ALBERT : A system for interactive analysis and display of voice source and acoustic parameters

David Rossiter and David M Howard
Parallel and Signal Processing Applications Research Group,
Department of Electronics, University of York, Heslington,
York, Y01 5DD, ENGLAND
Electronic mail : {dpr,dh}@ohm.york.ac.uk

Abstract

A number of systems have been reported in the scientific literature that are designed to provide voice users with visual feedback to help them develop some aspect of their vocal ability. These systems tend to be based on the derivation of a very small selection of vocal parameters. For example, a number of systems provide feedback on a measurement of the fundamental frequency of the voice alone. However, accurate monitoring and subsequent feedback of the majority of vocal acts requires the consideration of a number of voice source parameters in combination with a number of parameters derived from analysis of the acoustic output of the subject.

This paper discusses a computer program for real-time visual feedback of vocal performance. The software, called ALBERT (Acoustic and Laryngeal Biofeedback Enhancement in Real-Time), provides algorithmically configurable, interactive analysis and subsequent visual feedback of voice source and acoustic output characteristics. Primary parameters of the voice source and acoustic output of the subject are derived from microphone and electrolaryngograph outputs. Further parameters are calculated from user-defined algorithms which operate on the primary parameters. Real-time visual feedback is supported through a user-configurable selection of one, two, and three dimensional graphs. The user is also able to employ colour schemes for enhancing the visual presentation of information.

1 Introduction

The process of speech or singing can be considered as a circular flow of information in which the acoustic product is perceived by the brain as a form of feedback. This enables any necessary adjustments to be made to the mode of phonation in order to better match a target state. This circular pattern is illustrated in figure 1.

The ALBERT software serves to complement aural feedback with feedback based upon an assessment of vocal production presented in a visual form. The software is able to update a number of displays based upon an assessment of vocal performance. The domain within which the ALBERT software operates is shown overlaid on the normal feedback loop in figure 1.

The ALBERT software operates in the following sequence.

ACOUSTIC AND LARYNGEAL BIOFEEDBACK IN REAL-TIME

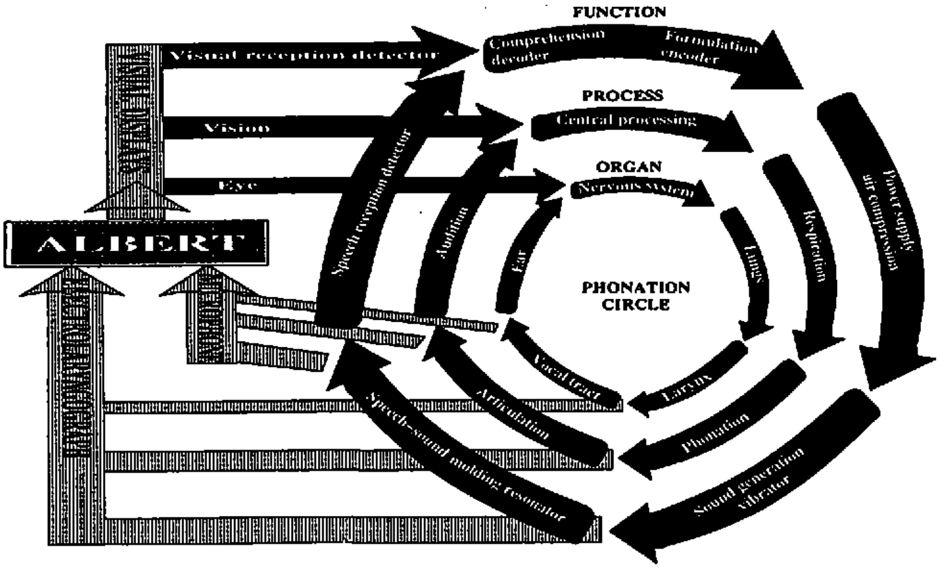

Figure 1: The speech circle with the domain of ALBERT superimposed

1. Analysis of the voice source and acoustic output signals in order to derive a set of primary voice parameter values.
2. Algorithmic assessment of primary voice parameters to create a set of secondary voice parameters.
3. The update of displays communicating the state of any number of primary or secondary parameters in a visual form.

The three steps operate consecutively, at a rate which is effectively perceived as instantaneous. The algorithmic and visualization processes mentioned in steps 2 and 3 are freely configurable by the user. The resulting displays are then perceived by the human visual system as complimentary information to the aural feedback normally present in the feedback loop.

The graphical user interface (GUI) used by ALBERT can also be configured by the user to suit the task for which the software is to be applied. The graphical user interface is divided into separate units, each of which is concerned with a particular function. The user is able to select the position, height and width of each unit, and whether or not it is to be displayed. An example of one configuration in which six units are employed is shown in figure 2. This example illustrates one possible configuration for investigative analysis of the acoustic output signal.

