MANUSCRIPT SUBMITTED FOR MAVEBA 2003

TOWARDS A NOVEL REAL-TIME VISUAL DISPLAY FOR SINGING TRAINING

David M Howard[1], Graham F Welch[2], Jude Brereton[1], and Evangelos Himonides[2]

[1]Media Engineering Research Group, Department of Electronics, University of York, Heslington, York YO10 5DD, UK
[2]School of Arts and Humanities, Institute of Education, University of London, 20 Bedford Way, London WC1H 0AL, UK

Abstract: **Real-time visual displays have found application to be tested as part of a recently funded pilot project to investigate the usefulness or otherwise of computer displays in the singing studio. Following previous work that suggests that simple displays of a small number of analysis parameters are generally the most effective, the system makes available analyses plotted against time that relate to: pitch, spectral ratio, larynx closed quotient and vocal tract area. These can be viewed singly, multiply or in combination. The algorithms used will be described as well as previous analysis experiments that indicate their potential usefulness. A number of example output screens will be illustrated to indicate how users interact with the system. The on-going testing paradigm will also be described which is designed to establish whether or not displays such as these can be used in the singing studio to any useful advantage.**

.

Keywords : **visual displays, singing, vocal tract display**

I. INTRODUCTION

This paper describes the technology to be emplyed in a project during which the application of real-time visual feedback technology in the singing studio will be investigated, both during lessons and outside during private practice. In general, science and artistic musical performance tend to use different language codes and symbolisation for knowledge, and often, their ontological standpoints are different. Whilst it is not known to what extent these two language codes are reconcilable, the benefits from the application of technology have been demonstrated in many other fields, including the arts. There is no longer a widespread culture of technology phobia in non-scientific fields of human endeavour.

The standard pedagogical model employed in the conservatoire studio typically involves weekly/twice weekly lessons with an expert, supported by private practice and performance. The teacher is engaged in a psychological translation of the student's performance, for example by turning musical gestures into language, and the student is engaged in a further translation of the teacher's verbal and visual feedback into adapted singing performance. A dual possibility thereby exists for the misinterpretation of information. Anything that can provide more robust and easily understandable feedback to both teacher and student would seem to be worthwhile, and this forms the basic premise to investigate the use of technology in the signing studio.

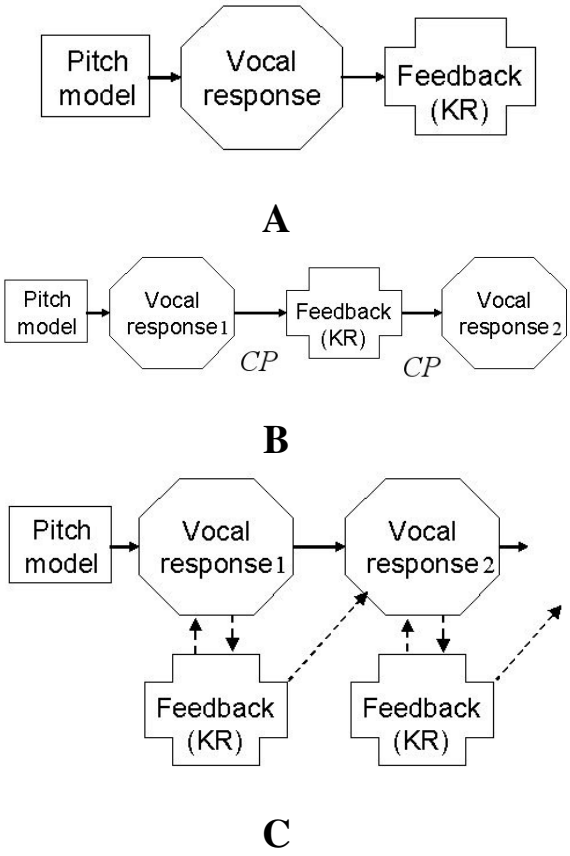

Figure 1: An illustration of the learning process for pitch in singing based on [1, 2]. Time is from left to right in these plots.
KEY: (A) the basic interaction between teacher and learner; (B) the on-going traditional learning process, and (C) the way in which real-time visual feedback can impact the learning process. KR = knowledge of results from an external source; CP = critical learning period.

