APPLICATIONS OF VISUAL FEEDBACK TECHNOLOGY IN THE SINGING STUDIO

Jean Callaghan[1], William Thorpe[2] and Jan van Doorn[2]
[1] University of Western Sydney   [2] University of Sydney

INTRODUCTION

This paper reports aspects of a continuing investigation into the use of computer-assisted visual feedback in the teaching of singing. The project is concerned with refining existing computer technology designed to provide visual feedback on acoustic parameters of the speaking voice and investigating how such feedback can be most effectively utilised in the singing studio.

Before embarking on a full-scale study, it was important to gain some understanding of the user's perspective. In a preliminary study conducted in 1999 teachers used commercially available speech technology in a singing lesson, incorporating the computer-assisted visual feedback in whatever way they found useful. Interview data from teachers and students indicated that it is certainly both feasible and productive to utilise computer technology in singing training (Callaghan, Thorpe & van Doorn, 1999). That study also clarified areas in which further investigation is needed and aspects of the speech technology that need modification for application to singing.

In 2000, the research was extended to further examine how computer technology can assist in the teaching of singing.[1] A specialised computer system that displays acoustic characteristics of a student's voice during singing was tested over a series of lessons in singing teachers' studios. Two types of results were obtained: those relating to the quality of the acoustic feedback and the computer display, and those relating to pedagogical uses.

The results indicated that such technology can provide useful assistance to the teacher and student, but the teacher needs to be able to interpret what the computer is showing and incorporate that understanding into the learning environment. We are now proceeding with a more extended study to develop new visual feedback technology for use in singing teaching, through investigation of:

1. acoustic analysis techniques for extracting perceptually relevant characteristics from the singing voice;
2. methods of presenting acoustic information in meaningful visual displays; and
3. pedagogical approaches that integrate this technology into the practice of singing teaching.[2]

[1] Supported by an Australian Research Council Small Grant
[2] Supported by an Australian Research Council Large Grant

ACOUSTIC ANALYSIS OF VOICE

Pitch and Vowels

The acoustic energy in voiced sounds such as vowels is generated by vibration of the vocal folds. The frequency at which the folds vibrate determines the pitch of the resulting sound, and the manner in which the folds vibrate (e.g. pressed, breathy) largely determines the acoustic power that is produced and the frequencies across which it is distributed. The acoustic energy is then modified by passage through the vocal tract so that the character of the final output sound is affected by a combination of laryngeal and vocal tract effects (see Figure 1).

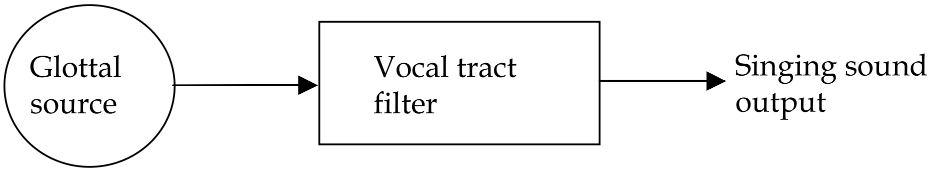


Figure 1: Simplified model showing the mechanisms involved in production of voiced sound.

Vowel production involves the shape of the lips, the opening between the jaws, the position of the soft palate and the shape of the tongue. Linguistic classification of vowels, however, has usually been done by reference only to the position of the main body of the tongue in the oral cavity—high-low and front-back (see, for example, O'Connor, 1973; Denes & Pinson, 1993). In the traditional vowel quadrilateral (Figure 2), /i/ is at the high front corner (i.e. the tongue is high towards the front of the mouth), /u/ at the high back corner (i.e. the tongue is high towards the back of the mouth) and / / at the low back corner (i.e. the tongue is low at the back of the mouth). Other vowels, such as / / (HERD), are classified as central. Vowels may also be classified as "closed" (the tongue near the palate) or "open" (the tongue low, at the bottom of the mouth.

