Defence Science Journa 43, No 3, July 1993, pp

Real-time Visual Displays in Speech and Singing

David M. Howard
Signal Processing, Voice and Hearing Research Group
Dept of Electronics, University of York, Heslington, York, YO1 5DD, UK

ABSTRACT

Many branches of the speech, hearing and singing research community use analyses of aspects of the voice source and overall spectral changes. By these means, patterned variations have been noted in both the voice source and the acoustic output from the vocal tract during speech and singing, the speech/singing of professional voice users and pathological speech. The widespread availability of fast computers and dedicated signal processing devices has enabled many of these techniques to be made available as real-time displays to a wide range of users. This paper describes a number of real-time visual displays of voice source, the acoustic output and articulatory gestures during speech and singing. The development of analysis techniques based on a model derived from the results of contemporary psychoacoustic experiments into the nature of the human peripheral hearing mechanism is discussed. This model depends on parallelism which is ideally suited to implementation on transputers which are specifically designed to operate in parallel. Developments such as these may further our understanding of the relative importance of voice source and acoustic cues in speech and singing, whether normal, professional or pathological, and enable a new generation of real-time visual displays to be implemented.

.. INTRODUCTION

Real-time visual displays provide a basis for human learning. They have the particular advantage that an image of an attempt at a task is seen to evolve in real-time on the screen, and then it is captured for more detailed study after the event. Providing the visual display presents data which is meaningful to and properly understood by the user, it can be beneficial to the development of skills relating to the task in hand which can develop rapidly.

Speech is one of the most complex tasks that humans have to master. Many situations are far from ideal to enable normal speech development, and real-time visual feedback systems which are appropriate to the task in hand can provide a highly useful tool for speech training and speech rehabilitation. There are a number of aspects of singing which are poorly understood by singing teachers in terms of their acoustic, voice source and physiological effects. Real-time visual displays can provide the teacher and student with feedback on particular aspects of the developing voice which are known to exhibit changing trends during voice development. It does not matter whether the teacher and student understand the details of the analysis upon which the display is based, provided that the changes to be observed actually occur during work on particular aspects of vocal training, such as posture or breathing.

This paper describes a number of real-time visual displays designed for use in speech and singing development which include: fundamental frequency against time, larynx closed quotient against time, a frequency/time/amplitude analysis of the acoustic output and a tube model display of vocal tract articulation.

2. REAL-TIME FUNDAMENTAL FREQUENCY DISPLAYS

Fundamental frequency (f_0) analysis has been the subject of research since the 1920s, and many techniques

Received 2 February 1993

have been investigated. No technique exists which works reliably for any speaker/singer in any acoustic environment, and the successful design and implementation of an f_0 analyser depends on a proper understanding of the limitations of any proposed technique with respect to the intended application itself. Hess gives a very comprehensive review of f0 analysis techniques[1].

2.1 Fundamental Frequency Display based on Electrolaryngography

A technique which is commonly used to give a reference f_0 measure is the electrolaryngography[2], otherwise known as electroglottography[3]. This makes use of a measure of the electrical impedance between two electrodes placed superficially on the neck at larynx level, one electrode behind each vocal fold. A constant voltage, high frequency signal is applied between the electrodes and changes in electrical conductivity due to the vibration of the vocal folds are monitored. During phonation, there is a change in the contact area between the vocal folds and hence a change in inter-electrode current flow. Figure 1 shows an example of few cycles from the electrolaryngograph output waveform[4], or Lx*.

The Lx waveshape remains essentially constant during voiced speech and singing, and is thus a relatively simple matter to derive a measure of f_0. This is generally achieved by finding the time between the maximum positive peaks in the time differentiated Lx waveform (these points are marked with a δ in Fig. 1) and taking the reciprocal.

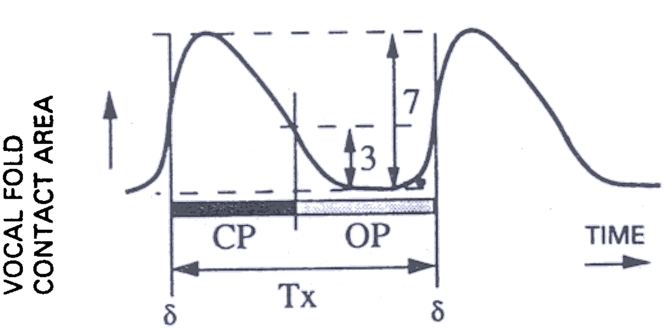


Figure 1. A few cycles of the electrolaryngograph output waveform, showing the measures made to calculate the period (Tx), closed phase (CP), and the open phase (OP).

*The output from the electrolaryngograph is usually plotted with increasing current plotted as a positive change. It should be noted that the electroglottograph output is generally plotted as the inverse.

Figure 2 shows the output from a real-time display of f_0 based on the Lx waveform against time for 'He wouldn't say yes and he wouldn't say no either' uttered by an adult male (Fig. 2(a)) and an adult female (Fig. 2(b)). Similarities can be observed in the general shape of the pitch patterns, which one would expect as both