Welch [1, 2] develops a model to characterise the learning process, taking pitch as an example, and this is illustrated in Fig. 1. During the traditional interaction between teacher and student, a model is provided, the

student makes an attempt vocally, and the teacher provides feedback to the student. A key issue in relation to this feedback is the gain the student makes in regard to knowing what s/he is supposed to be achieving in terms of a result, an external assessment being referred to as "knowledge of results" or "KR" as indicated in Fig. 1-A. Understanding what is required and how to recognise it is a vital aspect of the learning process.

Following feedback on a vocal response, the student subsequently will make another attempt as illustrated in Fig. 1-B. This is the nature of the traditional singing pedagogical process. The use of real-time visual feedback enables feedback to be provided *during* the student's vocal response, enabling modifications to be made immediately and their concurrent effect to be observed (see Fig. 1-C). 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 based on observations of the feedback provided during the previous attempt as appropriate.

Quantifiable parameters have been identified that vary with training and experience for: (a) actors [3], (b) adult singers [4, 5], as well as (c) girl and boy cathedral choristers [6]. Real-time visual feedback has been previously used successfully with primary school children [7, 8] and adult singers [9, 10]. 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 can then underpin feedback to provide more accurate formative and summative assessments.

II. DISPLAYS TO BE EMPLOYED

A. Consultation with the community

A one day workshop was held with a group of singing teachers, the authors, and interested colleagues who research in the areas of speech and/or singing. The purpose of this event was to review existing displays that might be useful in the context of the singing studio, and to produce a specification for the software to be employed in the project. Colleagues were reminded that the project is not about testing the effectiveness of the technology itself, but to establish its potential usefulness or otherwise. Specific research questions include:

- the extent to which teachers and students will accept and make use of technology in the studio
- the ease-of-use of the technology, both in the studio and elsewhere for private practice
- the nature of the data offered by the technology
- how the data can be integrated into singing teaching and learning
- the readiness with which the data can be interpreted and utilised
- whether the technology overly intrudes into the learning and teaching experience
- any potential perceived threat posed to the teacher and/or the student by the use of technology.

In order to make the technology be potentially widely applicable, a windows-based PC implementation was targetted. Existing possibilities for real-time displays were demonstrated, and the following were identified as being appropriate as tools for use in the singing studio for this project:

- fundamental frequency against time
- spectral ratio against time
- vocal tract area
- summary vocal tract area measures against time
- side view camera.

Each of these is described and illustrated below.

B. Fundamental frequency against time

The measurement of fundamental frequency (f0) has been the subject of considerable research [e.g. 10]. No one technique exists that is accurate for all subjects, covering the complete human pitch range uttered in any acoustic. The choice of a technique should be matched to the situation where it is to be used. A real-time display must not exhibit any delay to the user, it should be accurate operating over a wide f0 range for singers, of the order of C2(65Hz) to C6(1047Hz). A peak-picking system was employed that was originally developed in analogue form for use in cochlear implants [12], and subsequently applied in the SINGAD system [7, 8]. Each of the elements of its circuit has been implemented in C++, and an example plot of f0 against time is shown in Fig. 2.

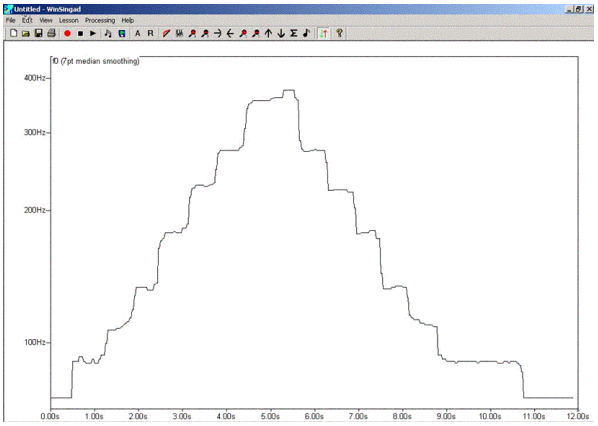

Figure 2: A display of fundamental frequency against time for a sung ascending and descending two octave arpeggio.