ACOUSTIC AND LARYNGEAL BIOFEEDBACK IN REAL-TIME

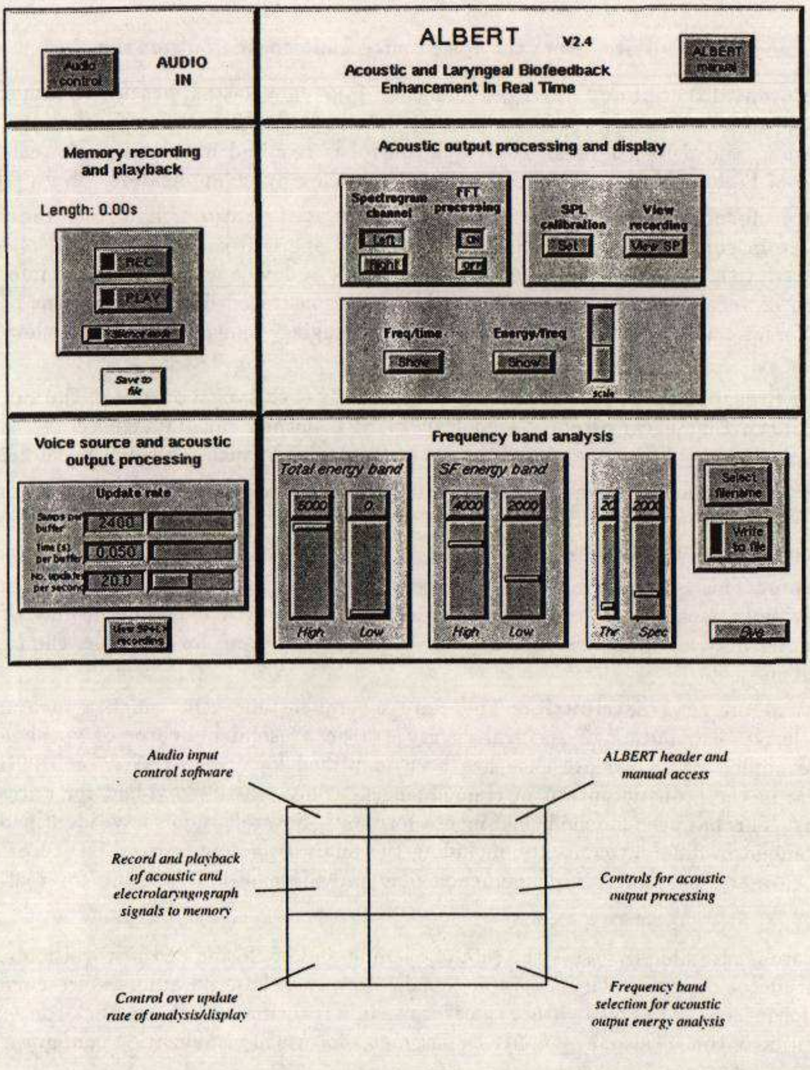

Figure 2: An example configuration of ALBERT for acoustic output analysis

ACOUSTIC AND LARYNGEAL BIOFEEDBACK IN REAL-TIME

2 Voice parameters

Several parameters are derived from the voice source and acoustic output signals.

- **Fundamental frequency (F0)** Fundamental frequency has application in almost all instances of voice useage. It has a primary role in linguistic communication through intonation, and a high level of conscious control is required for singing. A real-time display of F0 may be of use for tuition in both of these voice applications. E.g., [1] [2].
- **Larynx closed quotient (CQ)** This is defined as the ratio of how long the vocal folds are in contact relative to how long they are apart in each cycle. This is often measured with the aid of an *electrolaryngograph* [3], a device which is able to infer the pattern of vocal fold movement. Research has demonstrated different patterns of CQ for different styles of male [4] and female [5] [6] singing, and across different levels of training [7].
- **Sound pressure level (SPL)** The level of loudness of the vocal output of the subject. It is known that a consistently high level of SPL output can damage the voice. A real-time display of SPL may be of use in the treatment of such a form of vocal misuse.
- **Jitter** This is defined as the difference in F0 from one voice period to the next. Example applications include the encouragement of pitching stability for singers, and for use in post-operative voice therapy after surgery.
- **Shimmer** This is defined as the difference in SPL from one time instant to the next. It provides a measure of peturbation in voice production. A real-time display of this parameter may be of use for encouraging stable phonation in, for example, the tuition of deaf subjects.
- **Spectral energy distribution** There are several instances in which a measure of the relative distribution of spectral energy can be a useful indicator of vocal usage. For example, a spectral presence has been identified [8] in the region of 2100Hz to 3800Hz in the acoustic output of trained singers which was less evident for untrained singers. This has been labelled the 'singer's formant'. Several studies have identified this phenomena in different contexts, including the analysis of tenors from CD recordings [9]. An assessment of this phenomenon may provide a useful measure for real-time tuition.

