A BIOFEEDBACK SYSTEM FOR SINGING TUITION OF CHILDREN AND ADOLESCENTS

Christopher Barlow
School of Computing and Communications
Southampton Solent University

Jude Brereton
AudioLab
Department of Electronics
University of York
UK

Abstract

Developments in audio technology mean voice biofeedback is possible using standard home computers. There is insufficient data on children's voices on which to base such a system. A number of tools were assessed to provide the specification of a biofeedback system for young singers. In order to underpin such a system, the authors recorded and analysed 418 young singers from classical and theatre music backgrounds. Quantifiable, statistically significant links were found between gender, development, voice training and vocal genre. Results suggest that the measurements used could form the basis of a reliable voice pedagogy system for young voices.

Background

The teaching of singing presents some difficult pedagogical issues for both teacher and learner. It involves a re-iterative process of performance, analysis and teacher's feedback (Welch et al., 2005). Singing teachers often use visual and mental imagery to communicate improvements, requiring the students to translate before they can modify performance behaviour, leading to considerable opportunities for misinterpretation and misunderstanding between teacher and learner.

The traditional singing studio is beginning to undergo a dramatic evolution as real-time computer-based visual feedback now offers a powerful tool to aid teacher/ student communication (Howard, 1990), particularly if the student can work with it alone after appropriate demonstration and input with the teacher. Increases in computer processing speeds and development of inexpensive, high-quality audio hardware means that audio analysis and feedback is possible using standard home computers. Howard et al. (2004) demonstrated that the development of adult singing skills is enhanced when quantitative feedback is provided, involving the direct monitoring of vocal output during lessons. However, a lack of research on children's vocal development means that use of such systems with young people is problematic as there is no corpora of data on which to base such a system (Welch, 2002).

A child's vocal system undergoes rapid and dramatic change during adolescence, during which many singers are making demands on their voice as great as that of any adult. During the pubertal growth spurt, both male and female larynxes change dimensions rapidly, necessitating a constant reassertion of the muscle control skills needed for speech and singing (Pedersen, 1997). There are many conflicting ideas on how child singers should be treated during this period and very few of these concepts are supported by systematic research (Barlow, 2003). Although several computer based tools are available to provide feedback on the voice, for either voice pedagogy, or medical assessment, there is relatively little research using these tools which either attempts to quantify the young voice or to improve vocal pedagogy for young people.

Voice Analysis

The standard model of voice production is based on the *source-filter* theory (Fant, 1970), in which the *sound source* (vocal folds) produces a complex periodic waveform. The resonant properties of the vocal tract *(filter)* will cause certain frequencies within the source waveform to be amplified, causing resonant peaks (formants) in the vocal tract response (Howard, 2006), and changing the acoustic properties of the sound. The vocal tract can be manipulated to vary the position and strength of the formants (Howard, 2002). The relationship between the first two formants (F1 and F2) is a principle mechanism for differentiating vowels in speech (Welch et al., 2005), while the formants F3 to F5 having significant effects on vocal colour/tone and voice projection.

Voice analysis tools can be loosely divided into two key groups — those which analyse the acoustic output of the voice, and those which examine the voice source (Barlow, 2003).

Acoustic analysis. Acoustic analysis falls into two main categories: *time domain* (waveform) and *frequency domain* (spectrum). The time domain representation of a waveform displays amplitude against time of an analogue of the original pressure wave. Analysis capability using this representation is limited to fundamental frequency (pitch) and amplitude (loudness) against time (Kent & Read, 2002).

Frequency domain techniques are based on Fourier analysis of the waveform. Computer based real-time spectral analysis systems use a small section of the waveform (a *frame*), of usually around 20–30ms and divides it into discrete frequency components using a Fast Fourier Transform (FFT) (Kent & Read, 2002). The output is a frequency spectrum of the individual frame on which the

FFT is based, and as such contains no time information. Short-term and long-term average spectra (LTAS) can be generated by averaging the spectral values across a number of frames.