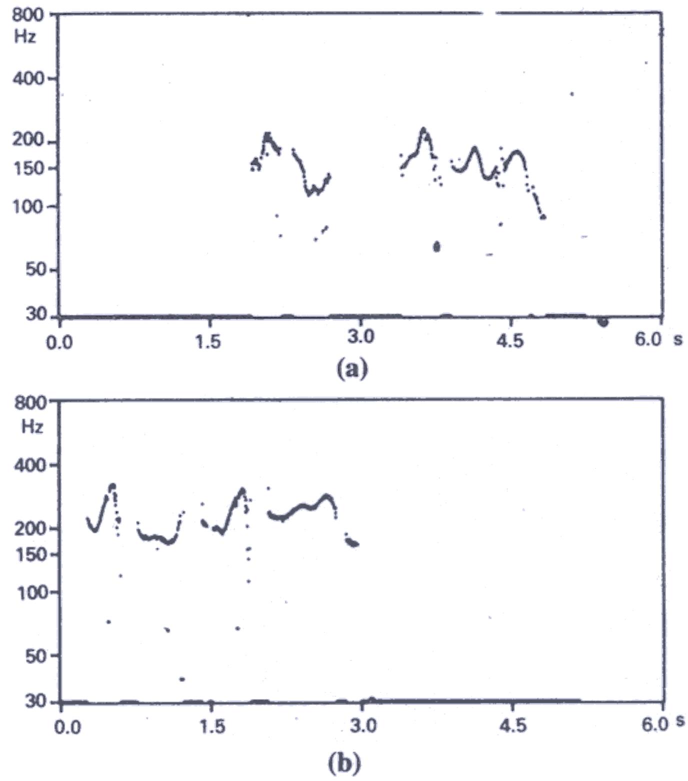


Figure 2. Real-time display of fundamental frequency derived from the electrolaryngograph output waveform for 'he wouldn't say yes, and he wouldn't say no either' uttered by (a) an adult male, and (b) an adult female.

are native received pronounciation (RP) speakers of English. This similarity is clear visually in the real-time display, since f_0 is plotted logarithmically to give a representation which is close to the manner in which humans perceive f_0 changes. This transformation is essential for an appropriate graphical f_0 patterning when presented as a real-time visual display. If a real-time display of f_0 is used during speech therapy or as part of a vocal training programme, statistical measures can be made from the f_0 data to monitor progress. In particular, a histogram can be produced for any length of utterance enabling the mean, mode, median, standard deviation, and range of f_0 used can be calculated[5].

The output from the electrolaryngograph provides a basis for a rigorous f_0 display. However, there are some situations in which it is either not a practical device to use or where it fails to give a usable output, due to for example, extensive layers of fatty tissue covering the larynx. It is also not easy to obtain a usable output from children whose vocal fold contact area is often small. In order to enable real-time f_0 displays to be made available in such situations, other methods for f_0 analysis have to be considered. These are usually based on an analysis of the acoustic pressure waveform picked up by a microphone.

2.2 SINGAD: A Real-Time f_0 Display based on Acoustic Analysis

The analysis of f_0 from the speech pressure waveform can be carried out in the time domain, the frequency domain or as a hybrid of the two.

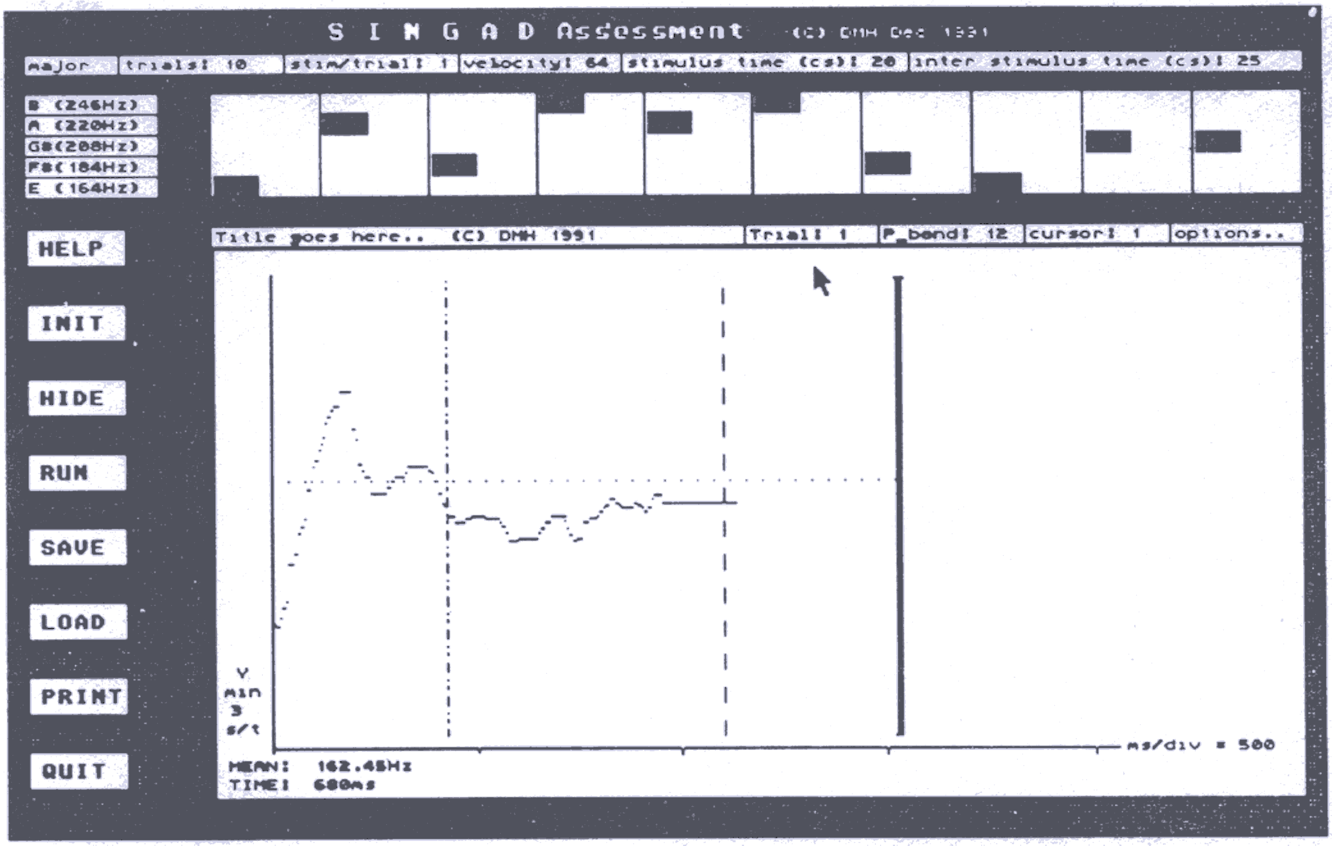
There exists a particular need for a real-time visual display of pitch to enable children (and adults) who are unable to sing accurately a note against a reference from, for example a piano, to develop conscious pitching control. Once the pitch is under conscious control, subjects can work on pitching notes against a reference as they begin to be able to move their pitch 'up' and 'down', or 'sharp' and 'flat' at will to 'tune-in' to a reference note.