The position of the tongue in the oral cavity in effect produces two acoustic chambers through which the sound must pass. Each chamber can be considered as an acoustic resonator, amplifying the acoustic energy near to its resonant frequency, and reducing the amount of energy at frequencies far from the resonance. So a simplified view of the vocal tract filter is to consider two resonances, called *formants* in acoustic phonetics (Figure 3).

The frequencies of the formants F1 and F2 can be identified from spoken or sung vowels with the appropriate type of analysis, and graphed on an *acoustic vowel chart* (Figure 4) similar to the articulatory vowel chart. Conventionally, F1 is represented on the vertical axis and F2 on the horizontal axis, with the direction of increasing frequency reversed to emphasise the relationship to the articulation.

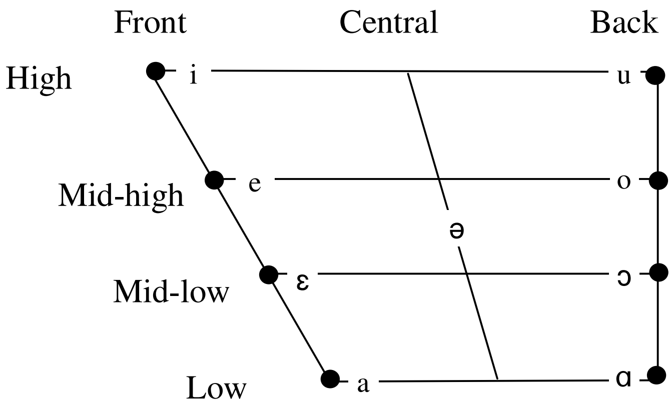


Figure 2: Articulatory Vowel Chart showing positions of tongue and jaw for different vowels.

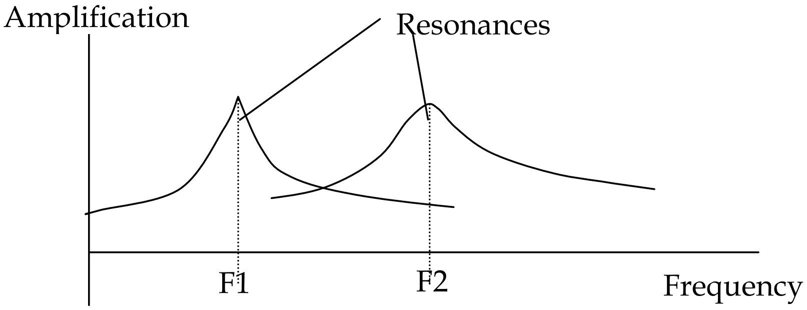


Figure 3: Formants resulting from resonances in the vocal tract.

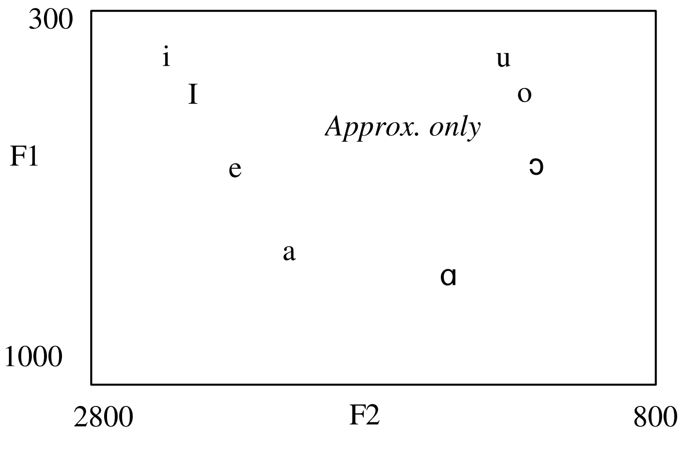


Figure 4: Acoustic Vowel Chart showing frequencies of first two formants for English vowels, indicating similarity to articulatory vowel space in Fig 2.