The *Spectrogram* uses a 3-dimensional time/intensity/frequency plot, in which time appears on the horizontal axis, frequency on the vertical axis and intensity is represented by darkness of colour on the plot (Howard, 2002). This enables the user to monitor *variation* of aspects of the acoustic output such as tone, formants and vibrato depth and rate. A difficulty with all acoustic forms of analysis, however, is their susceptibility to error caused by the acoustic of the room in which the recording is made — room resonances and external noise can be picked up my the microphone and will be represented on the acoustic analysis, potentially leading to erroneous assumptions (Barlow, 2003).

Voice Source Analysis

A principal concern in voice analysis is the sound source itself — the larynx. If this can be monitored while in operation, it is possible to derive a considerable amount of information about the production of voiced sound. The waveform denoting the vocal fold function (i.e., the rate of the opening and closing of the vocal folds) is called the "glottogram" (Hess & Indefry, 1984) and can be derived from a variety of means.

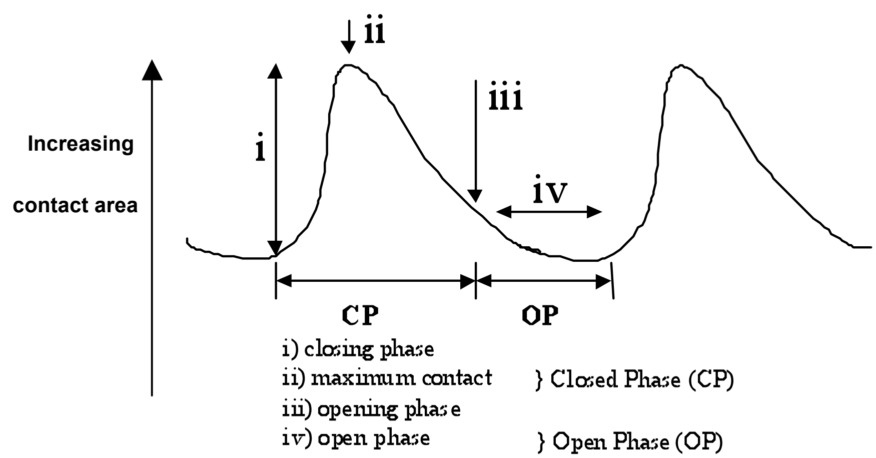
Direct analysis. Medical clinicians have several techniques at their disposal for analysing the laryngeal system. The primary concerns of these clinicians are usually tissue pathology or neurological dysfunction, and the equipment that they use is primarily based on direct visual or photographic observation (*laryngoscopy*) (Titze, 1994). Such analysis relies on invasive procedures which are of little use in pedagogy. Rather than examining pathological defects in the larynx, the voice coach is concerned with optimising its use as a musical instrument. Invasive equipment that prevents normal phonation removes any advantages that can be gained by having scientific feedback.

Indirect analysis. There are, however, a number of non-invasive procedures available to consider: *Inverse filtering* attempts to calculate the acoustic modifying function of the vocal tract and construct a filter that is the inverse (Chasaide & Gobl, 1997). The acoustic effects of the vocal tract are theoretically obviated, and therefore the original glottal waveform is restored. There is currently no available software which performs inverse filtering of the microphone input — speech or singing — in real time.

Human tissue is a fairly good conductor of electricity, and e*lectrolaryngography* uses a constant voltage high frequency electric current passed between two electrodes placed externally either side of the neck at the level of the larynx.

When the vocal folds are apart, the current flows along a longer path, increasing impedance of the signal (Garner & Howard, 1999). Current variation can therefore be analysed in terms of changes in vocal fold contact area (Figure 1). As the signal is resilient to the acoustic properties of the room, it is widely used for F0 estimation in laboratories and clinics (Barlow et al., 2002).