The Singing Assessment of Development (SINGAD) system is designed for assessing note pitching ability and developing conscious pitching control. Two versions currently exist, one for the BBC range of microcomputers[6] and another for the Atari computer[7]. The BBC version makes use of a specially modified version of a peak-picking device, originally designed for use with cochlea implantees[8], and the Atari version makes use of a Roland CP-40 pitch to Musical Instrument Digital Interface (MIDI) converter. A particular aspect of the design of the peak-picker is that it is a time domain device which does not incorporate any output smoothing. It has been demonstrated that its output is very similar to that obtained from the electrolaryngograph[9]. Thus it is well suited to the needs of more reluctant SINGAD users whose sung notes have a highly irregular f_0 pattern. An f_0 analyser which incorporates output smoothing would reject such data as being non-periodic and it would not produce any f_0 output. Such a device would not be suitable for application with SINGAD, since it is vital that some real-time trace is given for an input which is phonetically voiced.

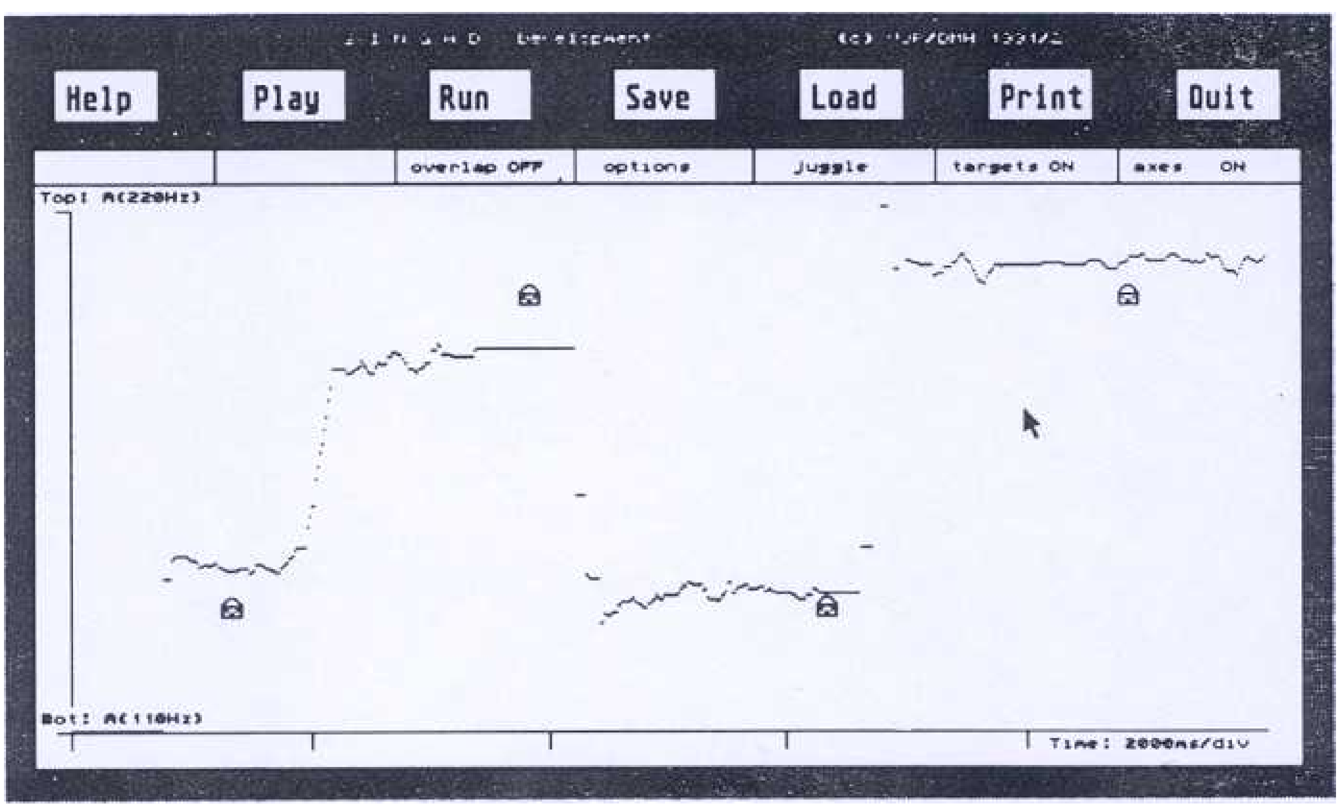
Assessment and development are the two phases of SINGAD. The Atari version has the most comprehensive assessment phase. The computer plays a series of notes via a standard music synthesizer via the MIDI, which the user is asked to sing back. The f_0 for the sung responses is measured and plotted against the reference note for future analysis, either in terms of the f_0 mean, maximum, minimum and standard deviation, or by visual observation of the nature of the contour. In the development phase, a real-time pitch line is displayed which moves from left to right across the screen with time, and up and down according to the sung pitch. Picture targets can be used to give added complexity to the development of conscious pitch control via visual feedback.

Figure 3 shows example SINGAD screens. In the example development screen (Fig. 3(a)), a set of four target houses can be seen with the pitch line produced by the user, which is of an appropriate shape but not quite reaching the appropriate levels to hit the houses. This pitching task is non-trivial, since appropriate pitch changes have to be made whilst the pitch live travels from left to right across the screen. In the example assessment screen (Fig. 3(b)), a set of ten trials are shown by ten separate thick black horizontal lines across the top of the screen. The large panel in the figure shows the results of the user singing the first trial, and the pitch level of the first trial is shown by the dotted horizontal line. Assessment of note pitching ability is achieved by reference to the mean frequency given, which is for the data between the two vertical cursors, in this case 162.45 Hz. This can be compared with the frequency of the stimulus itself, shown in the small panel of five note values on the top left of the screen, and for trial one, it is the bottom note whose value is 164 Hz. The assessment software enables one, three or five note trials to be used in a major, minor or pentatonic context, depending on the singing ability shown by the user.

