Education and Culture
Lifelong Learning Programme

evta
european voice teachers association

EUROPEAN VOCAL PEDAGOGY – DIGITAL RESOURCES TECHNOLOGY
sponsored by the Leondaro da Vinci Lifelong Learning Programme

DIGITAL TECHNOLOGY WORKSHOP
MONDAY, AUGUST 29, 2011

The world of the software developer – Using real-time displays intelligently

Professor David M Howard, Audio Laboratory, Department of Electronics, University of York, UK

The world of singing teaching makes use of a number of techniques to encourage student singers into the use of a proper and healthy technique in the context of whatever genre they wish to pursue. Increasingly, young singers are asking why they are not making use of technology such as computers, tablets and/or smart phones as part of their singing training. This short article explores the possibilities for using such technology in the context of singing training and the pros and cons of so doing.

Background

Traditionally, singing teaching is essentially a qualitative process using techniques handed down from one generation of teachers to the next. The student singer is relying on the teacher's ears to ensure that the student develops an appropriate sound; a process that is commonly supported through the use of imagery (Moorcroft, 2002) and what I call 'psychological hooks' or concepts designed to promote the use of postural gestures that are deemed to be appropriate the production of a sung output such as: *sing on the point of the yawn*, or *sing as if there is an orange stuck in the throat*, or *sing through an imaginary hole in the forehead*. Psychological hooks clearly do work for many students, enabling them to sing with a vocal output appropriate to their singing needs. However, these hooks usually neither describe the physical reality of the voice production process nor the nature of the acoustic output being achieved.

New generations of singing students express a desire to make use of their computer, tablet or smart phone technology in the context of singing training. Previous research has demonstrated that the use of such technology can have a significant benefit for student and teacher alike by enhancing the nature of the feedback that the student receives. The qualitative feedback provided by the teacher is reinforced by the quantitative feedback offered by the real-time visual displays (e.g. Howard and Rossiter, 1992; Rossiter and Howard, 1994; Rossiter et al., 1996; Garner and Howard, 1999; Thorpe et al., 1999). These studies show that added benefits enable the student to: (a) monitor directly their singing during lessons with their teacher; (b) quantify aspects of their progress lesson by lesson, and (c) enhance their practice time between singing lessons. It must, however, be stressed that such systems will never, in my opinion, replace singing teachers for two key reasons.

1. Often an identified change can be made to a displayed parameter by more than one means. For example, an increase in larynx closed quotient is associated with voice training (Howard, 1995), but it can also be increased through the use of a pressed phonation (Sundberg, 1987), which is a completely unsuitable and potentially damaging voice quality with which to sing.
2. Some aspects of the art of singing require the judgment of another human; any computer-based system cannot and will never be a substitute such as: stagecraft, performing musically, working with accompanists, working with conductors, working with directors, communicating with the audience, gesture, posture, ornamentation, etc.

One benefit here of using a real-time visual display would be that teachers would be able to spend more lesson time on these essential and often somewhat neglected musical aspects of performance.

Modern computers are now capable of carrying out audio processing in real-time. Powerful voice analysis techniques, which once were only found for example in specialist speech science laboratories, can be implemented in real-time on standard office or home PC machines. Real-time visual displays have demonstrated their usefulness for developing particular vocal skills in the past, but they have commonly involved the use of external hardware interfaces to process the input data. Modern PC machines are now capable of carrying out the processing of the input data and it is therefore becoming possible to dispense with external hardware.

Real-time displays

In order to provide useful information about the singing training process using technology, it is important that the technology does not overly intrude into the pedagogical process and that it is entirely under the teacher's control. A teacher needs to be able to set the scene for the activities and any technology should be used in the same way that a mirror might be employed; bring it out and use it when the situation demands and the process can be enhanced. So what can such displays offer?

Figure 1: A real-time waveform display from the author's "iSingad" iPhone/iPad App (WWW-2) for a sung 'ah' vowel.

The acoustic signal that reaches our ears when someone is speaking or singing is usually available in a real-time display such as that shown in figure 1. This particular plot is for an 'ah' vowel which is pitched and can therefore be sung. Notice that the display shows a repeating variation and this is indicative of the fact it is a pitched sound. If the pitch drops there will be less cycles per second and if the pitch rises the number of repeating cycles per second will increase.

