Taylor & Francis healthsciences

WinSingad: a real-time display for the singing studio*

David M. Howard[1], Graham F. Welch[2], Jude Brereton[1], Evangelos Himonides[2], Michael DeCosta[3], Jenevora Williams[4] and Andrew W. Howard[5]

From the [1]Media Engineering Research Group, Department of Electronics, University of York, UK, [2]School of Arts and Humanities, Institute of Education, University of London, UK, [3]Professional singing teacher, York, UK, [4]Professional singing teacher, Guildford, UK, and [5]SARAND Business IT Solutions, Haverhill, Cambridge, UK

Received 26 January 2004. Accepted 30 June 2004.

Logoped Phoniatr Vocol 2004; 29: 135–144

This paper describes the nature and implementation of a specially-designed integrated real-time display that is undergoing evaluation as part of a recently funded innovative pilot project to investigate the relative usefulness of computer displays in the singing studio. Following previous work that suggests that simple displays of a small number of analysis parameters are generally likely to be the most effective, the system makes available a range of complementary analyses that are plotted against time. These relate to: fundamental frequency, spectrum, spectral ratio, and vocal tract area. These can be viewed singly, multiply or in combination using a panel based design within the PC Windows environment, known as WinSingad. The algorithms used are described and the displays themselves are illustrated with results gained from the pilot phase of the research to indicate their potential usefulness.

Professor David M Howard, Department of Electronics, University of York, Heslington, York, GB-YO10 5DD, UK. Tel.: +44 1904 432405. Fax: +44 1904 432335. E-mail: dh@ohm.york.ac.uk

INTRODUCTION

Professional singing teaching is overwhelmingly based upon a qualitative process in which a technique continues to be handed down from generation to generation of practitioners using a master-apprentice model. The success or otherwise of this technique in developing an appropriate singing voice for an individual pupil relies almost exclusively on the teacher's 'craft knowledge'—a combination of the musical perception, imagination and experience, that normally derives from their own performance career. Imagery is a key technique that is commonly employed (1, 2) in the form of 'psychological hooks' or concepts such as: 'sing on the point of the yawn', or 'sing as if smelling a flower', or 'sing as if through the top of your head', which are used to promote postural gestures associated with appropriate outputs. Whilst such analogies or psychological hooks may enable students to produce an 'appropriate' vocal output, few if any describe the physical actuality of the voice production process (3). Callaghan's survey (4) of Australian teachers of singing, for example, reported three different types of imagery evidenced in practice (visual, kinaesthetic, aural). However, the experiential evidence of the efficacy of imagery continues to need corroboration in terms of well-controlled formal experimentation.

Over the past twenty years, there has been increased interest from the scientific community in quantifying a number of aspects of the singing and speaking voice with respect to voice education and development (5–7). One aim of such endeavour is to find some reliable, quantifiable measures of singing performance that can be used to replace and/or enhance the present rather diverse terminology, imagery and psychological devices used in the singing studio. In general, scientific and artistic approaches to musical activities tend to use different language codes and symbolization for knowledge, and whilst it is not known to what extent these language codes are reconcilable, the benefits from the application of technology have been demonstrated in many other fields, including the arts (8). Real-time computer-based visual feedback offers a very powerful tool to aid the development of any skill that requires scaffolded practice by repetition, particularly if the student can be left to work with it alone after appropriate demonstration and input with the teacher. Previous work (6, 9–11) has

* Paper presented at PEVOC-5 in Graz, Austria, August 2003

 ISSN 1401-5439

DOI: 10.1080/14015430410000728

demonstrated that singing development can be enhanced when quantitative feedback is provided, such as the direct monitoring of singing during lessons with the teacher. There is further advantage to be gained if students are able to use the technology to quantify aspects of their progress lesson by lesson, and enhance the quality of their practice time between singing lessons.

Taking pitch as an example, Welch (12, 13) develops a model to characterize the learning process as illustrated in Fig. 1. During the traditional teacher-student interaction, a target is provided by the teacher (e.g. verbal or sung), the student makes an attempt to imitate this vocally, and the teacher provides feedback. The gain the student makes by utilizing the teacher's feedback depends on knowledge of what s/he is supposed to be achieving in terms of the intended outcome, an external assessment referred to as 'knowledge of results' or 'KR' (see Fig. 1A).

The student will make another attempt following the feedback, as illustrated in Fig. 1B, and this process of attempt (such as a musical phrase, figure or section) followed by feedback forms the backbone of the traditional singing pedagogical process. In contrast, real-time visual feedback offers the possibility of providing feedback *during* the student's vocal response. Modifications can be made immediately and their concurrent effect observed immediately (see Fig. 1C). Apart from the more obvious advantage of removing the time lag between a vocal response and the feedback that is inevitable without real-time provision, the student is able to make another attempt immediately as observations of the feedback provided during the previous attempt give an immediate indication as to what needs to be altered.

Quantifiable parameters have been identified that vary with training and experience for: a) actors (14), b) adult singers (5), c) children (15), as well as d) girl and boy cathedral choristers (16). Real-time visual feedback has been previously used successfully with primary school children (11) and adult singers (10, 17). Our experience suggests that technological applications are only of potential benefit if they are easy to use by non-specialists and provide information that is meaningful, valid and useful. Such robust information, associated with robust knowledge of results can then underpin feedback to provide assessments that are more accurate in formative and summative evaluation contexts.