The software is also able to assess the rate of opening and of closing of the vocal folds, and the amplitude of the electrolaryngograph signal. However, these parameters are currently intended for reasons of analysis rather than for use in a real-time visual feedback context. A representation of time is also presented as a parameter for use in visualization configurations.

2.1 Primary voice analysis

Within ALBERT, the voice source and acoustic output signals are processed in the following ways.

- **Voice source signal** Several voice source parameters, including F0, CQ, and jitter are derived from analysis of the electrolaryngograph signal. The basic procedure for this is as follows. First, the polarity of the signal is first checked to ensure that positive changes reflect increased inter-electrode current flow. This is based on the assumption that the vocal fold contact area changes more rapidly when it is increasing than when it is decreasing (that is, the folds snap together more rapidly than they part). This is a known phenomenon. The points of maximum increasing vocal fold contact can be located by deriving the positive peaks in the time differentiated waveform. These are then used to define the start of the closed phase (CP) in each cycle. The time between these peaks is used to provide a measure of the fundamental period (Tx). The fundamental frequency is then assessed as F0=(1/Tx). The end of the CP is the instant when the negative-going electrolaryngograph waveform crosses a fixed ratio of the current cycle's amplitude. The ratio, set at 7:3, has been shown to exhibit a result close to that obtained by inverse filtering [7]. CQ is then obtained by the following equation: $CQ = ((CP/Tx)*100)$
- **Acoustic output signal** A 128 point Fast Fourier Transform (FFT) is carried out on the digitally sampled representation of the acoustic wave. This normally occurs at a rate of twenty times a second. Coefficient values for pre-determined energy ranges are then summed. These can be compared to provide an indication of vocal performance. For example, a measure of the overall tilt of the voice source spectrum can be derived by comparing the ratio of energy between 0-1kHz and 1-5kHz [10].

2.2 Algorithmic processes

The voice is a complex mechanism with many different biological, physical, and physiological qualities interconnected. Consequently, many voice parameters are influenced by the performance of other parameters. For example, it is a well-known phenomenon that increases in SPL are usually accompanied by proportional increases in F0. Any real-time display of a single vocal parameter should be able to consider the changes that would be expected to occur as a consequence of changes in other parameters. The ALBERT software is able to algorithmically assess the state of any number of vocal parameters in combination. For example, a real-time display of F0 for deaf subjects might take into account the degree of increase in F0 that would be expected to occur as a consequence of increasing SPL. The resulting display would then more accurately convey progress towards target performance.

2.3 Visualization

The ALBERT software is able to display and update a number of visualization configurations in real-time with which to communicate the state of the assessed vocal parameters. Visualization techniques include

- 1D bar graph displays
- 2D graph displays

- 3D displays
- The use of colour

Other real-time voice display software tends to present the user with only a small selection of visualization options such as frequency against time or spectral energy against frequency. The flexibility proferred by the ALBERT software enables the user to create the form of visualization which appears to be most suitable with regard to the vocal task and the assessment skills of the subject.

3 Example applications

3.1 Adult singers

A real-time display based upon quantitative evaluation of the singer's formant phenomenon previously discussed may be of use in the process of singing tuition. The ALBERT software could be configured to function as a tool for this purpose in the following manner. First, a new parameter is created which assesses the energy in the frequency range 2-4KHz relative to the energy in the frequency range 0-5KHz. The former range is that within which the presence of the singer's formant has been observed to occur. The latter range is an approximation to the frequency range across which acoustic output can reliably be assessed. The resulting parameter could then be displayed in a number of ways, such as a plot on a 2D graph with F0 or time parameters mapped to the x-axis.

This parameter, in addition to others, has been used in a formal test of the benefits of the ALBERT software when used as a complementary tool for the process of singing tuition [11]. Results support the proposition that use of the ALBERT software has reinforced the process of vocal tuition.

3.2 Child singers

A similar phenomena has also been observed in child singers. Current research has indicated that children phonating at high fundamental frequencies exhibit increased energy in the 700-1800Hz region as a function of vocal training [12] . A real-time display based upon quantitative evaluation of this phenomenon may therefore be useful in the tuition of child singers.