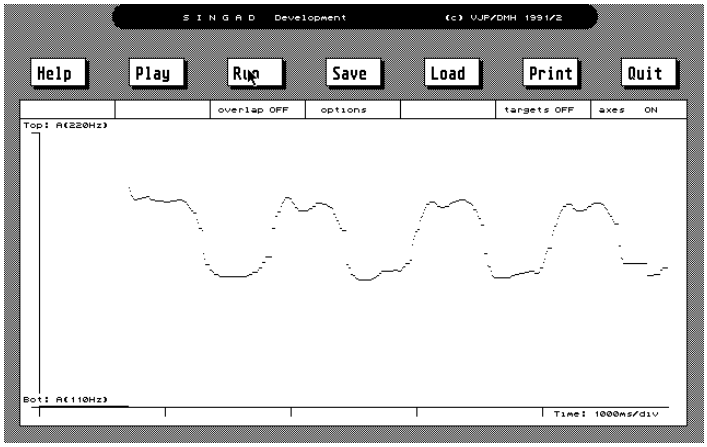

Figure 2: An early 1990s 'Singad' real-time display on an Atari computer for a child imitating a fire engine.

Pitch is a critical aspect of singing, and in figure 2 is plotted the pitch change for a child imitating the sound of a fire engine, an exercise that was used to encourage the use of 'up' and 'down' as words to describe pitch variations. It is worth a moment to think about this – on the keyboard pitch goes left and right but as musicians we talk about pitch going up and down. Pitch displays come in a number of formats including piano keyboards, note names with an indication as to how flat or sharp the sung note is in cents (hundredths of a semitone) with respect to that note tuned in equal temperament (or perhaps another temperament for subtle aspects of intonation control in for example, ensemble singing) or as a time plot of fundamental frequency as in figure 2.

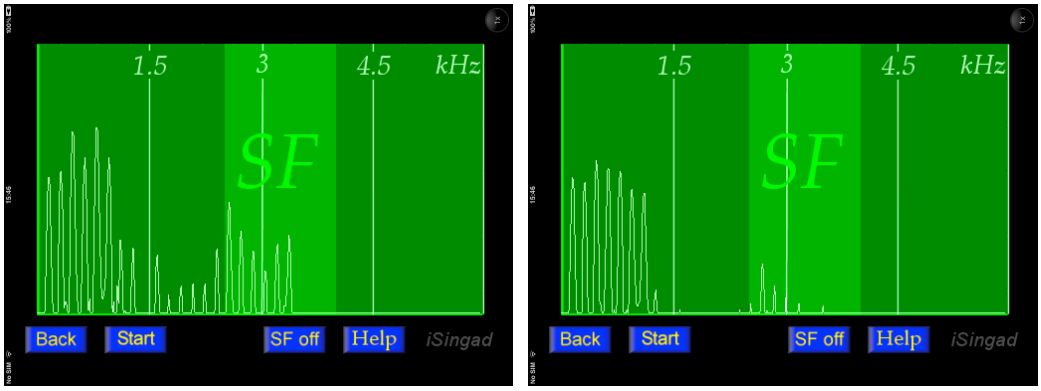

Figure 3: A real-time spectrum display from the author's "iSingad" iPhone/iPad App (WWW-2) for a sung 'ah' vowel in a projected (left) and unprojected (right) version to illustrate the presence and absence respectively of energy in the singer's formant cluster region denoted on-screen as the shaded region labelled 'SF'.

The output spectrum is a very useful display as it provides a direct link with the information provided to the brain by each ear. In figure 3 two spectra are shown, one for a projected and the other for a non-projected sung 'ah' vowels. The change in energy in the singer's formant cluster region (Sundberg, 1977) can be clearly seen. In this particular display the frequency and amplitude axes can be zoomed using a 'pinch' finger gesture. This display can also be used to show the differences between vowels in terms of their frequency spectra when their formant can be explored.

