PhD dissertation, University of Western Sydney, Nepean.

Table 1: Participant Teachers

	Age	Gen-der	Training for Singing Teaching*	Teaching	Years Teach-ing Sing-ing	Com-puter Skills	Famili-arity with Acoustic Analysis Displays	Famili-arity with Vowel Quad.
Teacher A	63	F	Grad. Cert. Continuing professional education	Private studio	15	No	Little	No
Teacher B	46	F	MA Continuing professional education	Private studio Fee-paying secondary school	10	Yes	Little	Yes
Teacher C	46	M	Diploma Continuing professional education	Private studio Conserva-torium	30	No	Yes	Yes
Teacher D	51	M	Continuing professional education	Private studio Univer-sity	20	Yes	No	Little

* In addition to musical training and performance experience

Table 2: Participant Students

	Age	Gender	Singing Training	Computer Skills	Familiarity with Acoustic Analysis Displays	Familiarity with Vowel Quad.
Student 1	15	F	4 years	No	No	No
Student 2	14	F	3 years	No	No	No
Student 3	16	F	18 months	No	Yes	Yes
Student 4	12	M	9 months	Yes	No	No
Student 5	28	F	5 years	Yes	Yes	No
Student 6	31	M	10 years	Yes	Yes	No
Student 7	40s	M	2 months (+ years earlier)	Yes	No	No
Student 8	28	M	9 months	No	No	Yes

Table 3: Results

	Lesson Aims	Usefulness	Difficulties	Serendipity
Teacher A/ Student 1	Vowels Twang Breath control	Limited Yes Yes	Not suitable for high voice; It. vowels not quite right	Leapfrog game useful for sustained voicing
Teacher A/ Student 2	Vowels Twang Dynamics	Limited Yes Yes	Not suitable for high voice; It. vowels not quite right	Leapfrog game useful for sustained voicing
Teacher B/ Student 3	Vowels Vowel modification Vibrato Onset Effect of postural changes	Limited Limited Very useful Useful Useful	Not suitable for high voice. Eng.-lang. vowels unsuitable	Learnt difference between pitch vibrato and amplitude vibrato
Teacher B/ Student 4	Voice change Effect of postural changes Italian vowels	No Yes Limited	Teacher unable to read relevant information Seemed incorrect for singing	Games proved useful for strengthening the voice
Teacher C/ Student 5	Vowels (at speech level)	Extremely useful	Would like singing vowels for a range of languages	
Teacher C/ Student 6	Vowels (at speech level) Singer's formant	Extremely useful Useful	 Would like real-time feedback and aural feedback	
Teacher D/ Student 7	Italian vowels Effect of postural changes	Limited Useful	Only suitable for speech Fixed microphone limiting	Useful for maintaining the singer's formant across vowels and across leaps
Teacher D/ Student 8	Pitch	Inadequate	Lack of specific pitch Fixed microphone limiting	Hummingbird game was of some use Effects of high larynx and poor posture showed on spectrograph