Speech and Singing

Perceptual characteristics of the singing voice include some that are common to speech applications, such as pitch, loudness, and vowel identity—with corresponding acoustic correlates fundamental frequency (Fo), sound pressure level, and formant frequencies (particularly the first two, F1 and F2). In singing, however, pitch accuracy is much more critical than in speech, and attaining the correct vowel pronunciation is an important part of learning because of the necessity to sing in different languages and also because vowel pronunciation affects vocal colour.

The singing voice also includes a number of other relevant characteristics with perceptual descriptors such as vibrato, “resonance”, and “colour” (Wapnick & Ekholm, 1997). Titze, 1997 asked “Are the corner vowels like primary colors?”, suggesting that, since these vowels “set the limits of vowel spaces in many (if not all) languages” (p. 35), they are like the primary colours, red, yellow and blue, from which all other colours are derived. Vowels are linked with colour in another way, too. Essentially, it is the “colour” of the vowel that determines the vocal colour, and accounts for the traditional association of particular vowels with particular emotions (as discussed, for example, in Vennard, 1967, and Manén, 1987).

Ekholm et al. (1998) compared perceptual ratings with a number of acoustic analyses performed on recordings of the singing voice. Their results suggest that the perceptual attribute of “resonance” is well -correlated to the acoustic power in and around the "singer’s formant”, a strong resonance between 2.5 and 3.5 kHz that appears to be caused by a clustering of the third, fourth, and fifth formants (Sundberg, 1994). There was also some relationship between the perceptual attribute of “colour” and the measurements of vibrato rate, extent, and onset delay at the start of a note. Other acoustic measures that appear to represent perceptual characteristics include the overall spectral slope (Bloothooft & Plomp, 1988) and the levels of individual harmonics within single critical bands (Sundberg, 1994).

THE AVAILABLE TECHNOLOGY

The visual feedback currently available is able, essentially, to give feedback on pitch, vowel quality, and resonance. Five displays are available: pitch trace, immediate pitch display, spectrographic display, spectrum, and vowel chart.

1. Pitch trace (Figure 5)

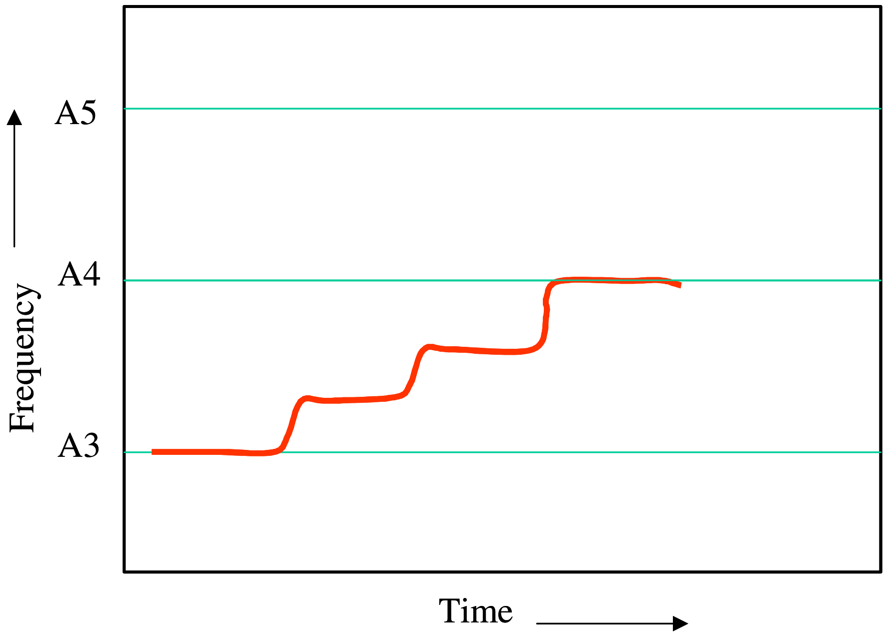
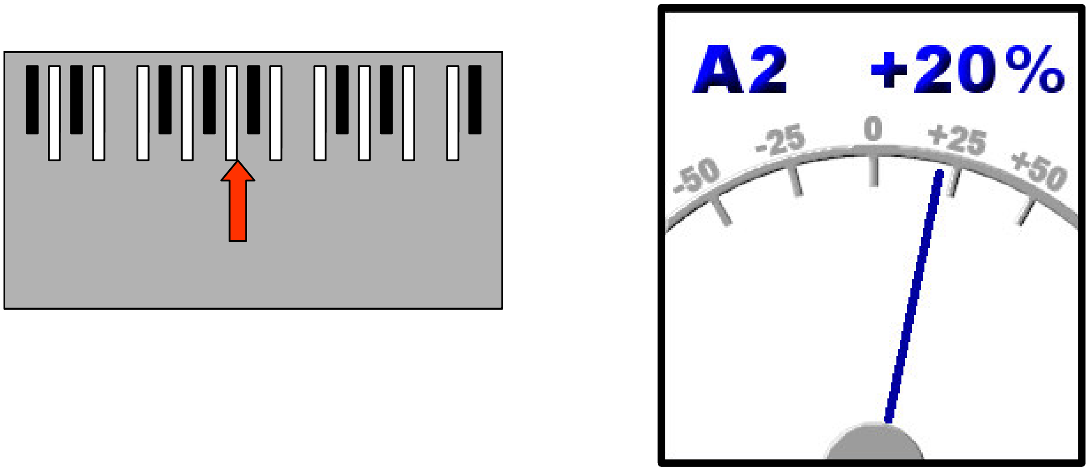