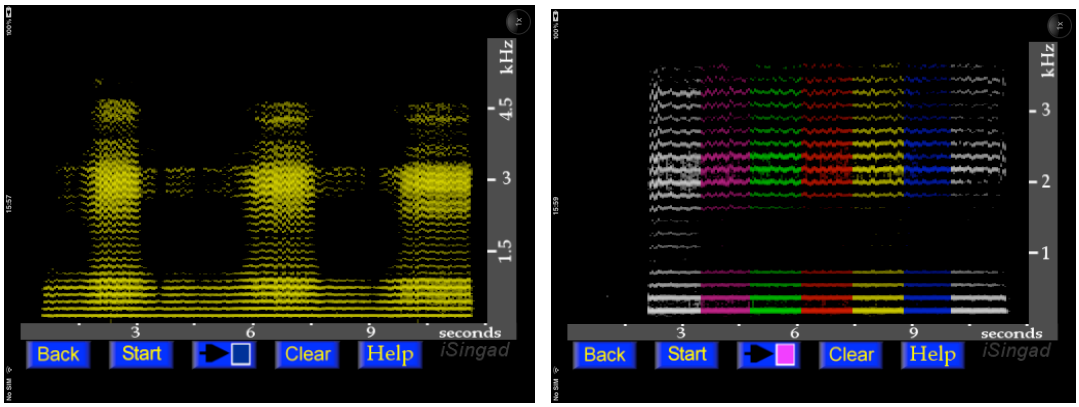

Figure 4: Real-time spectrogram displays from the author's "iSingad" iPhone/iPad App (WWW-2) for a sung 'ah' vowel for three projected and unprojected versions to illustrate the presence and absence respectively of energy in the singer's formant cluster region denoted on-screen as the shaded region labelled 'SF' (left) and to illustrate the author's view of a colour spectrogram (right).

When one wants to look at the changes in spectrum with time a spectrogram is commonly employed, Figure 4 shows a spectrogram for three projected and non-rojected sung 'ah' vowel pairs. Here the onset and offset details can be explored. Notice in these examples that the projected versions do not start and end with all harmonics appearing/disappearing in the singer's formant cluster region at exactly the same time. The right hand spectrogram illustrates what, for the author, is a colour spectrogram. Commonly, colour spectrograms make use of what is known as the 'red-hot poker' colour map which indicates relative levels in the signal based on how the

colours of a poker change in a furnace from black (cold) to white (hot). When a spectrogram is plotted in colour using such a colour mapping, different areas in the spectrogram appear in different colours and the eye sees a definite boundary between areas where the colour changes, say from orange to yellow. But such a colour change does not indicate anything important about the underlying signal; rather it is just where the level has changed from one dB value to the next which is completely arbitrary in terms of the input itself. Indeed, it would vary if the overall input level was changed, for example by moving closer or further from the microphone. So a set of colour spectrograms are illustrated in the right hand panel of figure 4 where the shading is used to indicate level variation.

Figure 5: Real-time display of dynamic variation in musical terms (ppp, pp, p mp, mf, f, f, ff, fff) from the author's "iSingad" iPhone/iPad App (WWW-2).

The overall dynamic level can be displayed and is likely to find particular application in the control of dynamic level across the musical dynamic range (ppp, pp, p, mp, mf, f, ff, fff) as shown in figure 5. In addition, this display can offer a sound level meter function which enables the local level to be checked in terms of a health and safety check of th elevels in the rehearsal or performance space in relation to the prevailing local noise exposure regulations. This is an oft forgotten aspect of singing where there is always the possibility of hearing damage in spaces that are particularly reverberant.

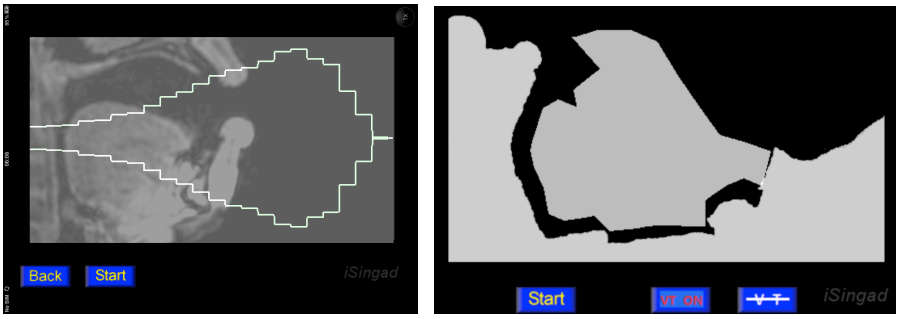

Figure 6: Real-time display of vocal tract area from the author's "iSingad" iPhone/iPad App (WWW-2). On the left is a representation of the tube area from the glottis to the lips for an 'ah' vowel, and on the right is a quasi-magnetic resonance image display of the vocal tract diameter from the glottis to the lips for an 'ah' vowel.