OVERALL PROGRAM DESIGN

One key feature that emerges in order to ensure that displays employed for real-time visual feedback are meaningful, valid, useful, and fit for purpose, is that they are kept simple, so that they can be readily understood

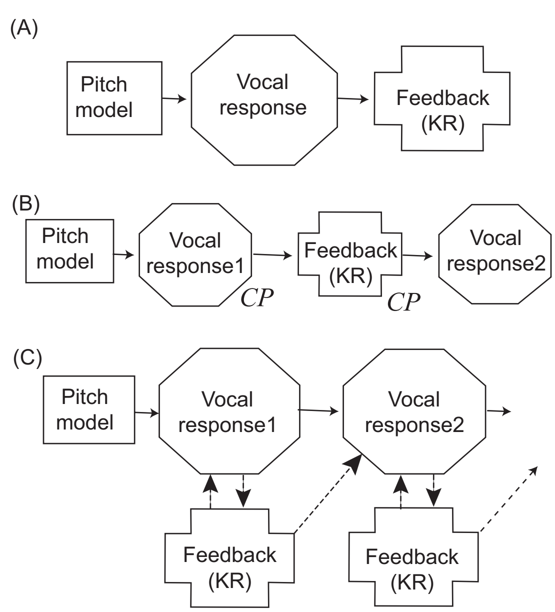

Fig. 1. An illustration of the learning process for pitch in singing based on Welch (11, 16). Key: (A) the basic interaction between teacher and learner; (B) the on-going traditional learning process, and (C) the way in which real-time visual feedback can impact the learning process. KR = knowledge of results from an external source; CP = critical learning period; time is from left to right.

by users. Simplicity underpins the current implementation which takes inspiration from the SINGAD (18) and ALBERT systems (6). The basic requirement is for displays of individual, measured parameters against time, where users can select either single or multiple displays depending on their current needs. Presentation of each display is designed to be clear and uncluttered, and whenever multiple displays against time are employed, the time axis remains common to all parameters on display across the screen. The feedback provides support to the user on a temporary basis during lessons and/or practice; it does not become embedded in the feedback loop and it can be removed at any time.

The software is written in Microsoft Visual Studio C++, and known as WinSingad. Audio input is via a typical soundcard and a standard PC multimedia electret microphone. Displays are presented individually in panels within the program window, and the user can decide which are visible and in what order they are offered on the screen. When there is more than one panel visible on the screen, any individual panel can be made the most prominent in size terms by enlarging it with respect to the others via a single mouse click. Each panel has a properties dialog associated with it through which alterations can be made to any user-adjustable parameters associated with a particular display, as well as the thickness of the plot lines and the colours of the background, foreground, text and plot lines.

WinSingad maintains a standard and an advanced set of menu options for panel selection. In the standard mode, the panels considered to have direct application for users are made available. The advanced mode is provided primarily for use by programmers, but it is left available to all users. It makes available a number of additional waveform panels to enable the operation of the various stages through the analysis algorithms to be confirmed by viewing their output waveforms. These displays also enable checks to be made in the event of unexpected display outputs.

Audio input is via the soundcard, and recorded audio data can be saved in *wav* format to provide a permanent data record as well as the possibility of further analysis in the future. Full compact disk bandwidth is used (16 bits at 44.1 kHz sampling rate), and the overall length of recording is limited only by the available memory. Input data is currently monophonic to enable a speech pressure output from a microphone to be recorded. There are plans to use a stereophonic input in the future to enable the output from the electrolaryngograph to be recorded, enabling the most robust fundamental frequency estimation (19) to be made as well as larynx closed quotient (5) to be measured.

The choice of parameters to display is based primarily upon experience with SINGAD and ALBERT, where choices were made that were based in turn on quantitative experimental analysis of singers and actors in training. The majority of the displays are plotted against time, since it is the time course of a parameter that is generally of interest when taking account of how it varies as the vocal output is changed. The displays currently available are: Input waveform; Fundamental frequency against time; Short-term spectrum; Narrow band spectrogram; Spectral ratio against time; Vocal tract area; and Mean/min vocal tract area against time.

Input waveform

The input waveform is captured in 20 ms buffers from the soundcard and processing is carried out on a buffer-by-buffer basis. In order to enable users to view the raw input waveform, a real-time display of the current input buffer is available in its own panel. This oscilloscope-like display is often used to illustrate: a) periodicity and non-periodicity, b) the effect of making louder and softer sounds, c) the effect of changing pitch, d) the result of projecting a sound, and e) the differences between various vowels (see Fig. 2). Under advanced mode, a number of other waveform panels are available that are associated with the processing algorithms, and these are discussed below.

Fundamental frequency

Fundamental frequency (f0) measurement of speech signals has been the subject of research since around the 1920s and, to date, there is no one algorithm that will measure f0 reliably for any speaker in any acoustic situation. The appropriate choice of f0 analysis technique should be made with respect to the nature of the input signal, the acoustic conditions in which the system will be used and the analysis output errors that are acceptable and those that are not (19).

WinSingad currently makes use of a time-domain peak-picking algorithm based on (20), which was originally developed for use in cochlear implant hearing aids (21). Particular features of this algorithm that make it