Work with SINGAD in UK primary schools has demonstrated that 5-7 year old pupils using the development phase on a weekly (10 minutes per week) basis over several weeks show a statistically significant improvement in their note pitching ability over pupils in the same class who do not use the development system[10]. It is further demonstrated that pupils do not require teacher's intervention during SINGAD development work, they can successfully be left to work with the system in pairs.



(a)



(b)

Figure 3. Example SINGAD, (a) assessment screen, and (b) development screen.

2.3 Larynx Closed Quotient Analysis

Other aspects of the voice source can be measured from the *Lx* waveform. One which changes as a function of vocal training is the electrolaryngographically derived larynx closed quotient (CQ). CQ can be defined with reference to Fig. 1 as the percentage of each larynx cycle for which the vocal folds are in contact*.

CQ is measured on a cycle-by-cycle basis. The period (*Tx*) is first determined from the time differentiated *Lx* waveform as the time between the points marked with a δ in Fig. 1. These points also mark the start of each closed phase (CP). Then a local threshold is set as 3/7 of the peak-to-peak amplitude for that cycle (Fig. 1) and the point at which the negative-going *Lx* waveform crosses the threshold is taken as the end of the closed phase. CQ is calculated as given in Eqn (1).

$$CQ = 100^* \left\{ \frac{CP}{Tx} \right\} \% \qquad (1)$$

Howard *et al*[11] studied CQ for the speech and singing of eighteen adult males, and found that those who had received vocal training spoke and sang with statistically significant higher CQ values than those for the untrained group. They suggested the following interpretation in terms of a more efficient voice usage. Firstly, the time for which there is an acoustic path to the lungs via an open glottis is reduced. This results in a reduction in the total acoustic energy transmitted to the essentially anechoic environment on the lungs, an effect known as sub-glottal damping, and therefore lost to the listener. Secondly, less stored air is vented in each cycle due to the decrease in open phase. This enables longer breath groups to be used and/or notes to be held for a longer time and improves the efficiency of power source usage. Lastly, the perceived voice quality is less breathy. A similar study of adult female voices[12] suggests no such clear trend. Rather, there appears to be a patterning of CQ change with f_0 which varies as a function of vocal training such that the more trained exhibit rising CQ values with f_0, while some of the untrained show the reverse trend.

*It should be noted that even during normal speech/singing, the glottis (the space between the folds) may not be reduced completely to zero. There often remains a glottal 'chink' through which air constantly escapes, giving rise to a breathy voice quality. Larynx cosed quotient is measureable whenever the electrolaryngograph picks up changes in vocal fold contact area, whether or not the glottis is closed).

Electrolaryngographically derived larynx CQ appears to be a quantity which varies with professional voice training and a pilot longitudinal study supports this view[13]. Figure 4 shows a display from a real-time display of CQ (%) and f_0 (Hz) against time developed by Howard and Garner[14] for an adult female and an adult male, singing a major scale. Both are receiving singing training. Note the appropriate changes in f_0 in each case as the scale ascends, but that CQ for the male remains essentially constant, while for the female, CQ rises with f_0. It should be noted that in both cases the singers are making use of the same vocal register. This system is based on the *Lx* output waveform, and real-time processing is carried out on a Motorola DSP-56001 integrated circuit which is currently on an Ariel PC-56D board in a PC compatible system. Informal trials with the system in a singing school suggest that this display could be of great interest and usefulness to professional voice users.

3. REAL-TIME DISPLAY OF THE ACOUSTIC OUTPUT

3.1 Spectrographic Displays

The display most commonly used to give a representation of the overall acoustic output during speech and/or singing is the spectrogram[15]. Much of the current knowledge of acoustic pattern variations during human speech[16] and singing[17] is based on detailed studies of spectrograms. Earlier spectrograms were produced by means of a spectrograph, which stores a short snippet (often approximately 2.5 s) of the input acoustic signal and replays it many hundreds of times via a band-pass filter. The centre frequency of the band-pass filter is increased slightly with each replay to cover the analysis ranges, usually 80-8000 Hz, of the instrument. Any signal present at the filter output causes the output picture to be blackened as a function of its amplitude in time and frequency at that point. Generally, the user is provided with a choice of analysis filter bandwidths, of which 300 Hz (wide-band) and 45 Hz (narrow-band) are the most commonly used. These enable the signal to be analysed with good frequency resolution with the narrow band filter, or good time resolution when the wide band filter is selected. Excellent overviews of the operation of the spectrograph can be found in works of Baken[3] and Rosen and Howell[18].