A very important aspect of singing training relates to vocal tract shape variations that are used to achieve for example, projection, increase in overall level and extremes of pitch range. These vocal tract area displays offer a way of seeing into the mouth whilst a sound is being sung; both displays work in real-time to provide their output pictures. They provide the underlying vocal tract area from glottis to lips. Current research on articulatory speech synthesis (e.g. Speed et al., 2009) makes use of similar data, usually gleaned from magnetic resonance images of a vocal tract, but there is no reason why in the future a synthesis option shoul dnot be provided so that the sound that the analysed tract shape would produce can be listened to.

Caveats and practical advice

Engineers create solutions to real-world problems and the implementation of real-time visual displays for singing trainging is no exception. However it should be born in mind that as an end-user, your expectations of a program might not be the same as those of the person who implemented it in practice. Always think about what results you would expect from an analysis system and make or generate (for example from an appropriate App for a smart phone such as the authors' "HarmSyn" for iPad, or "8ve Oscillator" for iPhone/iPad (WWW-1)) a sound for which the expected nature of the display is known (e.g. Howard and Angus, 2009). A sinewave at a known frequency is a good test input for example with which the frequency axes of a display can be checked. Just because a display system produces an output does not mean that it is doing 'what is says on the tin'!

Another point relates to real-time operation; these displays are usually called 'real-time' visual displays. In practice it is not possible to calculate parameters to display in zero time so there is actually no such thing as a 'real-time' display. In the context of voice analysis though, this is rarely now an issue provided we are not perceptually aware of a delay when we are observing a display of a sound we are listening to.

It is important to ensure that a clean version of the input is recorded and this means making use of a microphone that is placed reasonably close to the singer's lips (around 10-30 cm) off axis at about 45 degrees to prevent 'popping' sounds often associated with plosives such as 'p', 't' and 'k'. The distance is important to ensure that the signal picked up is predominantly that from the singer and not that from the room itself in terms of other sounds ('noise') or the muddying effects of the local room acoustic. The other key point is to ensure that the recording level is set such that it does not reach a maximum where it would clip the input or be so low that no useable signal is recorded. Ensuring the signal recorded is predominantly the signal of interest is otherwise described as avoiding 'garbage in = garbage out'.

Conclusions

Real-time displays are now becoming available for the support of the singing training process and this is to be welcomed and used where appropriate. It is important to check that the results observed are as expected. No computer program is free from bugs and the program itself is only as good as the abilities and knowledge of the programmer her/himself. Bear in mind that if the program is free to download and use then it may not be being maintained properly and bugs are possibly not going to be fixed; there is no such thing as a free lunch. Ultimately, real-time visual displays need to be used intelligently – think about the output in terms of whether it makes sense in the context of what you are observing and listening to.

References

Garner, P.E., and Howard, D.M. (1999). Real-time display of voice source characteristics, Logopedics Phoniatrics Vocology, 24, 19-25.

Howard, D.M. (1995). Variation of Electrolaryngographically derived closed quotient for trained and untrained adult female singers, Journal of Voice, 9, (2), 163-172.

Howard, D.M. and Rossiter, D. (1992). Results from a pilot longitudinal study of electrolaryngographically derived closed quotient for adult male singers in training, Proceedings of the Institute of Acoustics, 14, 529-536.

Howard, D.M. and Angus, J.A.S. (2009). Acoustics and psychoacoustics, 4th Ed., Oxford: Focal Press.

Moorcroft, L. (2002). Embracing alternative methodologies: Science and imagery in the teaching and performance of singing, In: *Proceedings of the 7th International Conference on Music Perception and Cognition*, Stevens, C., Burnham, D., McPherson, G., Schubert, E., and Renwick, J. (Eds.), Adelaide: Casual Publications, 561-654.

Rossiter, D., and Howard, D.M. (1994). ALBERT: A system for interactive analysis and display of voice source and acoustic parameters, Proceedings of the Institute of Acoustics, 16, (5), 301-308.

Rossiter, D.P.,Howard, D.M. and De Costa, 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.

Speed, M., Murphy, D.T., and Howard, D.M. (2009). Acoustic coupling in multi-dimensional finite difference schemes for physically modelled voice synthesis, *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, **WASPAA-09**, New Paltz, NY, 18-21 October.

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

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.

WWW-1: http://www.davidmhoward.com/iPhoneApps.htm (last accessed 4th September 2012).

WWW-2: http://itunes.apple.com/us/app/isingad/id545041820?mt=8 (last accessed 4th September 2012).