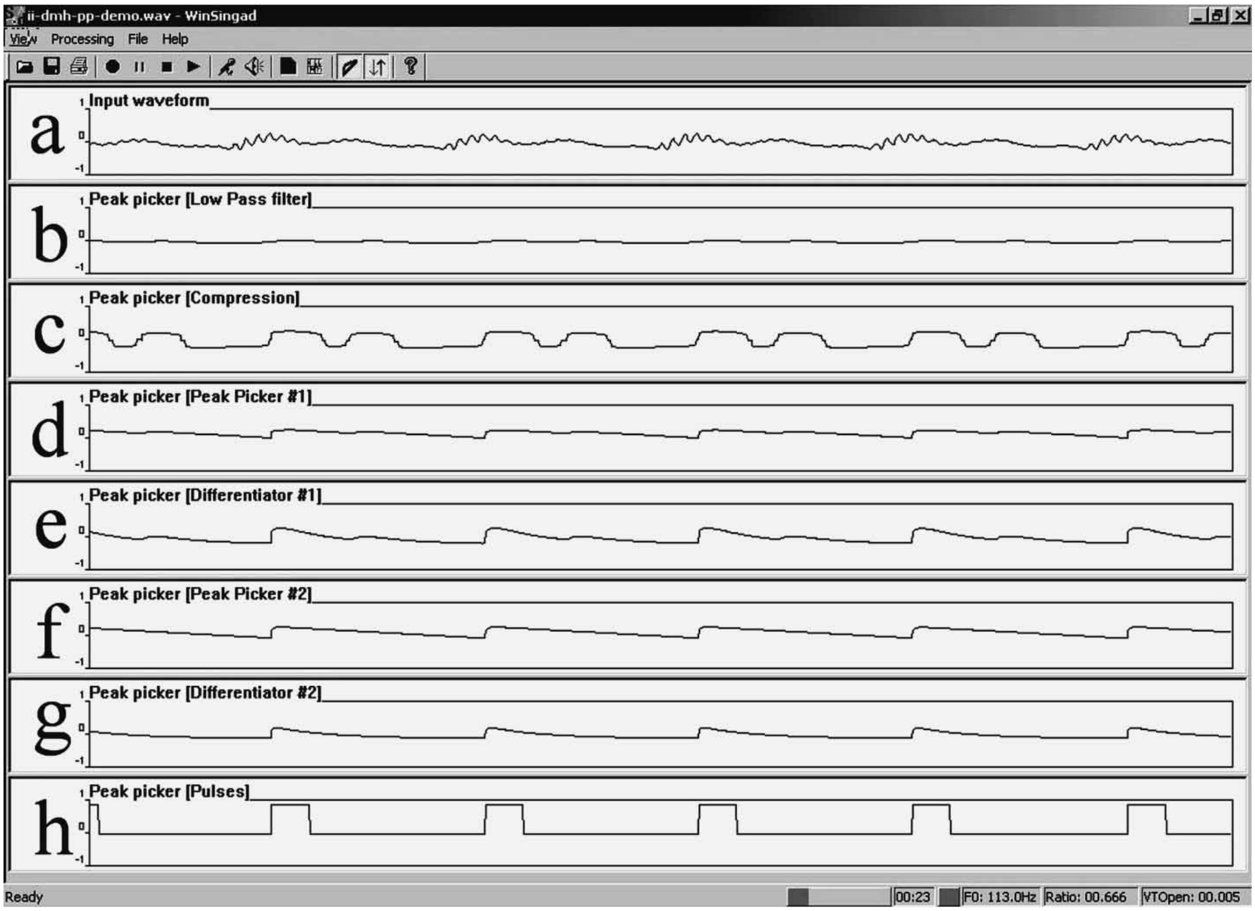

Fig. 2. Waveform panels for the vowel in 'heard' sung by an adult male to illustrate the operation of the peak-picker for fundamental frequency estimation. Key: (a) input waveform, (b) low-pass filtering, (c) amplitude compression, (d) positive peak-picking, (e) time differentiation, (f) positive peak-picking, (g) time differentiation, (h) pulse output.

appropriate for use in this application include: a) accurate cycle-by-cycle measurement of f0; b) an f0 output is always given for a sound input even if the voice quality is irregular (e.g. rough or hoarse); and c) the f0 output is essentially instantaneous—the algorithm incorporates only a four-sample memory (~90 µs). Potential difficulties include: a) excessive reverberation—avoided by placing the microphone off-axis, close to the lips of the subject, and b) competing local acoustic noise. An output f0 smoothing option has been implemented in WinSingad, providing a choice of 3, 5, 7, and 9 point median smoothing. This has the additional advantage of offering more resilience to background acoustic noise in the context of its application in the singing studio.

The operation of the peak-picking algorithm can be summarized as follows with reference to Fig. 2, which is a screen from WinSingad to show the peak-picker operating waveform panels (labelled a–h). This screen is provided for illustration purposes and to enable the operation of the peak-picker to be reviewed. The raw input waveform (a) is first low-pass filtered (b) to leave a few harmonics in voiced sounds and little or no energy for voiceless sounds, and then amplitude compressed using a logarithmic amplifier (c) to reduce the effect of significant amplitude variation associated with, for example, formant or fundamental frequency changes. This waveform is then subject to positive amplitude peak-picking (d) and time differentiation (e) to reduce the relative amplitude of any secondary peaks with respect to the major peak in each cycle. Another peak-picker (f) and time differentiation (g) follows, and finally a comparator is used to produce one pulse per cycle (h). The f0 display is produced from this pulse train output (h) as the reciprocal of the period of each cycle (f0 = 1/period).

Figure 3 shows a plot of the logarithm of f0 against time for a tenor singing a rising and falling glide and a two octave ascending scale on /i:/; exercises that might typically be employed when warming up the voice. No smoothing has been employed.