Figure 1: Idealised electrolaryngograph output



(Howard et al., 1990, p.206)

Voiced sounds are generated by the oscillation of the vocal folds as air passes through the glottis (the opening between the folds). The percentage of the cycle for which the vocal folds are in contact is known as the “larynx closed quotient” or CQ. This measure can be used to quantify the function of the vocal folds, and has been related to training (Howard, 1995), genre (Evans & Howard, 1993), development (Barlow, 2003) and vocal efficiency (Howard, 1990).

Method

To build a dataset to underpin development of a biofeedback system for use with young singers, the authors have made a number of recordings of young singers and non-singers from a range of disciplines and geographical areas. These include ‘professional’ choristers from seven UK cathedral choirs, pupils from one specialist music school, pupils from two specialist stage schools, as well as four conventional schools. A total of 418 students have been recorded speaking and singing.

Students were grouped, for the purposes of classification, into three categories of training, development, and gender as outlined in Table 1 below. All subjects were aged between 8 and 18 years at the time of recording.

Table 1: Categorisation of subjects

	Category/Criteria
Training duration	**1. Trained:** >2years with professional teacher or director (after Barlow and Howard, 2002) **2. Untrained:** <3 months singing lessons (ibid)
Voice development	**1. Unchanged:** Age up to 12.9 years (after Pedersen, 1990; Harries, 1997) **2. Changing:** Age 13.0 to 15.9 years (ibid) **3. Changed:** Age 16.0 years and above (ibid)
Gender	**1. Male** **2. Female**

Materials

Subjects were recorded speaking and singing using a headset mounted omnidirectional reference microphone and a Laryngograph® which was used to record the laryngographic signal (Lx). The Lx signal was viewed on an oscilloscope during the recording to maintain correct electrode positioning. Recordings were made digitally in Wave format (.wav) using the Laryngograph system onto a PC laptop with 16 bit resolution and 22kHz sampling rate.

Procedure

Subjects were recorded using a standard protocol (Howard, 1990: Pedersen, 1997):

a) Reading aloud a passage of spoken text approximately 90s in duration to determine mean spoken F0.
b) Singing a two-octave, ascending and descending scale over the range of G major.
c) Singing a verse of *Happy Birthday* in the key of C major.
d) A small group of students who were studying classical and theatre singing were recorded singing a verse of *Happy Birthday* in C major in each of the two styles.

Analysis

1. Laryngographic data was analysed from all subjects to identify a possible relationship between the assessment parameters and measurable data from the analysis system. This included plotting CQ/Log F0 scatterplots and measuring mean CQ data across 3rd octave intervals of the scale.

2. Long Term Average Spectra (LTAS) were analysed for a sample of 40 female classical and theatre music voices singing *Happy Birthday* in the key of C major (Figure 2). SpectraLab© was used to generate LTAS using a Hanning window with an FFT size of 4096 samples.

3. For the same group of singers, Laryngographic data was used to calculate mean CQ values for each note circled in the phrase (Figure 2) to assess if differences in vocal production could be measured between the two genres.

Figure 2: *Happy Birthday* — syllables selected for analysis circled

Results

CQ/pitch changes with training. Analysis of CQ/log F0 scatterplots of young 'classical' singers demonstrated a developmental pattern. CQ value varied against pitch with increasing levels of training in 'classical' voice for unchanged male voices and for all female voices within the study (Table 2). This suggests that the relationship of CQ and F0 (pitch) can be an indicator of training level. Howard (1995) found a similar developmental relationship of CQ/F0 amongst adult female singers, although the pattern variation was different.

Table 2: Composite model of vocal groupings according to slope of Qx plot of trained and untrained male and female adolescent singers