C Spectral ratio against time

A key element in singing training is that of voice projection, and one acoustic consequence of this is the appearance of a peak in the output spectrum in the region 2.5kHz to 4kHz, known as the singer's formant [e.g. 13]. The ratio of the energy in this band to the energy in the total signal is calculated. This measurement is constrained between 0 and 1 providing the full band extremes encompass the singer's formant band. In this implementation, these are set to (100Hz to 4000Hz) and (2500Hz to 4000Hz) respectively. These values can be changed by the user. Fig. 3 shows an example plot of this ratio against time for the vowel /a:/ sung in a projected and non-projected style.

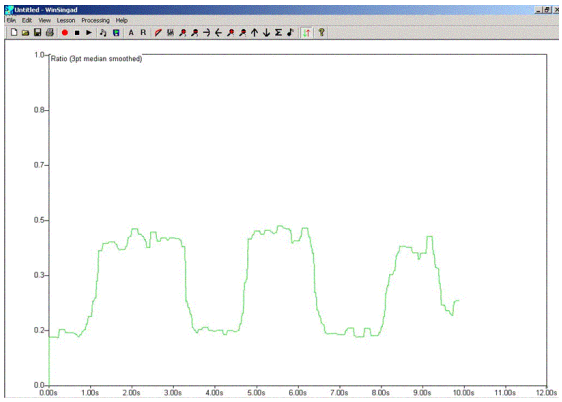

Figure 3: Example ratio against time display for/a:/ sung alternating between a non-projected.(lower ratio values) and a projected style (higher ratio values).

D 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 the areas (or diameters/radii) of a set of equal length tubes between the glottis (space between the vocal folds) and the lips. Fig. 4 shows an example vocal tract area display for a sung /a:/ vowel, where the glottis and lips are at the left and right edges of the display respectively.

There are, however, limitations associated with this representation. Firstly, it strictly only models non-nasal voiced sounds, due to the assumptions employed in linear prediction. Secondly, the output area values have no absolute area reference, and therefore they are arbitrary. They are usually therefore normalized either to a fixed glottis width (this is adopted in Fig. 4), 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 process, it is not obvious how it might be constrained, for example, to vocal tract configurations that are physically possible.

It is for this reason that summary plots of the average, minimum or maximum vocal tract area against time will be incorporated.

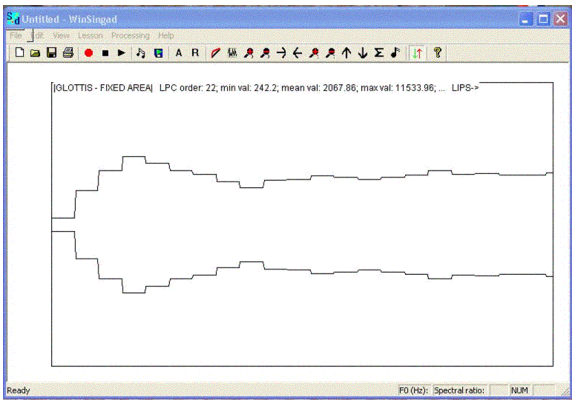

Figure 4: Example vocal tract area display for a sung /a:/ vowel . The glottis and the lips are at the left and right hand side of the plot respectively.

E Summary vocal tract area against time

The mean, minimum and maximum vocal tract area is calculated for each frame of input data, and these can be plotted against time. A plot of the mean area against time is shown in Fig. 5. An important aspect of singing training relates to the degree of perceived openness of the vocal tract, or the degree of constriction, and it is suggested that some indication of this may be given through reference to minimum vocal tract area against time.

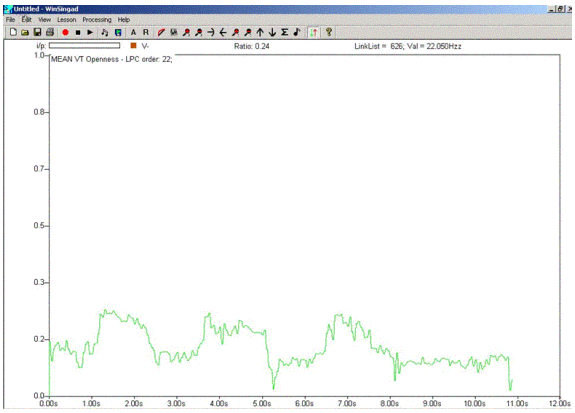

Figure 5: Example display of mean vocal tract area against time for /a:/ sung alternating between a non-projected.(lower values) and a projected style (higher values).