Figure 5: Pitch Trace, in which the temporal pattern of pitch is displayed by means of a “moving” line drawn across the screen. Reference lines indicate fixed pitch targets.

This shows the temporal pattern of pitch, typically by means of a line that scrolls across the screen. Time is therefore represented along the horizontal axis, and the pitch at any instant by the height of the line at that time. Typically, the display shows the pitch of the current phonation as the head of the line, with the recent history scrolling away to the left. The singer therefore has feedback of both the immediate phonation as well as its context. The size of a pitch change corresponding to a particular interval is therefore very apparent. However, due to the resolution of the display screen, it can be difficult to represent the exact pitch, apart from at a small number of reference frequencies, so it may not be possible to show if the interval was accurate.

2. Immediate pitch display (Figure 6)

Another way to display the pitch is to show just the immediate value, without previous history. This simplifies the display because the temporal dimension is removed, thus allowing an improvement in the frequency resolution. The two ways in which the pitch can be portrayed are as a point on a scale (e.g. as represented by a keyboard) or by showing the error relating to a specific target pitch (e.g. as in a musical tuning device). The error feedback could be associated with an auditory stimulus that the student is asked to match.



(a) **(b)**

Figure 6: Immediate pitch display showing **(a)** sung pitch with reference to a keyboard representation of the musical scale, and **(b)** the error between the sung and target pitch.