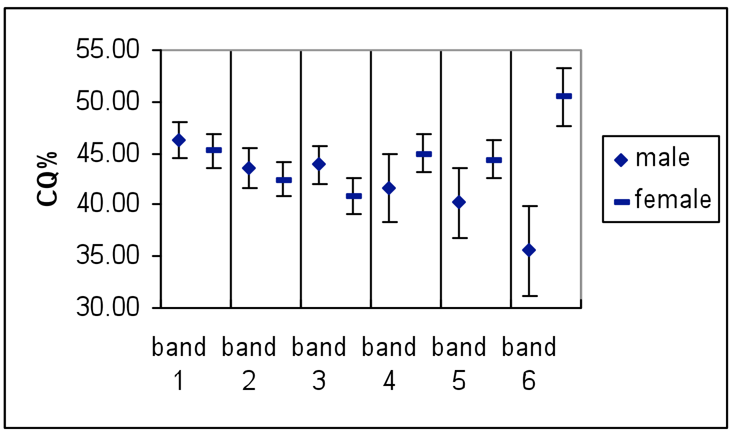
Classification	*Pattern*	*Idealised Plot*
Group 1	negative gradient slope throughout plot	CQ F0
Group 2	horizontal (zero gradient) plot.	
Group 3	varying gradients (positive-negative positive or negative-positive-negative).	
Group 4	initially negative then switching to positive gradient ('V' shape).	
Group 5	positive gradient throughout plot	

Increased training

Mean closed quotient. The scale was divided into 3rd octave bands (Barlow, 2003) and mean CQ was calculated across the sample group for each pitch band. A further statistical analysis of mean closed quotient values against sung pitch found significant ($p < 0.05$) relationships between the effects of gender, training and vocal development and closed quotient values across the two octave range.

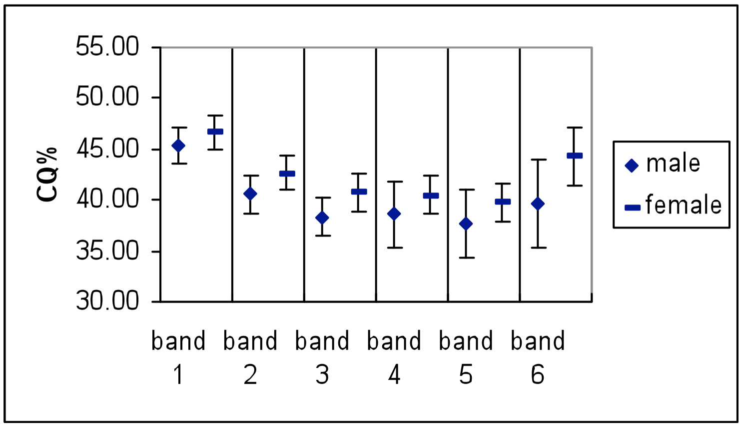
Significant ($p < 0.5$) differences were found between exhibited mean CQ values of male and female children with unchanged voices. In particular boys exhibited a decreasing mean CQ against pitch, while CQ in girls' voices decreased with pitch over the low octave and increased over the upper octave (Figure 3). It is suggested that given the minimal difference in physiological structure of the male and female prepubertal larynx, this difference is principally behavioural.

Figure 3: Mean CQ with standard error for unchanged male and female child untrained singers in third octave bands



This hypothesis is further supported by the analysis of CQ against training. Trained boys and girls, exhibited significantly similar patterns to each other (Figure 4), suggesting that training can overcome innate behaviours in children's vocal production.

Figure 4: Mean CQ with standard error for unchanged male and female trained singers in third octave bands



Larynx closed quotient for 'classical' vs 'theatre' styles. Mean CQ was calculated for each of the selected notes for both singing styles. Table 4 demonstrates that mean CQ and standard deviation for 4 of the 5 notes analysed is higher in 'mix', with the mean for C5 being nearly identical. Analysed as individual singers across all notes, 76% of mean CQs were higher in 'mix' voice than in 'head'. A one tailed student's T-test of the means demonstrated a significant difference between the two data sets ($p = 0.018$).

