Vocal Process

Sharing Information
Promoting Expertise

Advanced Vocal Training with Vocal Process: Computer Voice Analysis

Jeremy Fisher, Gillyanne Kayes and Matthew Reeve

Abstract

Computer voice analysis is used in three areas of our work:

- Explanation: by clarifying the relationship between vocal function and acoustic output.
- Biofeedback: as a visual tool helping clients develop proprioceptive feedback of their instrument.
- Assessment: as a practical and objective mechanism to assess progress and development of accuracy in voice production.

Computer voice analysis provides valuable visual feedback for our diverse range of clients. These include classical and non-classical singing teachers, voice and speech teachers, speech and language therapists, and performers from all genres.

Computer voice analysis also forms part of the assessment protocol of personal vocal practice on the Vocal Process 'Integrated Voice' training programme.

This poster presents a range of vocal tasks. Each task is illustrated with spectrograms produced using Kay Elemetrics MultiSpeech software (Model 3700 Version 2.2), running on a Pentium 2.66GHz PC under Microsoft Windows XP, and a handheld Creative Labs MC1000 microphone. Recordings use a sampling rate of 11,025MHz.

There is an additional introductory guide to reading and interpreting spectrograms.

Reading Spectrograms: A Guide for Vocal Practitioners

When reading a spectrogram, you are usually presented with two screens. When you input a sound, either by loading a pre-existing sound or using a microphone, it appears on the upper screen.

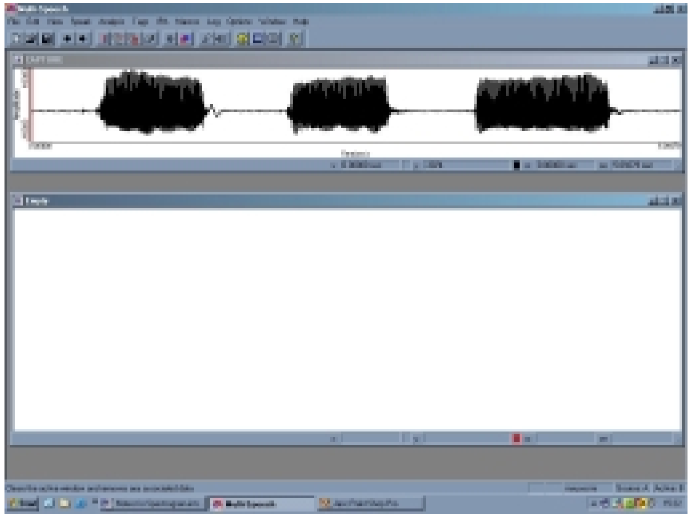
Figure 1

Figure 1 shows the vowel /i:/ or 'EE' repeated three times at the same intensity. Time (in seconds) is shown horizontally across the page and intensity (amplitude in decibels) is shown vertically.

On the lower screen in Figure 2 a spectrographic view of the same inputted sounds is shown. Time is again displayed horizontally, with pitch (frequency in Hertz) now shown vertically. The intensity of signal is represented on this screen by how black each area is—the stronger the harmonics, the blacker the horizontal line. A white area within a signal means that the harmonics are missing or too weak to be read. Using the spectrographic reading you can discover information including fold mass activity, the relationship between phonation and airflow, activity of the false vocal folds and laryngeal set-up.

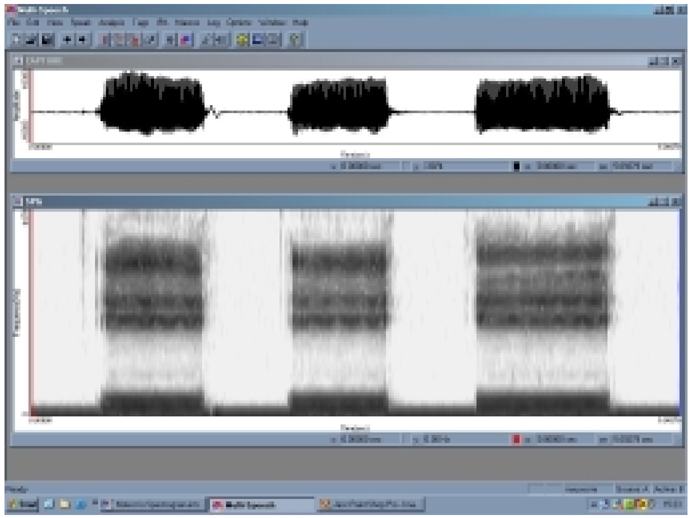
Figure 2: Spectrographic view

Figure 3 shows the formant view. Here the position of the formants in each sound are shown. Formants are clusters of harmonics that are stronger than the surrounding areas. Again, time is shown horizontally and frequency is shown vertically. In spoken or sung examples, formant readings are used to identify vowel shaping and also the relationship between laryngeal height and the tongue.