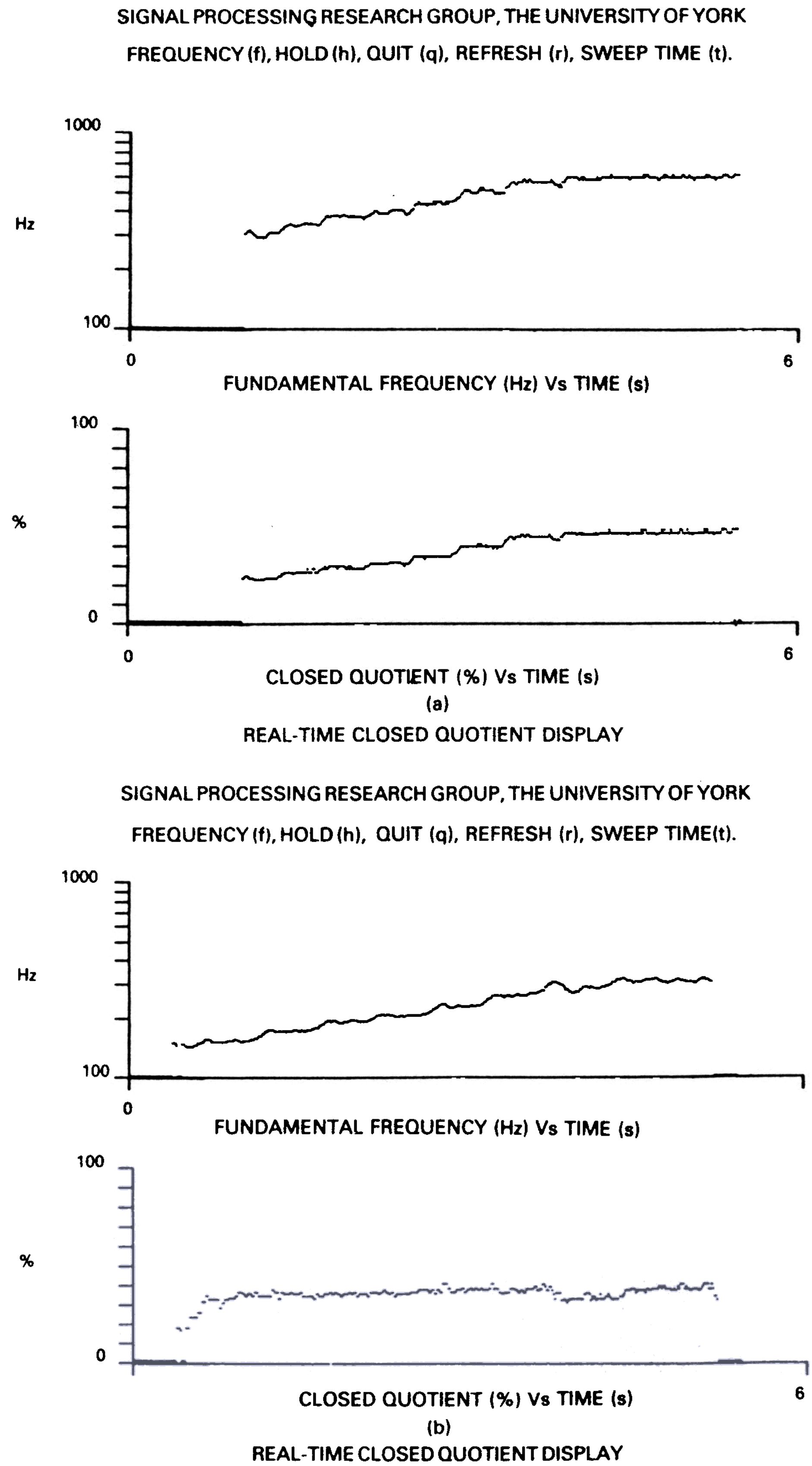


Figure 4. Real-time larynx closed quotient display for a major scale sung by (a) trained adult female, and (b) trained adult male.

Spectrograms are now commonly produced digitally using modern digital signal processing (DSP) techniques. Such spectrograms are produced by sampling the pressure waveform from a microphone, at a frequency which is at least double the upper analysis frequency required, and calculating the discrete Fourier transform over a preset window of the signal. The calculations can be carried out in real-time with DSP integrated circuits like the Motorola 56000 and 96000 series and the Texas Instruments TMS320 series, and real-time spectrogram displays are possible with many types of computers. It is also possible to display individual spectra (amplitude/frequency plots) for single spectrographic analysis windows, or as a long-term average, commonly known as long-term average spectra (LTAS), over any desired number of analysis windows. LTAS analyses are used to study the acoustic nature of steady-state sounds such as sustained vowels. While fundamental frequency information is effectively lost, since no human uttered sound has an absolutely constant pitch, this inherent variation in f_0 serves to illuminate the frequency response of the vocal tract during the steady-state sound.

Figure 5 shows a narrow-band, and wide-band spectrogram for real-time displays uttered by an adult male, along with a short-term and a long-term spectrum in the initial steady portion of the diphthong in time. The loss of fundamental frequency information but gain in clarity of the formant peaks can be clearly observed in the LTAS as opposed to the short-term spectrum. Each vertical striation in the wide-hand spectrogram represents one vocal fold closure, and the thick generally horizontal bars are the changing vocal tract resonances or formants, which move in frequency as different articulation gestures are made. The thin horizontal lines in the narrow-band spectrogram are the individual harmonics of the excitation. The formants are not as clearly shown here, but the changing f_0 can be seen, especially in the upper harmonics. The production of speech/singing sounds can be studied and developed with reference to a real-time spectrographic display, which shows the relevant acoustic patterns. Changes in acoustic patterns when a professional singer sings in different styles are discussed by Howard[19].