3. Spectrographic display (Figure 7)

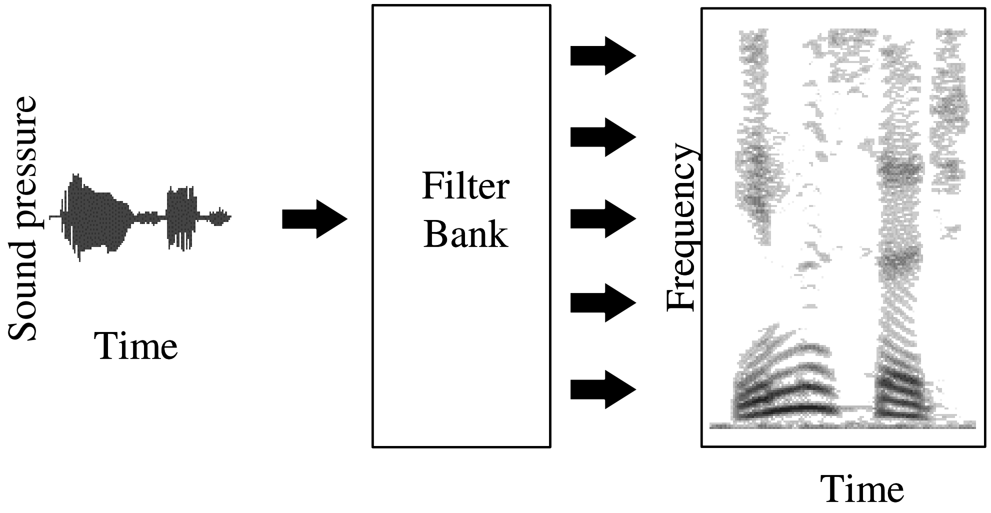


Figure 7: Spectrographic analysis of a time-domain waveform into a two-dimensional image representing the temporal evolution of the voice frequency spectrum.

The spectrogram of a voice signal portrays the time-varying nature of the spectrum of frequencies present in the sound. Because of the highly structured nature of voice signals, this reveals some very distinctive patterns. In particular, it is possible to see the broad patterns made by different vowels and consonants, and also the individual harmonics and their variation with pitch and vibrato. Non-harmonic noise (such as in breathy or other distorted phonation, or associated with glottal onsets) can be observed. However, due to the large amount of information present, it can be difficult to usefully interpret this during real-time visual feedback. Some teachers who have spent time examining the different patterns have been able to make use of this display to explain aspects of phonation to their students, such as vibrato, onset, and intensity of the singer's formant.

4. Spectrum (Figure 8)

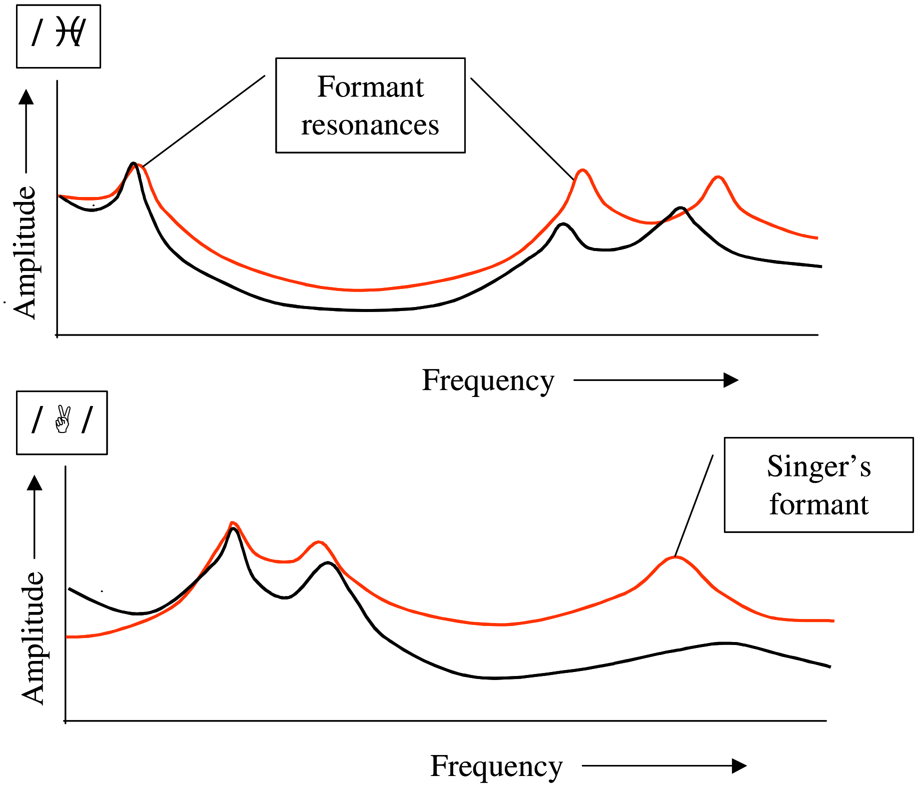
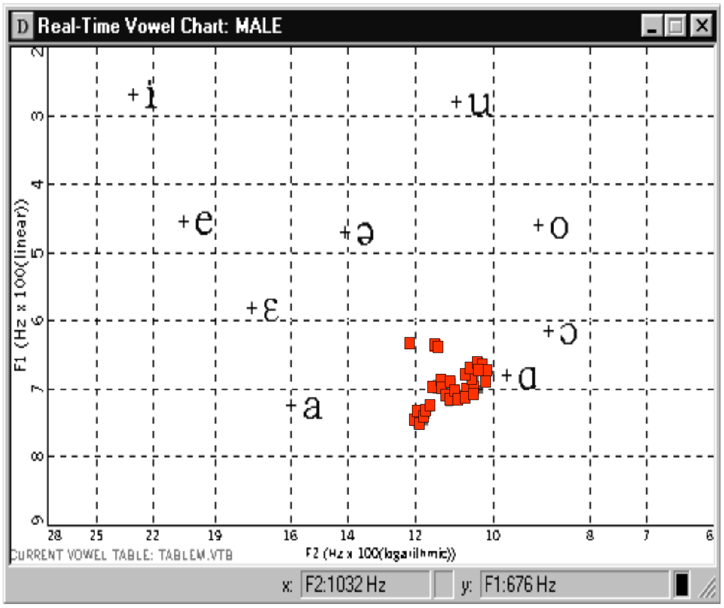