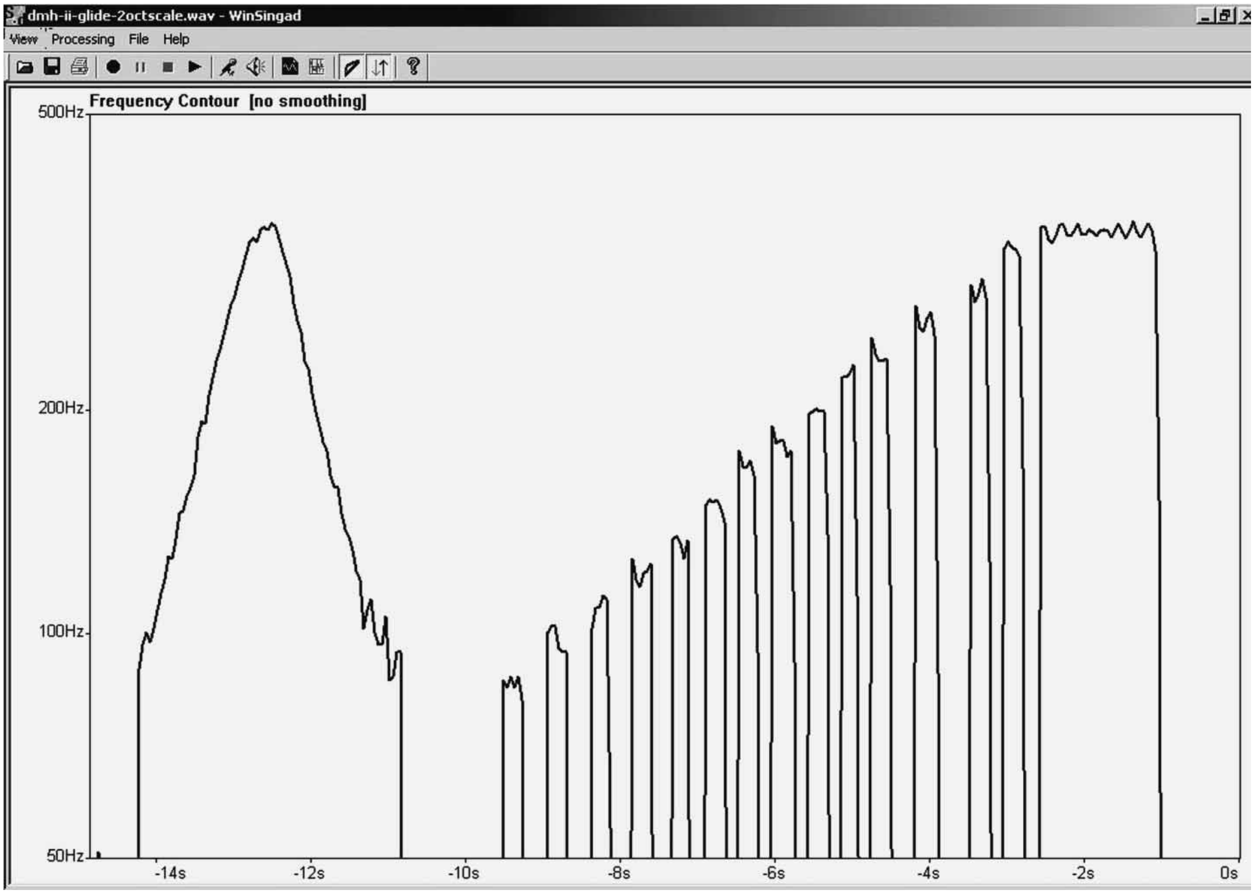

Fig. 3. WinSingad panel showing fundamental frequency against time for a glide and a 2-octave ascending scale on the vowel /i:/ sung by a tenor.

Spectrum, spectrogram and spectral ratio

Spectral analysis is a commonly applied technique to enable the nature of the acoustic output to be explored (22). WinSingad offers both a spectrographic display (frequency against time, with amplitude shown as the darkness of markings) and a spectral display (amplitude against frequency). Following a) the importance of developing this skill in spectral modification for opera singers (7, 23), and b) the design objective behind WinSingad of keeping displays simple, a display of the ratio of acoustic energy in the 2–4 kHz band to the energy in the full band signal is also available. This measurement has the operational advantage in that it will always remain between 0 and 1, providing that the full band frequency extremes are not within the singer's formant band. The frequency band extremes can be set by the user—the program ensures that the full band upper and lower frequency limits are equal to or outside the singer's formant frequency band limits. The default setting is for the full band and the singer's formant band to be set to (100 Hz to 4000 Hz) and (2000 Hz to 4000 Hz) respectively. These values can be changed by the user.

Figure 4 shows a WinSingad spectrogram (upper) for the vowel /i:/ sung by a tenor in a non-projected and a projected style on two notes an octave apart (lower note, upper note, lower note). The spectrum of the final vowel (projected) is shown in the centre panel, and the spectral ratio plot is at the bottom of the figure. The spectrogram and spectrum displays indicate the frequency range used as the singer's formant band in the ratio display with horizontal and vertical lines respectively, and in this example they are set at 2 kHz and 4 kHz. (The 4 kHz line is not showing on the spectrogram figure above as it is beyond its upper plotted boundary here.) The appearance of acoustic energy in this range is apparent for the projected production compared with the non-projected production, and this is made singularly clear in the ratio display.