3.2 Modelling the Human Peripheral Hearing System

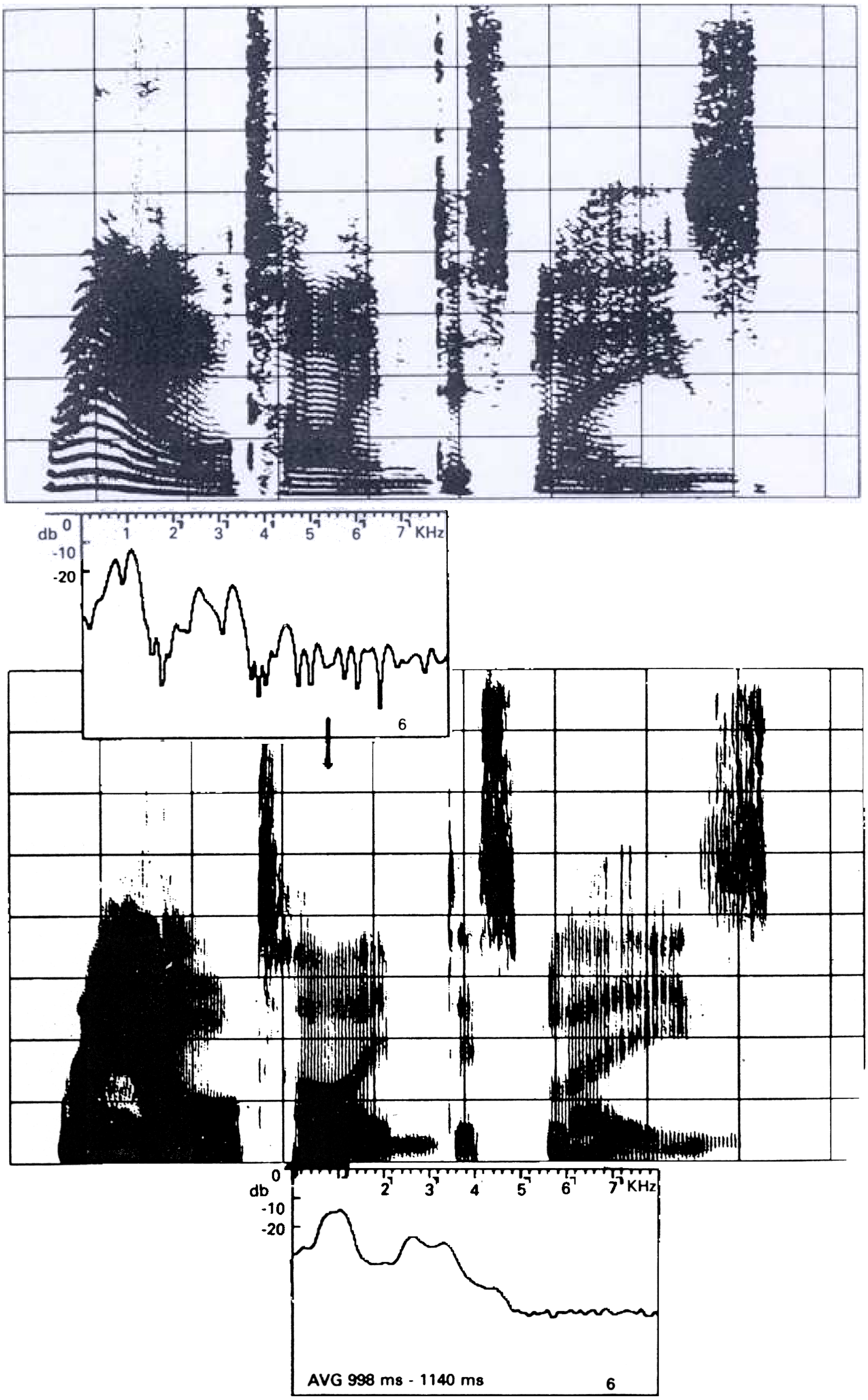
Conventional spectrographic analysis uses an analysis bandwidth which is both constant and fixed. However, psychoacoustic research has demonstrated that measured bandwidths of the human hearing system essentially increase as a function of increasing filter centre frequency, a relationship known as the critical band mechanism. Moore and Glasberg[20] deduced Eqn (2) based on psychoacoustic measurements (Fig. 6) for calculating the equivalent rectangular bandwidth (ERB) value, for a given centre frequency of analysis:

$$\mathrm{ERB}\,[H_z] = 6.23.10^{-6} f_0^2 + 93.39\, e^{-3} + 28.52$$

$$(100\mathrm{Hz} < f_0 < 10000\mathrm{Hz}) \qquad (2)$$

The critical band mechanism indicates that the human peripheral hearing system makes a good frequency analysis at the low end of the frequency spectrum and a good time analysis at the high end. A proper understanding and application of this aspect of the critical band mechanism has, for example, given rise to robust contemporary models of human perception of pitch and timbre[21]. There is considerable interest in modelling the functionality of the peripheral human hearing system to gain same of the advantages over the existing speech/singing analysis systems. A spectrograph which make use of an ERB bandwidth analysis filter system would enable new insights to be gained into the relative importance (in human perceptual terms) of acoustic patterns in speech and singing. It could be a means by which new dynamic and/or static relationships between these patterns are uncovered. New insights may be gained into, for example, improving the naturalness of formant synthesised speech[22], understanding fully the differences between trained and untrained singers[23] and singing in different styles[19], and real-time versions of such a system could provide the basis for a new generation of visual displays for use in voice training and rehabilitation[24].

The computational demands of an analysis and graphics system required for a real-time ERB bandwidth spectrogram system would be close to the limits of many currently available microcomputers. The human peripheral hearing system is modelled as a parallel mechanism, and with current technology, an adequate real-time response can only be achieved by exploiting any parallelism inherent in signal processing algorithms. Transputers provide an economical method for the design and implementation of parallel processing algorithms. Each transputer is a high performance microprocessor, giving floating-point performance



(a)

(b)

Figure 5. Spectrogram for 'real-time displays' uttered by an adult male, (a) a narrow-band (window length of 512 samples), and (b) a wide-band (window length of 96 samples), along with a short-term (above) and a long-term (below) spectrum in the initial steady portion of the diphthong in 'time' (Calibration: sampling rate: 16 kHz; Y-axis: 1000 Hz/division; X-axis: 250 ms/division).

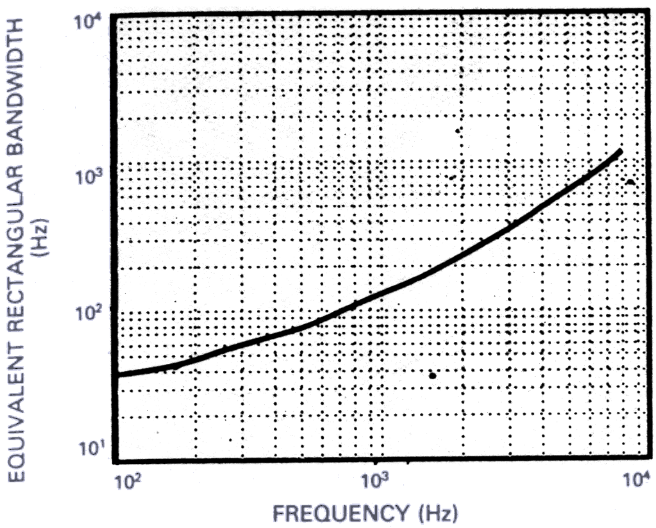


Figure 6. Graph of equivalent rectangular bandwidth against centre frequency.

better than that of a 68020 system, and the inherent design of the transputer ensures that inter-process communications are easy to design and relatively fast.