Figure 8: The instantaneous spectrum, appropriately smoothed to reduce the harmonic structure, indicates the pattern of resonances in the voice. These reflect the vowel identity, and characteristics such as the singer's formant.

By smoothing the spectrum and removing the temporal aspect, the overall spectral shape of the voice can be displayed in real-time—during sustained vocalisation. This allows one to focus on the frequencies of the formants and their relative intensities. The lower two formants are typically considered to convey the vowel identity, whereas the third and higher formants represent aspects of voice quality. It is possible to use this type of feedback as an aid in adjusting voice quality and vowel identity. For instance a target can be set and the student asked to change the voice quality towards a greater level of singer's formant, while maintaining the vowel formants at their appropriate locations.

5. Vowel chart (Figure 9)



CURRENT VOWEL TABLE: TABLEM.VTB

x: F2:1032 Hz y: F1:676 Hz

Figure 9: “Acoustic” Vowel Chart representing vowel identity with respect to frequencies of first two formant frequencies. The superimposed dots show the visual feedback provided during vocalisation.

Phoneticians (e.g. Fant, 1960) have determined that there is a correspondence between the articulatory vowel chart, defining the positions of the tongue and jaw for each vowel (see Figure 3), and the acoustic vowel chart that shows the frequencies of the first two formants (see Figure 2). This provides a straightforward way to represent the vowel identity from an acoustic analysis of the voice sound only. Again there is no temporal dimension, with the analysis result being immediately displayed as a dot on the screen, overlayed by a chart indicating the desired vowel target(s). The student can adjust his or her articulation and observe how the point moves about the chart, and thereby attempt to approach the target vowel. This type of feedback appears to be particularly useful in accent modification tasks, because it bypasses the auditory perceptual categories that can limit our ability to perceive small differences in vowel quality. However, there are some major difficulties with this type of feedback at high pitch, because the wider harmonic spacing as pitch increases means that the accuracy with which formant frequencies can be resolved become much less.

Requirements for Refining the Visual Feedback

In considering the usefulness of existing speech technology, it is apparent that singing, which in general has a much greater range of pitch and intensity variation, can place some extreme demands on algorithms originally developed and optimised for speech.

For visual feedback to be useful for learning, it must be both accurate and relevant to the required task (Welch, 1985). Accuracy implies that the analysis algorithms are correct and relatively insensitive to noise such as extraneous acoustic inputs. In particular, the visual representation of some feature of the acoustic signal should change little for similarly small changes in the acoustic signal. Relevancy implies that prominent features of the visual display represent perceptually important aspects of the acoustic signal (in terms of the particular learning task) and that when a student approaches a desired acoustic result, the visual representation also converges towards the desired target. If feedback is inconsistent or incorrect, learning may actually be degraded (Buekers et al., 1994).