This could be achieved by using the same process as that described for adult singers, but with selection of the frequency range 700-1800Hz instead of 2-4KHz. A further modification may be to de-select the drawing of the axes in the display, which may serve to confuse or distract the subject.

3.3 Subjects with short attention span

A number of voice users are unable to concentrate for long periods of time. This includes children and adults with learning difficulties. The ALBERT software can be configured to

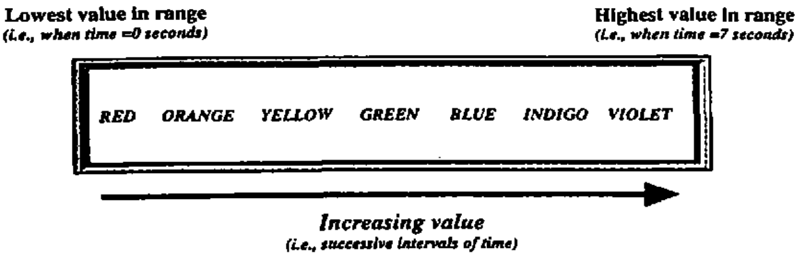

Figure 3: An example colour scheme for use in mapping *time* as a mechanism to encourage attention

complement the display of a parameter with colour in a number of ways. This facility can be used to create a real-time display which continually changes colour.

This first requires the design of an appropriate colour scheme. An example is shown in figure 4. All the colours of the rainbow have been employed in the colour set. The lowest data values are then mapped to the first colour in the range (red), with the highest data values mapped to the last colour in the range (violet). Values between these extremes are mapped to the other colours according to their relative position between the two extremes.

This colour scheme can then be mapped to reflect the value of a parameter. In this example, the colour set would be mapped to *time*. In this way the display would continually change colour. The time for which the colour set would cycle can be predetermined. An appropriate example may be seven seconds. The display would then change to a different colour with the passing of each second.

4 Conclusions

This paper describes computer software which has been created for real-time analysis and visual feedback of a number of voice source and acoustic output qualities. Parameters may be algorithmically processed to produce further measures representative of some measure of vocal development. The software has found application in a number of voice assessment contexts.

5 Acknowledgements

Thanks to Peta White of the Roehampton Institute for help with information concerning the spectral output of developing child singers. This work was funded by SERC research grant GR/H14878.

6 Further information

Further reports are freely available through the process of anonymous ftp transfer from *ftp.york.ac.uk*, in directory */pub/users/elec10/voice*. World wide web (WWW) access is available via *file://ftp.york.ac.uk/pub/users/elec10/voice/Welcome.html*.

References

[1] SPAAI, G.W.G., STORM, A., & HERMES, D.J. A visual display for the teaching of intonation to deaf persons: some preliminary findings. *Journal of Microcomputer Applications*, 16, pp277-286

[2] HOWARD, D.M., & WELCH, G.F. Microcomputer based singing ability assessment and development. *Applied Acoustics*, 27:2, pp89-102

[3] FOURCIN, A.J., & ABBERTON, E.R.M. First applications of a new laryngograph. *Medical and Biological review*, 21, pp172-182

[4] HOWARD, D.M. Quantifiable aspects of different singing styles - a case study. *Voice*, 1:1, pp47-62, 1992

[5] ESTILL, J. Belting and classic voice quality: some physiological differences. *Medical problems of performing artists*, Philadelphia : Hanley and Belfus, pp37-43, March 1988

[6] EVANS, M. & HOWARD, D.M. Larynx closed quotient in female belt and opera qualities: a case study. *Voice*, 2:1, pp7-14, 1993

[7] HOWARD, D.M., LINDSEY, G.A., & ALLEN, B. Toward the quantification of vocal efficiency. *Journal of Voice*, 4:3, pp205-212, 1990 {see also Errata:*Journal of Voice*, 5, pp93-95, 1991}.

[8] SUNDBERG, J. Articulatory interpretation of the 'singing formant'. *Journal of the Acoustical Society of America*, 55, pp838-844, 1974

[9] HOWARD, D.M. & LINDSEY, G. Spectral features of renowned tenors in CD recordings. *Proceedings of the International Conference on Speech Research*, pp17-20, 1989

[10] LOFQVIST, A. The long-term average spectrum as a tool in voice research. *Journal of Phonetics* 14, pp471-5, 1986

[11] ROSSITER, D.P. & HOWARD, D.M. Vocal development with and without the influence of real-time visual feedback. *To be submittted to the Journal of the Acoustical Society of America*, 1994

[12] WHITE, P.J., WALKER, R., & WELCH, G.F. Spectral energy measures of trained and untrained children's voices. *Paper in preparation, Leverhulme Project, Roehampton Institue, London*