Tyrrell *et al*[25] describe the design of a transputer-based implementation of a model of the human peripheral hearing. The human hearing model employed in that study uses ERB bandwidth Gamma tone filters, each describing the shape of the impulse response function of the auditory system. These were introduced to describe the shape of the impulse response function of the auditory system as estimated by the reverse correlation function of neural firing times[26]. The Gamma tone filter is defined in the time domain as,

$$gt(t) \propto t^{n-1} \exp(-2\pi bt) \cos(2\pi f_0 t + \phi) \quad (t \geq 0) \qquad (3)$$

The form of the function is that of an amplitude modulated carrier tone of frequency f_0(Hz), with an envelope proportional to $t^{n-1} \exp(-2\pi bt)$, which is the Gamma distribution. The parameters of the Gamma tone filter are n, the order, which controls the relative shape of the envelope, becoming less skewed as n increases; b which controls the duration of the impulse response function, increasing b leads to shorter duration; f_0 which determines the frequency of the carrier; and $\varnothing$ the carrier phase, which determines the relative position of fine structure of the carrier to the envelope. Tyrell *et al*[25] showed that a bank of 20 Gamma tone filters working in real-time in parallel on 16 transputers gave a best sampling frequency of 10 kHz. This gives a 5 kHz bandwidth for real-time operation, which will be sufficient in most cases, since most of the acoustic cues essential for speech production/perception are to be found within this frequency range.

4. ARTICULATION

4.1 Vocal Tract Cross-Sectional Area

The settings of the super laryngeal vocal tract are responsible for the acoustic modifications made to the voice source waveform. It is often difficult for voice teachers to explain exactly in what position articulatory modifications should be made for professional voice users. The output from a real-time visual display which could help here with this aspect of vocal training is shown in Fig. 7 for the vowel of 'spa' sung in a trained and an untrained style by a trained singer. This display shows a cross-sectional area of the oral vocal tract from the glottis to the lips (their positions are marked) which changes in real-time as different sounds are articulated.

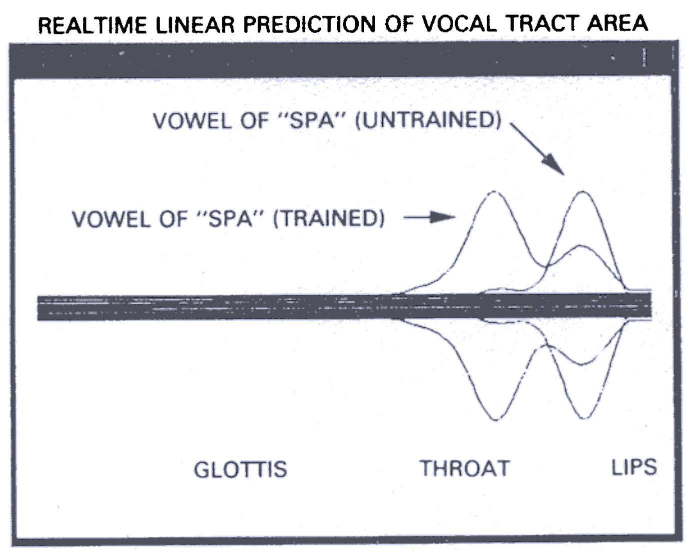


Figure 7. Screen dump from real-time display of vocal tract cross-sectional area for the vowel of 'spa' sung in a trained and an untrained style.

The display is based on a linear predictive coding (LPC) analysis, with user-controlled variable analysis order, of the input speech waveform in terms of an equivalent acoustic tube model for the speech. The algorithm is based on conventional all-pole LPC analysis and thus the resulting tube model has no side branches, and hence it does not model the nasal cavity. The system runs in real-time on a Silicon Graphics Indigo computer[27]. The output has been compared with published X-ray data of vocal tract settings for different vowels and the trends in its output appear to be appropriate. It is not at present possible to compare the output directly with published data for x-ray and other soft tissue imaging techniques, since this would require a reasonable quality synchronously recorded speech pressure waveform on which to base the LPC analysis. Informal observation of the output suggests that it does vary as one would expect to a variety of articulations.

Initial trials with this system have demonstrated changes in articulation shape for different vowel sounds. The effect of an adult male singing a vowel in a trained and quasi-untrained manner (Fig. 7) produces changes in the displayed vocal tract shape which are consistent with those documented by Sundberg[17]. In particular, the trained adult male singer exhibits a large peak in his spectral output when singing in the 2.5 — 4 kHz region. This is known as the singers formant which enables a singer to be heard over large accompanying forces. Sundberg has modelled acoustically this effect and suggested that it was due to an enlarging of the pharyngeal space accompanied by a lowering of the larynx. The model shows an enlarging of the tube just above the glottis for a sung vowel as against its spoken counterpart (Fig. 7).

5. SUMMARY AND CONCLUSIONS

A number of real-time visual displays of fundamental frequency against time have been described, which have been designed for pitch development. A reference display, based on the output from the electrolaryngograph, gives the basis for a highly robust display, but there are situations when it is not either practical or possible to use it. A number of displays based on analysis of the acoustic output are available, and the choice of a particular analysis techniques should be based on the knowledge of the application itself and errors which are acceptable and those which are not. SINGAD is a system designed to help with vocal pitch development, and its input is based on a peak-picker analysis of the acoustic waveform from a microphone.