USING THE TECHNOLOGY IN TEACHING

It is important for the teacher to be aware that essentially what we see on the screen is a map —a visual representation of some aspects of the vocal sound. In order to guide the student in the right direction, the teacher needs to be able to interpret the map and relate its features to physical and auditory landmarks. The teacher needs to know how vowels are produced and what effect physical manoeuvres have on the acoustic output.

Pitch

Pitch in singing may be regulated by laryngeal control or breath management, mechanisms which are often interdependent.

Vowel Quality and Resonance

In singing, vocal resonance and word articulation are interdependent parameters reliant on the movements of the articulators (pharynx, jaw, soft palate, tongue, lips). While vowel identity is largely determined by the first two formants, the colour of the voice is also influenced by the overall spectral distribution of the acoustic energy. This is influenced by the position of all the formants, and also by the characteristics of the glottal vibration (Sundberg, 1987).

The first formant is particularly affected by the mandible (jaw), the second formant by the tongue shape, and the third formant by the position of the tip of the tongue or, when the tongue is retracted, to the size of the cavity between the lower incisors and the tongue. Additionally, all formant frequencies decrease uniformly as the length of the vocal tract increases, which can be accomplished by lowering the larynx and/or increasing the degree of lip rounding (Pickett, 1980). Hence, the combination of larynx height adjustment and lip rounding or spreading provides an effective means of darkening or brightening the vowel colour.

Titze (1998) points out that widening the pharynx produces a darker, stronger sound quality. The first formant frequency is lowered and lower partials are emphasised by the vocal tract. However, for the vowel to be identified, the location and degree of constriction of the vocal tract characteristic of the vowel must remain relative to the mouth configuration.

The singer's formant, the high spectrum envelope peak in the vicinity of 3 kHz in all vowels, is the acoustic correlate of what we perceive as "ring" or "focus", cultivated to balance the lower harmonics and to assist in register-blending and legato line, and to achieve carrying power of the voice. Production of the singer's formant requires a wide pharynx and an aryepiglottic constriction.

VOCAL PEDAGOGY

Teaching musical performance skills is always a complex task requiring the translation of cognitive understandings and sound images into physical co-ordinations. This is a particular challenge in singing, where the musical instrument is the whole person. The singer may be thinking language and music while apprehending internal sensations of vibration, movement, and sound, and while attending and responding to external sensations such as the sound of the voice and the sight and sound of instrumental accompaniment, other singers, and an audience. The process of hearing, perceiving, and remembering sound forms a loop with the production of sound.

Teaching singing uses modelling, coaching, exploration, reflection and feedback. 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).

Learning to sing involves training in two fundamental elements: musical concepts and psychomotor skills. These are interdependent in that content (e.g. pitch) cannot be studied without applying some specific level of skill (e.g. the ability to co-ordinate the vocal mechanism to produce the requisite pitch).

The process of skill acquisition can be broken down into three stages: the cognitive stage, the associative stage, and the autonomous stage (Anderson, 1982). In both the cognitive and associative stage, modelling and external feedback have been identified as important. Most commonly, modelling is supplied by the teacher demonstrating the task and feedback by the teacher’s verbal comments.

In conventional singing training, interpretation of the teacher’s feedback can be problematic, partly because of the difficulties in verbally explaining perceptual and production aspects of the voice, but also because of the delay between when the student produces the vocalisation and when the feedback is made. This delay makes it difficult to learn the motor control programs because the feedback provided by the teacher is disassociated from the proprioceptive and auditory sensations accompanying vocalisation (Welch, 1985). Several studies in cognitive psychology and cross-modality perception have provided evidence that visual feedback can dramatically improve musicians’ learning of skills and may, in combination with verbal feedback, be more effective than verbal feedback alone (e.g. Marks, 1978; Walker, 1981; Welch et al., 1989; Butler & Winne, 1995).