Table 4: Mean larynx closed quotient and standard deviation for each note

Voice		C4 / æ	E4 / u	G4 / u	Bb 4 / æ	C5 / ɝ
'classical'	mean CQ%	26.2	25.6	24.6	28.3	28.2
	Sd	6.2	6.2	5.9	4.8	5.9
'theatre'	mean CQ%	31.1	31.1	29.4	33.6	27.8
	Sd	2.9	5.4	4.4	3.6	4.8

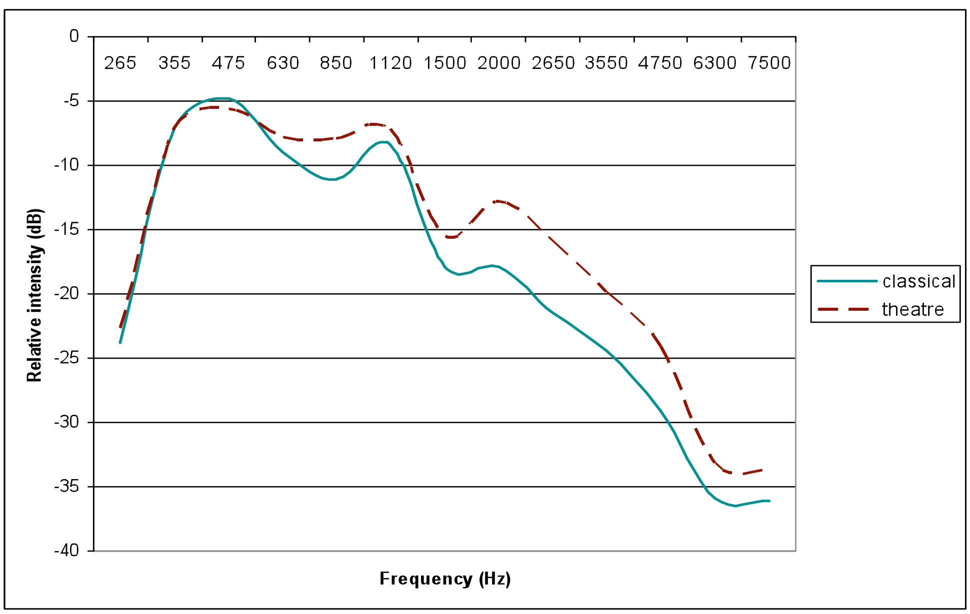
Long term Average Spectra of 'classical' and 'theatre' styles. LTAS were calculated for each of 40 students. Twenty classically trained students sang *Happy Birthday* in a 'classical' style and 20 theatre music trained students sang the same verse in a 'theatre' style. The verse was sung in the key of C major, between C4 (261 Hz) and C5 (522 Hz). Intensities were normalised to give equivalent intensity between genres over the bottom bands, and mean LTAS data was calculated for each group (Table 5). The LTAS curves are shown in Figure 5.

Table 5: LTAS for female adolescents in singing in 'classical' and 'theatre' styles

Frequency Boundary / Voice type	Classical	theatre
265	-23.7	-22.7
355	-7.0	-7.0
475	-4.8	-5.5
630	-8.9	-7.8
850	-11.1	-7.8
1120	-8.3	-7.1
1500	-17.8	-15.
2000	-17.9	-12.8
2650	-21.4	-15.9
3550	-24.4	-19.8
4750	-29.1	-24.1
6300	-35.9	-33.2
7500	-36.0	-33.7

Results demonstrated that the spectral slope is shallower for theatre voice than for classical, supporting the results taken from measurements of average vowel spectra, reported by Barlow et al. (2007).

Figure 5: LTAS of adolescent females singing in 'classical' and 'theatre' styles



This suggests a voice resonance strategy which differs between styles. The spectrum shows a considerably stronger intensity for all frequencies above 355 Hz, and in particular the spectral slope of the 'classical' voices is steeper in the lower frequencies.

Current Systems

There are a number of voice biofeedback software packages making claims that they can be used to 'teach singing' which have appeared on the commercial market in recent years, including: Singing Coach® and Singing Coach Kidz®, Sing and See®, Music Masterworks® and Vimas Singing Tutor®. The voice science community, in addition, use a variety of sophisticated tools to analyse the voice. The available outputs from a range of these tools are summarized in Table 6.