Other aspects of the vocal source are important, and larynx CQ appears to alter when the voice is trained. A real-time display of CQ appears to have potential application in voice training classes. Spectrographic displays provide frequency/time/amplitude data relating to the acoustic output. These have been available for a number of years and they form the mainstay of speech, hearing and singing research. Real-time versions are becoming widely available with the widespread availability of dedicated DSP integrated circuits and fast computer systems. Of particular interest is the modelling of the human hearing system itself in order to provide a real-time display of acoustic patterns which are closer to those actually received by the brain. Transputers are ideal devices for this, since the inherent parallelism in their operation closely matches the inherent parallelism in contemporary models of human peripheral hearing.

A real-time display of the acoustic tube gives an indication of important areas in the vocal tract with respect to variations in place of articulation. This is achieved by a linear predictive coding analysis of the input acoustic waveform in terms of an acoustic tube model. Results closely match with patterns shown in published X-ray data.

At present, real-time visual displays are being used with those developing their voices for professional use. These could be used with those exhibiting vocal pathologies of human voice production which are essential for efficient voice production. The advantage of working with professional voice users at the outset is that they tend to exhibit trends in their vocal production which are clearly distinguished from non-professional voice users. These trends can be extrapolated in reverse to enable the knowledge and experience gained by their use and the use of other real-time visual displays to be applied in the rehabilitation of those with vocal pathologies.

6. ACKNOWLEDGEMENTS

The author is thankful to the University of York Innovation and Research Pump Priming Fund, and the UK Science and Engineering Research Council grant (No. GR/H14878) for their support for this work.

REFERENCES

1. Hess, W. Pitch determination of speech signals: algorithms and devices. Springer-Verlag, Berlin, 1983.
2. Hess, W. & Indefrey, H. *In* Accurate pitch determination of speech signals by means of a laryngograph. Proceedings of the IEEE International Conference on Acoustics. Speech and Signal Processing, 1984, **ICASSP-84**, 1-4.

3. Baken, R.J. Clinical measurement of speech and voice. Taylor and Francis, New York, 1987.

4. Fourcin, A.J. & Abberton, E.R.M. First applications of a hew laryngograph. *Medical and Biological Review*, 1972, **21**, 172- 82.

5. Abberton, E.R.M.; Howard, D.M. & Fourcin, A.J. Laryngographic assessment of normal voice: a tutorial. *Clinical Linguistics and Phonetics*, 1989, **3**, 281-96.

6. Howard, D.M. & Welch, G.F. Microcomputer-based singing assessment and development. *Applied Acoustics*, 1989, **27** (2), 89-102.

7. Howard, D.M. & Welch, G.F. Visual displays for the assessment of vocal pitch matching development. *Applied Acoustics*, **41**. (in Press).

8. Howard, D.M. & Fourcin, A.J. Instantaneous voice period measurement for cochlear stimulation. *Electronics Letters*, 1983, **19**(19), 776-79.

9. Howard, D.M. Peak-picking fundamental frequency estimation for hearing prostheses. *J. Acoust. Soc. Am.*, 1989, **86**(3), 902-10.

10. Welch, G.F.; Rush, C. & Howard, D.M. A developmental continuum of singing ability: evidence from a study of five-year-old developing singers. *Early Child Development and Care*, 1991, **69**, 107-19.

11. Howard, D.M.; Lindsey, G.A. & Allen, B. Toward the quantification of vocal efficiency. *Journal of Voice*, 1990, **4**(3), 205-12. (see also errata: *Journal of Voice*, 1991, **5**(1), 93-95).

12. Howard, D.M.; Lindsey, G.A. & Palmer, S.K. *In* Larynx closed quotient measures for the female singing voice. Proceedings of XIIth International Congress of Phonetic Sciences, 1991, **5**, 10-13.

13. Howard, D.M. & Rossiter, D. Results from a pilot longitudinal study of electrolaryngographically derived closed quotient for adult male singers in training. *Proc. Inst. Acoustics*, 1992, **14**(6), 529-35.

14. Howard, D.M. & Garner, P. DSP-56000 based real-time electrolaryngographically derived closed quotient. *Proc. Inst. Acoustics*, 1992, **14**(6), 375-81.

15. Howard, D.M. Speech: measurements. *In* Concise Encyclopedia of Biological and Biomedical Measurement Systems, edited by Payne, P.A. Pergamon Press, Oxford, 1991. 370-76.

16. Potter, R.K. & Kopp, G.A. Visible speech. Dover, New York, 1966.

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

18. Rosen, S.R. & Howell, P. Signals and systems for speech and hearing. Academic Press, London, 1991.

19. Howard, D.M. Quantifiable aspects of different singing styles - a case study. *Voice*, 1992, **1**(1), 47-62.

20. Moore, B.C.J. & Glasberg, B.P. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *J. Acoust. Soc. Am.*, 1983, **74**(3), 750-53.

21. Moore, B.C.J. An introduction to the psychology of hearing. Academic Press, London, 1977.

22. Palmer, S.K. & Howard, D.M. *In* Dynamic voice source synthesis. Proceedings of XIIth International Congress of Phonetic Sciences, 1991, 5, 442-45.

23. Howard, D.M. & Collingsworth, J. Voice source and acoustic measures in singing. *Acoustic Bulletin*, 1992, **17**(4), 5-12.

24. Hirson, A. & Fawcus, R. Visual feedback in the management of dysphonia. **In** Voice disorders and their management, Ed. 2, edited by M. Fawcus. Chapman Hall, London, 1991.

25. Tyrell, A.M.; Howard, D.M. & Beasley, N.A. Transputer model of the human peripheral hearing system. *Microprocessing and Microprogramming*, 1992, **35**, 619-24.

26. Patterson, R.D. Auditory filter shapes derived with noise stimulti. *J. Acoust. Soc. Am.*, 1976, **59**, 640-54.

27. Downes, M. Real-time determination of vocal tract articulation for vocal training using linear predictive coding. University of York, Heslington, York, 1992. MSc Project Report (Unpublished).