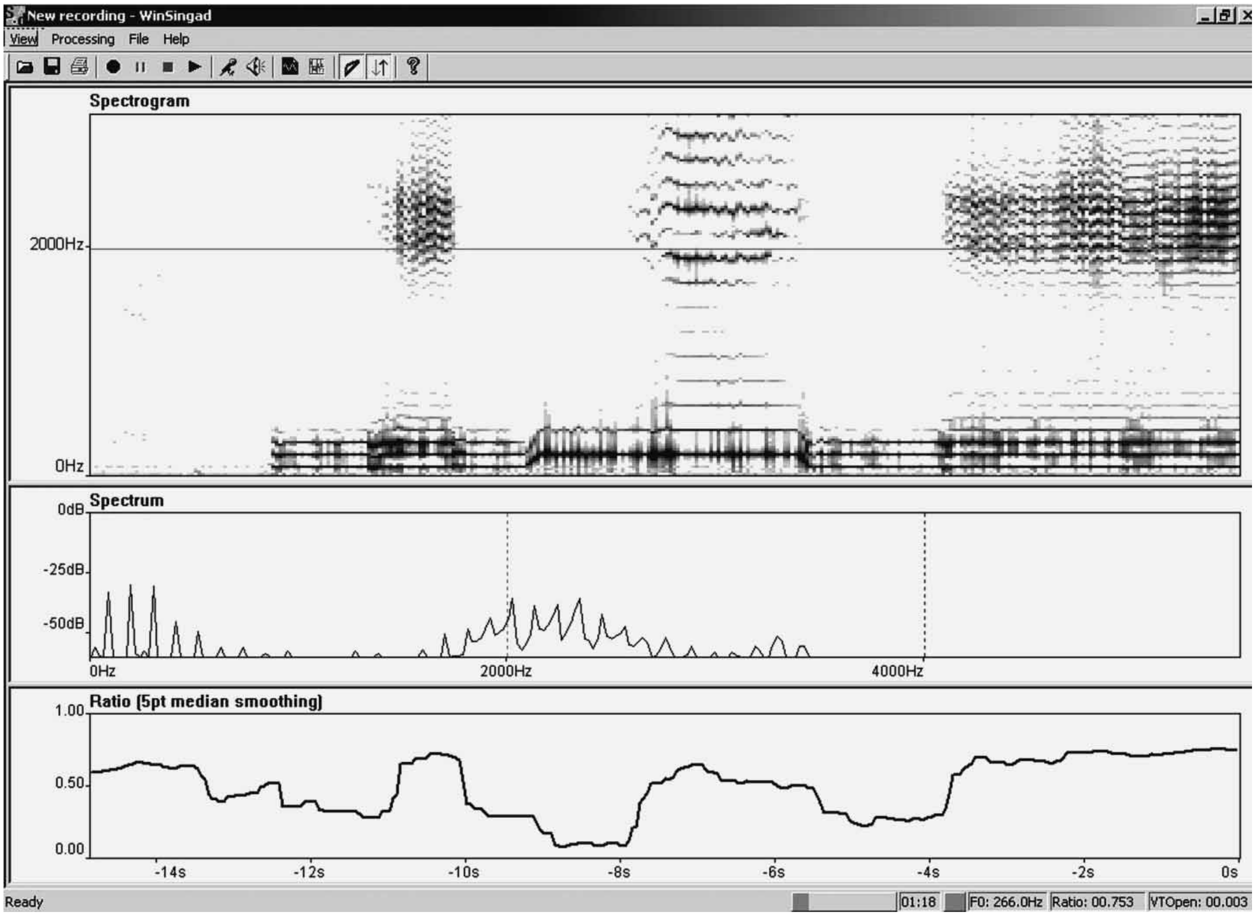

Fig. 4. WinSingad panels showing a spectrogram (upper), spectrum for the final vowel (centre) and ratio (lower) for the vowel sequence /u:/, /a:/, /i:/ sung by a tenor in a non-projected and projected style. Key: Solid and dotted lines indicate the position of the non-projected and projected output respectively. Ratio is a display of the ratio between the acoustic energy in the 2–4 kHz band and that in the full band signal.

Vocal tract area

A display of the vocal tract area can be obtained via a lattice filter model derived from a linear predictive analysis of the vocal output (14). This models the vocal tract in terms of either the areas, diameters or radii of a set of equal length tubes between the larynx and the lips. The lower plot in Fig. 5 shows an example vocal tract area display for the vowel in the second syllable of the word 'comfort' from bar 8 of 'Comfort ye' from Handel's Messiah, which is sung unaccompanied by a tenor on B3 (247 Hz). The glottis and lips are at the left and right edges of the plot respectively.

There are, however, limitations associated with this representation. Firstly, it only strictly models non-nasal voiced sounds, due to the assumptions employed in linear prediction. Secondly, the output area values are arbitrary in absolute terms. They are usually, therefore, normalized either to a fixed glottis width, or to a fixed maximum value. Finally, there are situations where more than one set of tube areas provides a solution and results can be presented that could not be articulated by a human vocal tract. Due to the integrated nature of the solution itself, it is not obvious how it might be constrained, for example, to vocal tract configurations that are physically possible. It is felt for this reason that summary plots of the average, minimum or maximum vocal tract area against time would be potentially more beneficial, and these are also available in WinSingad. The upper plot in Fig. 5 shows a plot of the mean vocal tract area against time.