Table 6: Outputs from current voice science and 'singing teacher' software

Output / Software (Voice science / Singing teacher)	Real time	Waveform	Real time Spectrum	Average Spectrum	EGG	CQ/F0 plot	Spectrogram	Formant analysis	Voice Range Profile	Pitch trace	Piano keyboard	Sheet music view	MIDI song import	Interactive games	Statistics/post processing	Maximum recording	Vocal tract modelling	Cost (£GBP)
Speech studio	X	x	x		x	x	x	x	X	x					x	60 mins		>5K
WinSingad	X	x	x				x			x						3 mins	X	N/a
Praat		x	x				x	x	X	x					x	60 secs		Free
Voce Vista	X	x	x		x		x	x	?	x					x	8 secs		Free
CSL	X	x	x	x	x		x	x	X	x				x	x			>10K
Music Masterworks	X									x			x			N/a		~20
Singing Coach	X									x		x	x			N/a		~15
Vimas Singing Tutor	X									x		x	X			N/a		~10
Sing and See	X	x								x	x	x						~30
Sing and See Pro	X	x	x				x			x	x	x				N/a		~50

It is evident that the software marketed as educational tools for children focus on very different aspects than do the Voice science tools. All the software surveyed above includes some sort of pitch tracking and indeed it this tool alone which is common to all educational and voice science analysis programmes.

Some packages allow the display of real time spectral information which may then be linked into a graphical display of musical score, piano keyboard or visual targets. Some software also allows the importing of musical files in Musical Instrument Digital Interface (MIDI) format, to enable the singer to see how 'accurately' he or she has sung a particular song.

Such tools, whilst useful for teaching pitching accuracy, are rather limited in the analysis of the complex physiological and musical act of singing. For the teacher to fully integrate technological tools into their pedagogic practice it is necessary to incorporate more complex tools into a voice biofeedback system, including detailed reporting of data and analysis of more fundamental parameters other than pitch.

Research on an early vocal feedback tool for children, based on pitch tracking (Welch et al., 1989) suggested inherent problems in any educational package which concentrates on only one parameter. As the software concentrated on feedback of pitching accuracy, singers concentrated on getting 'accurate' pitch scores and ignored other essential aspects of voice production, resulting in high scores while producing a perceptually 'poor quality' singing voice. The advantage of these systems, however, lies in two things: an intuitive user interface and the ability to set easily definable targets for the pupil to work towards.

The scientific tools on the other hand, generally include statistical analysis, data logging and a variety of different displays, but suffer from two key disadvantages: some of the software is very expensive, while even the freeware tools tend to demand a high degree of technical competence by the user, and lack a user friendly interface which would make the learning experience intuitive and enjoyable. Some software is only able to provide feedback after post-processing. Real-time feedback is essential for a pupil to make the links between the action and the feedback response.

Discussion

There are a number of key aspects of singing performance which can be addressed by the teacher: pitching accuracy, tone, loudness, posture, vibrato depth/rate and breathing. The majority of these can be analysed either in real-time or by post-processing data by currently available technology. The data examined above demonstrates that it is indeed possible to generate significant measurable data by analysing the voice source and acoustic output of child singers. Furthermore, the use of a combination of acoustic analysis and voice source analysis can be used to give detailed breakdowns of the progression of young voices with training within a vocal genre, which can be used to differentiate between the effects of training, gender, development and musical genre for young singers. In addition, a combination of biofeedback tools and a suitable user interface could be used effectively when integrated into a pedagogic system. Due to the issues with current systems the majority of software currently available is less than ideal for private use by even adult students and is particularly unsuited to unsupervised use by children. However, a new system especially for child voices could encourage the student to use the software at home. Most usefully, the system would allow the teacher to manipulate different settings, and/or set "homework" for the student to work on before the next tuition session, to check progress and monitor the student's use of the system resulting in an improved learning cycle with formative feedback enabled even for private practice.