F Side view camera

Singers often make use of a mirror during training for feedback on their posture. With a computer display, it is

possible to make use of a camera with the result displayed on screen. We shall employ a camera to enable singers to view their posture from the side to enable the straightness of their spine to be observed. The screen will be placed at head height to encourage a vertical head position.

III. Discussion and conclusions

A set of displays to be employed in real-time in singing studios has been described. These are being integrated by a professional programmer into a complete system with the side view camera output, in which the user is given control over which single or arbitrary set of displays s/he wishes to use. Appropriate control over processing and display parameters will be provided to the user via standard menus and dialog boxes. In this way, attention can be drawn to individual parameters displayed alone, or to multiple parameters as familiarity and confidence grows, and areas of interest can be zoomed in on as desired. This system will provide the computer-based display system to enable the usefulness or otherwise of technology in the singing studio to be assessed.

An action research methodology is to be employed for this assessment, in which the teachers, students and the research assistants, acting as observers, keep diaries of progress and activities during lessons. Two teachers will be involved, each with an experimental and control group with two students in each.

The system will also allow both the audio signal (microphone) and video signal (side-view camera) to be recorded to enable vocal responses to be reviewed and/or archived for progress tracking.

IV. ACKNOWLEDGEMENTS

This project is supported by the Arts and Humanities Research Board (AHRB) in the UK under an innovation award numbered B/IA/AN8885/APN15651. The authors thank the singing teachers and other professional colleagues who contributed to the initial workshop.

References

[1] Welch, G.F. (1985a). A schema theory of how children learn to sing in tune. Psychology of Music, **13**, (1), 3-18.

[2] Welch, G.F. (1985b). Variability of practice and knowledge of results as factors in learning to sing in tune. Bulletin of the Council for Research in Music Education, **85**, 238-247

[3] Rossiter, D.P., Howard, D.M., and Comins, R. (1995). Objective measurement of voice source and acoustic output change with a short period of vocal tuition, *Voice*, **4**, (1), 16-31.

[4] Rossiter, D.P., and Howard, D.M. (1998). Observed change in mean speaking voice fundamental frequency of two subjects undergoing voice training, *Logopedics Phoniatrics Vocology*, **22**, (4), 187-189

[5] Howard, D.M. (1995). Variation of electrolaryngographically derived closed quotient for trained and untrained adult singers, *Journal of Voice*, **9**, (2), 163-172.

[6] Welch, G.F., and Howard, D.M. (2002). Gendered voice in the Cathedral choir, *Psychology of Music*, **30**, (1), 102-120

[7] Howard, D.M., and Welch, G.F. (1993). Visual displays for the assessment of vocal pitch matching development, *Applied Acoustics*, **39**, (3), 235-252.

[8] Welch, G.F., Howard, D.M., and Rush, C. (1989). Real-time visual feedback in the development of vocal pitch accuracy in singing, *Psychology of Music*, **17**, 146-157.

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

[10] Thorpe, C.W., Callghan, J., and van Doorn, J. (1999). Visual feedback of acoustic voice features for the teaching of singing, *Australian Voice*, 5, 32-39.

[11]. Hess, W. (1983). *Pitch determination of speech signals*, Berlin: Springer Verlag.

[12] Howard, D.M. (1989). Peak-picking fundamental period estimation for hearing prostheses, *Journal of the Acoustical Society of America*, **86**, 902-910.

[13] Sundberg, J. (1987). *The science of the singing voice*, Dekalb: Northern Illinois University Press.

[14] Rossiter, D.P., Howard, D.M., and Downes, M. (1995). A real-time LPC-based vocal tract area display for voice development, *Journal of Voice*, **8**, 4, 314-319.