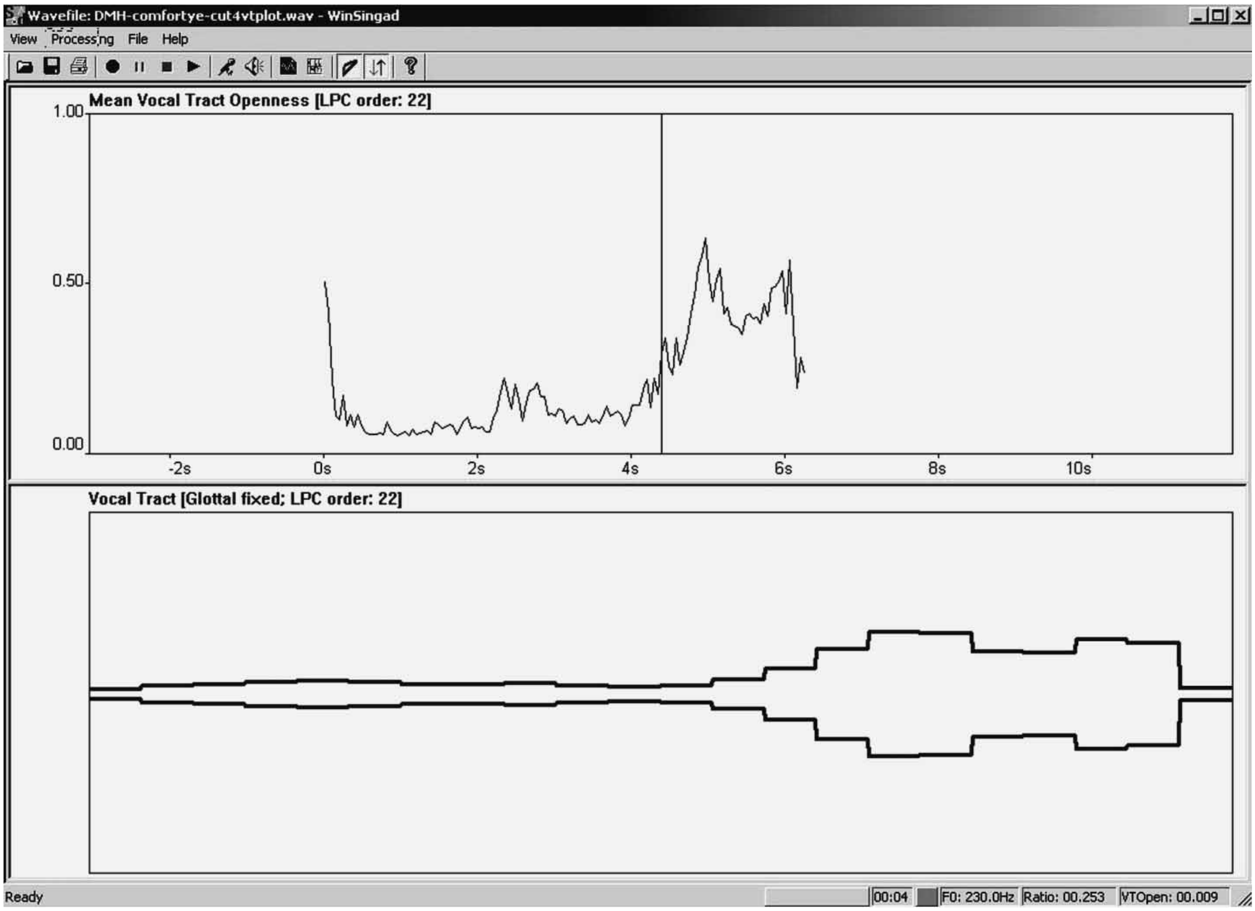

Fig. 5. Upper: display of the mean vocal tract area for the vowel in the second syllable of the word 'comfort' in 'comfort ye' in bar 8 of the tenor recitative of that name from Handel's Messiah. Lower: vocal tract area display for the vowel in the second syllable of the word comfort, which is sung unaccompanied by a tenor on B3 (247 Hz)—the glottis and the lips are at the left and right hand side of the plot respectively.

INITIAL RESULTS IN THE SINGING STUDIO

The WinSingad system has been in use in a York singing studio with four adult students (two female, two male) for an initial period of 2 months. Three of these students are studying mainly operatic repertoire and the other has been working on classical songs and songs from the theatre musical repertoire. All started learning singing as adults and all sing as soloists at an amateur level and/or perform in amateur opera or musicals. Each has been learning with this teacher for over a year. The teacher's main aim is to instil 'good' vocal technique by providing knowledge of vocal production through clear explanations, the use of diagrams from standard text-books, anecdotal evidence and demonstration. The teacher himself is an operatic baritone with many years experience both as a teacher and internationally renowned performer.

The lessons take place in a singing studio located within a medieval bartizan tower on York's city walls. This is a two-storey structure, and it has been converted to provide a waiting area on the ground floor which is separate from the main singing studio on the upper floor. The studio area is reasonably large and is furnished with a baby grand piano. The teacher is aided by a piano accompanist in most lessons, except when working on basic vocal technique with beginner students. Generally, the student stands at the far end of the piano facing toward the pianist and the teacher sits or stands to the student's left–hand side. The laptop computer running WinSingad is mounted on a specially commissioned stand, which is placed to the student's right-hand side. The teacher is able monitor the display from some distance and draw attention to particular points of interest during exercises or pieces. The laptop stand is adjustable to eye-level so that the student can also monitor the display whilst s/he is singing without adversely affecting normal singing posture and retaining a vertical spine position. For the first five weeks of the trial, the microphone was placed approximately 1 m from the singer on the same stand as the laptop or on an adjacent chair. Occasionally it was tried on the piano, but interference from the piano tended to dominate the displays. For the last three weeks of the trial, a tie-clip microphone was used.

The teacher generally uses WinSingad to monitor a student's vocal performance during both warm-up exercises and work on repertoire, all of which are usually accompanied by the piano. The displays generally employed by the teacher are the spectrogram, vocal tract area and ratio displays. The fundamental frequency contour is not typically used due to output fundamental frequency errors caused by the presence of acoustic interference from the piano accompaniment. The spectrogram, vocal tract area and ratio displays are also subject to potential errors due to the presence of the piano accompaniment, but this is not considered to be a problem since 1) the singer's acoustic output at the microphone is at a much higher level that that of the piano, and 2) the piano has no significant spectral energy (and certainly no spectral peak) in the spectral region of particular interest (2 kHz to 4 kHz).

To date in this study, work is progressing towards the general improvement of vocal resonance with all four of his students. When using the spectrogram display, students are directed to aim for 'bright colours' which equates to increased high frequency energy relative to the low frequency region (and therefore higher spectral ratio values in the singer's formant frequency region). This frequency region is indicated on the WinSingad display by two ratio boundary lines (lower and upper), set at 2000 Hz and 4000 Hz respectively. The teacher prefers to use the spectrogram display in colour as he finds that yellow and white, which are used to plot high-energy peaks, correspond well with the imagery he commonly uses when talking about the 'brightness' and 'bright colour' of the voice. He also makes use of the vocal-tract area display to check that the student's vowel sounds are 'placed at the front'. This is indicated in the display as a broadening towards the lip end of the vocal tract area display (e.g. towards the right hand side of Fig. 5). After initially finding the ratio display more difficult to understand, he now makes use of it more often to reinforce the information given by the spectrogram display.