Requirements for a Biofeedback System

The key functions of a vocal feedback tool must be to enhance the learning experience, to be easy to use and to perform a useful musical function, which necessitates a number of key requirements for its specification:

- Assess a wide variety of performance parameters in order to allow maximum feedback on the singing voice and not cause pupils to focus on certain aspects of learning to the detriment of others.
- Allow real time analysis in order to give a direct link between action and feedback.
- Allow storage and post processing of data to assess performance development over time.
- Make use of 'simple' and 'pro' interfaces for use by student and teacher respectively.
- Incorporate user friendly interfaces which engage students and allow them to set targets and view development.

Acknowledgements

The authors would like to thank all the participants and organisations which made this possible, including: Abbey Gate College, Chester; Abbey Gate School, Chester; Blackburn Cathedral Choir; Bootham School, York; Brooklyn Youth Chorus Academy, USA; Chester Cathedral Choir; High Storrs School, Sheffield, Ripon Cathedral Choir; Salisbury Cathedral Choir, Sheffield Cathedra Choir, Stagecoach UK Ltd; Sylvia Young Theatre School; St Mary's Music School, Edinburgh; Wells Cathedral Choir; York Minster Choir.

This research project is supported by the Arts and Humanities Research Council: Grant number AH/E000721X/1.

References

Barlow, C., & Howard, D. M. (2002). Voice source changes in child and adolescent subjects undergoing singing training. *Logopedics Phoniatrics Vocology, 27*(2), 66–73.

Barlow, C., LoVetri, J., & Howard, D. M. (2007). Voice source and acoustic measures of girls singing "classical" and "contemporary commercial" styles. In A. Willamon, & D. Coimbra (Eds.), *Proceedings of the International Symposium on Performance Science, Casa da Musica, Porto, Portugal* (pp. 195–200).

Barlow, C. A. (2003). *Electrolaryngographically derived voice source changes of child and adolescent subjects undergoing singing training.* Unpublished doctoral dissertation, University of York, UK.

Brereton, J. (2000). A *voice source analysis of chest and head registers in singing*. Unpublished master's thesis, Trinity College, Dublin, Ireland.

Chasaide, A. N., & Gobl, C. (1997). Voice source variation. In *The handbook of phonetic science.* Oxford: Blackwell.

Evans, M., & Howard, D. M. (1993). Larynx closed quotient in female belt and opera qualities: A case study. *Journal of Voice, 2,* 7–14.

Fant, G. (1970). *The acoustic theory of speech production.* The Hague: Mouton.

Garner, P., & Howard, D. (1999). Real-time display of voice source characteristics. *Logopedics Phoniatrics Vocology, 24,* 19–25.

Hess, W., & Indefry, H. (1984). Accurate pitch determination of speech signals by means of a laryngograph. In *Proceedings, International Conference on Acoustics, Speech and Signal Processing 1984* (pp.1–4).

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

Howard, D. M. (2002). The real and non-real in speech measurements. *Medical Engineering and Physics*, *24*, 493–500.

Howard, D. M., Lindsey, G. A., & Allen, B. (1990). Towards the quantification of vocal efficiency. *Journal of Voice, 4,* 205–212.

Howard, D. M., Welch, G. F., Brereton, J., Himonides, E., DeCosta, M., Williams, J., & Howard, A. W. (2004). WinSingad: A real-time display for the singing studio. *Logopedics Phoniatrics Vocology, 29*(3), 135–144.

Kent & Read. (2002). *The acoustic analysis of speech* (2nd ed.). San Francisco: Singular.

Pedersen, M. (1997). *Biological development and the normal voice in puberty*. Unpublished doctoral dissertation, University of Oulu, Finland.

Titze, I. (1994). *The principles of voice production.* Denver: NCVS.

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

Welch, G. (2002). *Report from the ESF/SCSS Exploratory workshop on: Voice development, assessment, education and care in childhood and adolescence.* European Science Foundation.

Welch, G. F., Howard, D. M., Himonides, E., & Brereton, J. (2005). Real-time feedback in the singing studio: An innovatory action-research project using new voice technology. *Music Education Research*, *7*(2), 225–249.