Using acoustic analysis displayed on a computer screen can provide another kind of visual feedback which, in conjunction with the teacher's analytical ear, can be used as a tool in diagnosing problems and achieving correction of those problems. Through visual and auditory feedback, the singer's awareness of the quality of the sound can be “made real”, helping to establish the mental images which build to learning the task.

An advantage of using computer-assisted visual feedback is that students can judge their own progress. It is even possible—after the teacher has clearly defined the task—for them to use the equipment in the teacher's absence and still be confident about what they are achieving. In our preliminary study referred to above (Callaghan, Thorpe & van Doorn, 1999) one participant teacher felt that his students achieved more, and faster than usual, because they were able to respond to the visual feedback cues without any of the feeling of being judged that sometimes comes with verbal feedback from the teacher.

Expert teaching relies on two types of knowledge: content knowledge (knowledge of the subject matter to be taught) and pedagogical knowledge (knowledge of how to teach) (Shulman, 1987). In a qualitative research project evaluating the voice pedagogy of singing teachers in Australian tertiary institutions, Callaghan (1998) found that most singing pedagogy in Australian tertiary institutions is practised with incomplete content knowledge, i.e. knowledge of vocal physiology and acoustics. One way of incorporating knowledge of acoustics into a traditional master-apprentice approach is to supply computer-assisted visual feedback.

REFERENCES

Anderson, J.R. (1982). Acquisition of cognitive skill, *Psychological Review*, 89, pp. 369-406.

Buekers, M.J., Magill, R.A., & Sneyers, K.M. (1994). Resolving a conflict between sensory feedback and knowledge of results, while learning a motor skill, *Journal of Motor Behavior*, 26(1), pp. 27-35.

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

Callaghan, J. (1998). Singing teachers and voice science: An evaluation of voice teaching in Australian tertiary institutions, *Research Studies in Music Education*, 10, pp. 25-41.

Callaghan, J., Thorpe, W. & van Doorn, J. (1999). Computer-assisted visual feedback in the teaching of singing. In M.S. Barrett, G.E. McPherson & R. Smith (Eds), *Children and Music: Developmental Perspectives*, pp. 105-111. Proceedings of the 2nd Asia-Pacific Symposium on Music Education Research and the XXI Annual Conference of the Australian Association for Research in Music Education. Launceston: University of Tasmania.

Denes, P.B. & Pinson, E.N. (1993). *The speech chain: The physics and biology of spoken language* (2nd ed.). New York: Freeman.

Ekholm, E., Papagiannis, G.C. & Chagnon, F. (1998). Relating objective measurements to expert evaluation of voice quality in western classical singing: Critical perceptual parameters, *Journal of Voice*, 12, pp. 182-196.

Fant, G. (1960). *Acoustic thedory of speech production*. The Hague: Mouton.

Manen, L. (1987). *Bel canto: The teaching of the classical Italian song-schools, its decline and restoration*. Oxford: Oxford University Press.

Marks, L.E. (1978). *The unity of the senses*. New York: Academic Press.

O'Connor, J.D. (1973). *Phonetics.* Harmondsworth, Middlesex: Penguin.

Pickett, J.M. (1980). *The sounds of speech communication*. Baltimore, MD: University Park Press.

Shulman, L.S. (1987). Knnowledge and teaching: Foundations of the new reform, *Harvard Educational Review*, 19(1), pp. 4-14.

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

Sundberg, J. (1994), Perceptual aspects of singing, *Journal of Voice*, 8, pp. 106-122.

Titze, I. (1997). Voice research: Are the corner vowels like primary colors? *Journal of Singing*, 54(2), pp. 35-38.

Walker, R. (1981). The presence of internalised images of musical sounds, *Bulletin of the Council for Research in Music Education,* 66-67, pp. 107-112.

Wapnick, J. & Ekholm, E. (1997). Expert consensus in solo voice performance evaluation, *Journal of Voice*, 11(4), pp. 429-36.

Watson, P.J. & Hixon, T.J. (1985). Respiratory kinematics in classical (opera) singers. *Journal of Speech and Hearing Research*, 28, pp. 104-22.

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

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