An example

Figure 6 shows the spectrogram and ratio display of an excerpt from the end of 'Love, could I only tell thee', words by C Bingham and music by JM Capel sung by one of the students (baritone). The thin and thick black horizontal bars on the figure represent the position of the words 'thou' and 'art' respectively. When this recording was made, the teacher was monitoring the screen as well as listening in order to both see the visual correlate whilst also hearing the resonance in the voice.

Shortly after the start of the last note, at the point marked 'A', the teacher gave verbal feedback and encouragement to the student (saying 'go on') whilst pointing to the screen display. The student then turned to look at screen and focused on improving his resonance in the note. The increase in resonance towards the end of last note can be readily appreciated in both the spectrogram and the ratio display. In this case, the student was able to gain a clear indication from the display in real-time when more resonance in the sound had been achieved, and the feedback enabled him to hold that vocal tract posture for the remainder of the note.

DISCUSSION AND CONCLUSIONS

A set of displays for real-time feedback in the normal working context of a singing studio has been described. Appropriate control over processing and display parameters is provided to the user via standard Windows menus and dialog boxes. The displays themselves are organized as individual panels that can be viewed alone or in any combination. In this way, attention can be drawn to individual parameters or to multiple parameters as familiarity and confidence grows, and areas of interest can be zoomed in on as desired. The displays are designed to provide visual feedback that is simple to interpret whilst being relevant to the needs of the developing singer and the teacher.

The research methodology employed involves the teachers and their students working as co-researchers with the research team by keeping diaries of progress and activities during lessons, practice sessions (students) and lesson planning (teachers), known as an action research methodology. Huge enthusiasm has been shown on the part of the teachers and students to use the software on a regular basis. Their initial feedback has all been very positive, and in many cases an element of surprise has been expressed when discovering the potential of the use of technology in this context. All students using the software have said that they have found the software easy and interesting to use. The facility to record and play back the sung sounds has brought many advantages, such as enabling the teacher to play back to the student points of interest in a particular piece or exercise, when they can

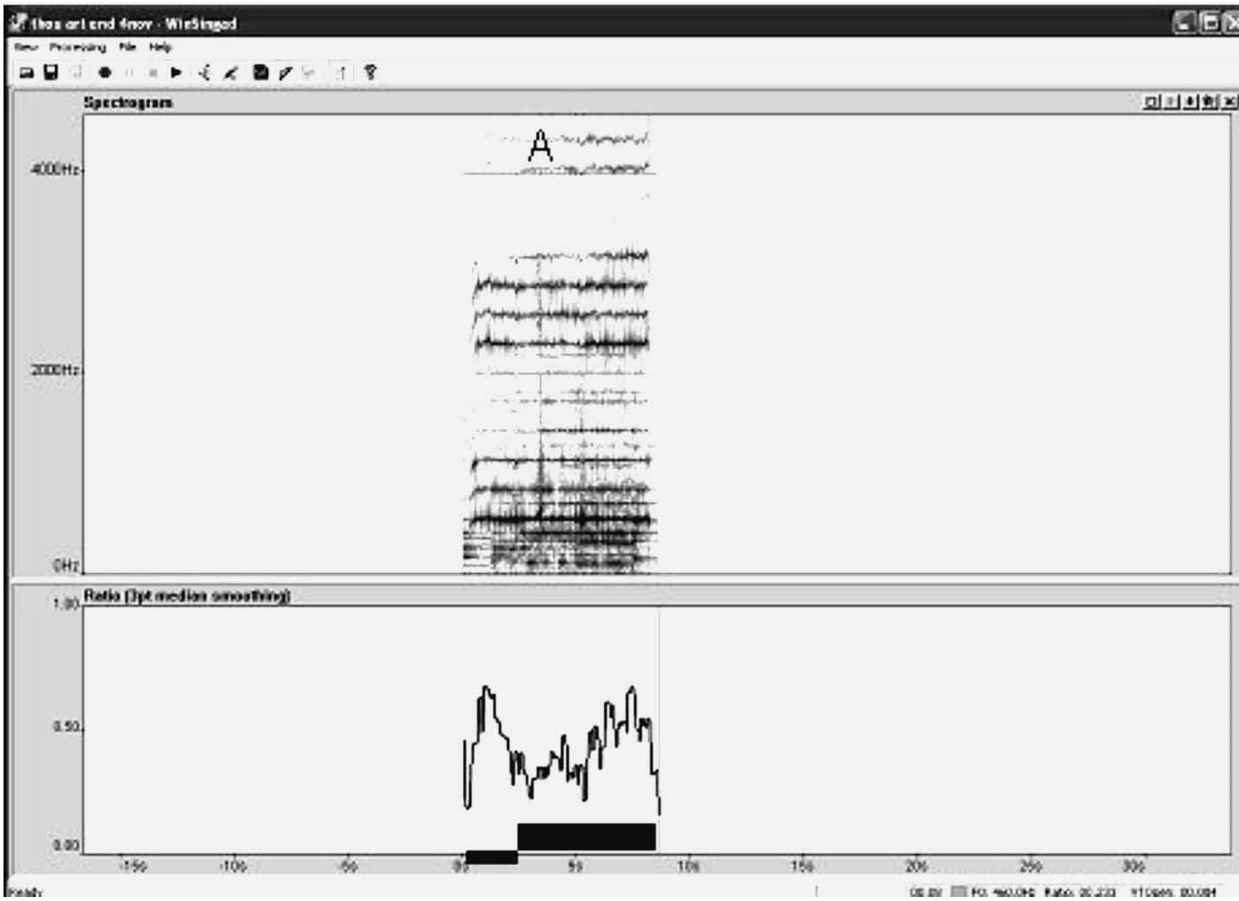

Fig. 6. Example of spectral ratio (lower) and spectrogram (upper) display panels against time for sung D4 (approximately 273 Hz) on the words 'thou art' from the end of 'Love, could I only tell thee', words by C Bingham and music by JM Capel.

both attend to the display and discuss what it reveals. It has been noted that there are times when students find that watching the display is rather too distracting when they are concentrating on performing a piece when their visual attention is better served elsewhere (to the score, teacher or accompanist). However, this has never been identified as an issue when rehearsing warming-up exercises, or practising short sections of a piece, as these are generally either known by heart or done by repetition.

Initial pilot results have highlighted examples where the software has been able to provide useful visual feedback in support of the immediate goals of the teaching process, generally in the context of developing resonance for projection and indicating where in the vocal tract its shape is changing. The intention is not for systems such as WinSingad to replace the teacher, but rather to provide a means whereby idiosyncrasies in individual teaching styles may be framed within a closer approximation to acoustic realities and in a common format that is accessible to both student and teacher. Teachers are always likely to be needed for the development of artistic performance. There are aspects of vocal training and education where the judgment of another human will always be required and where the use of a real-time display would not be appropriate, such as: stagecraft, performing musically, working with accompanists, working with conductors, working with directors, communicating with the audience, gesture, posture, ornamentation, etc. Nevertheless, initial results from this new practitioner-focused display suggest that singing teachers should be able to spend more lesson time for these essential and often somewhat neglected building blocks of performance.

ACKNOWLEDGEMENTS

This project is supported by the UK Arts and Humanities Research Board (AHRB) under innovation award number B/IA/AN8885/APN15651. The authors thank the students who took part in this phase of the study, the reviewers for their most helpful comments, as well as the teachers, students and other professional colleagues who contributed to the initial workshop.

REFERENCES

1. Moorcroft L. Embracing alternative methodologies: Science and imagery in the teaching and performance of singing. In: Stevens C, Burnham D, McPherson G, Schubert E, Renwick J, editors. Proceedings of the 7th International Conference on Music Perception and Cognition. Adelaide: Casual Publications; 2002. p. 561–654.
2. Persson RS. Concert musicians as teachers: On good intentions falling short. In: Cropley AJ, Dehn D, editors. Fostering the Growth of High Ability: European Perspectives. Norwood, NJ: Ablex; 1996. p. 303–20.
3. Welch GF, Sundberg J. Solo Voice. In: Parncutt R, McPherson GE, editors. The Science and Psychology of Music Performance: Creative Strategies for Teaching and Learning. New York: Oxford University Press; 2002. p. 253–68.
4. Callaghan J. Voice Science and Singing Teaching in Australia. Unpublished PhD thesis. University of Western Sydney, Nepean, Australia; 1997.
5. Howard DM. Variation of Electrolaryngographically derived closed quotient for trained and untrained adult female singers. J Voice 1995; 9: 163–72.
6. Rossiter D, Howard DM. ALBERT: A real-time visual feedback computer tool for professional vocal development. J Voice 1996; 10: 321–36.
7. Sundberg J. The science of the singing voice. Dekalb, Illinois: Northern Illinois University Press; 1987.
8. Williamon A, editor. Musical Excellence: Strategies and Techniques to Enhance Performance. Oxford: Oxford University Press; 2004.
9. Rossiter DP, Howard DM, Comins R. Objective measurement of voice source and acoustic output change with a short period of vocal tuition. Voice 1995; 4: 16–31.
10. Thorpe CW, Callaghan J, van Doorn J. Visual feedback of acoustic voice features for the teaching of singing. Australian Voice 1999; 5: 32–9.
11. Welch GF, Howard DM, Rush C. Real-time visual feedback in the development of vocal pitch accuracy in singing. Psychology of Music 1989; 17: 146–57.
12. Welch GF. Poor pitch singing: a review of the literature. Psychology of Music 1979; 7: 50–8.
13. Welch GF. A schema theory of how children learn to sing in tune. Psychology of Music 1985; 13: 3–18.
14. Rossiter DP, Howard DM, Downes M. A real-time LPC-based vocal tract area display for voice development. J Voice 1995; 8: 314–9.
15. Howard DM, Angus JAS. A comparison between singing pitching strategies of 8 to 11 year olds and trained adult singers. Logoped Phoniatr Vocol 1997; 22: 169–76.
16. Welch GF, Howard DM. Gendered voice in the Cathedral choir. Psychology of Music 2002; 30: 102–20.
17. Rossiter DP, Howard DM, De Costa M. Voice development under training with and without the influence of real-time visually presented biofeedback. J Acoust Soc Am 1996; 99: 3253–6.
18. Howard DM, Welch GF. Visual displays for the assessment of vocal pitch matching development. Applied Acoustics 1993; 39: 235–52.
19. Hess W. Pitch determination of speech signals: Algorithms and devices. Berlin: Springer; 1983.
20. Gruenz OO, Schott LO. Extraction and portrayal of pitch of speech sounds. J Acoust Soc Am 1949; 21: 487–95.
21. Howard DM, Fourcin AJ. Instantaneous voice period measurement for cochlear stimulation. Electronics Letters 1983; 19: 776–9.
22. Baken RJ. Clinical measurement of speech and voice. London: Taylor and Francis; 1987.
23. Lindsey GA, Howard DM. Spectral features of renowned tenors in CD recordings. Proceedings of Speech Research-89, 1989, Budapest, 17–20.