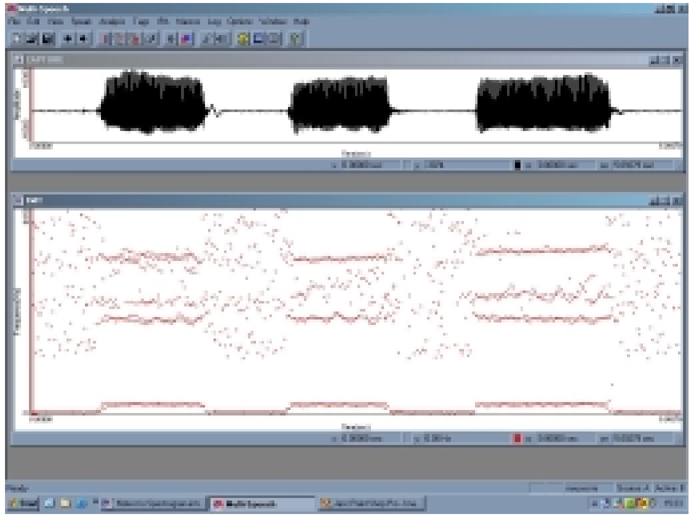

Figure 3: Formant view

Figure 4 shows how the formant view may be overlaid onto a spectrographic screen for ease of analysis.

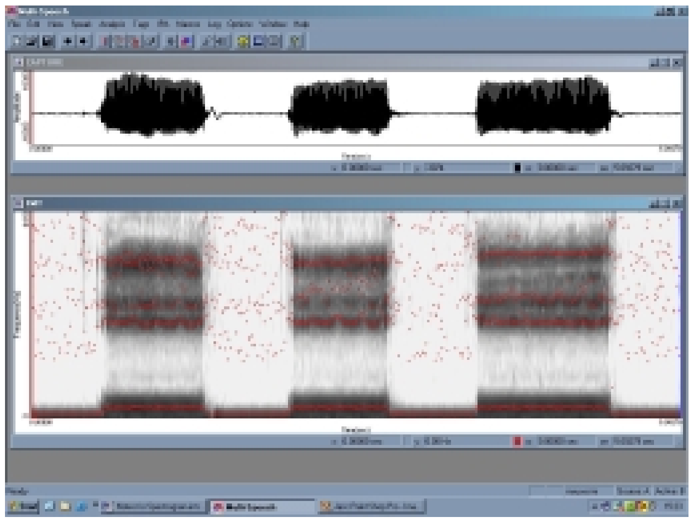

Figure 4: Spectrographic view overlaid with formant display

Tongue Positioning & Vowel Production

The sample demonstrates different tongue positions that may be used in vowel production and their effect on vocal function. A by-product is an increased awareness on the part of the vocalist on the differences between vowel and vocal quality. We have found this task especially useful for those classical singers who depress the tongue dorsum in search of a dark sound quality.

Figure 5 and 6 show a chain of five vowels, /i:/ to /u:/ and then back again, using two different tongue positions. In Figure 5 the tongue is retracted and the dorsum depressed (backed, low and flat). In Figure 6 the tongue position is close and fronted (higher in the mouth and more forward).

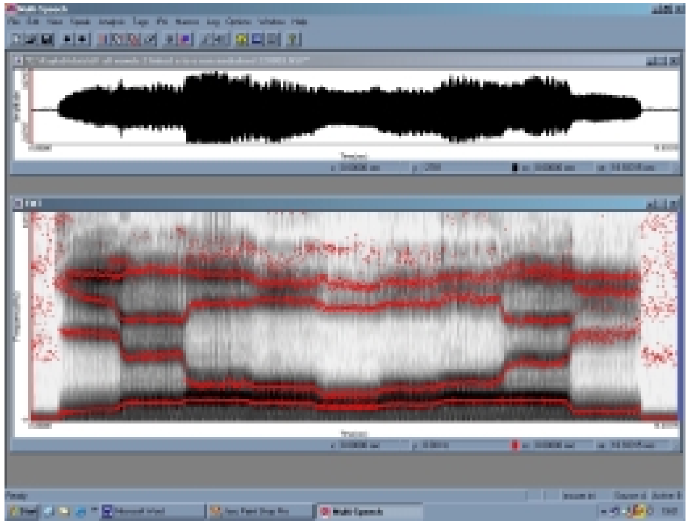

Figure 5: /i:/ to /u:/ and back again with tongue retracted and the dorsum depressed

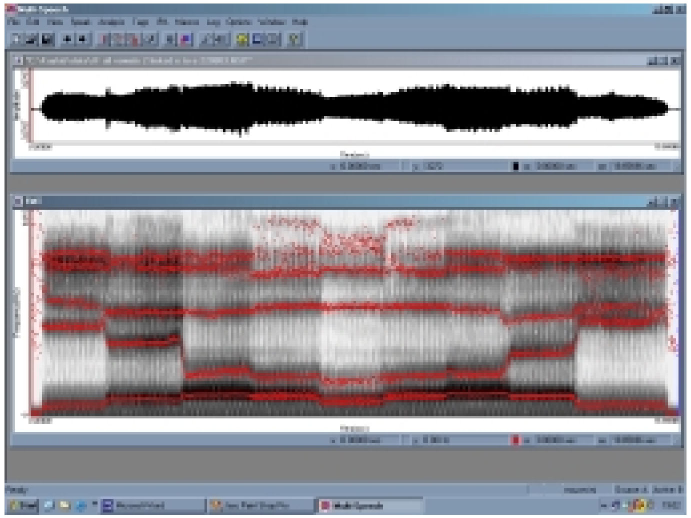

Figure 6: /i:/ to /u:/ and back again with tongue 'medialised'

Note the effect on the first and second formants, which move closer together in Figure 5 and stay further apart in Figure 6. In Figure 5 the third formant raises on the back vowels. More energy seems to be required below 500Hz—the signal behind the first and second formants is more intense and appears blacker on the spectrogram. We consider this reading to be showing a degree of pressed phonation, which might lead to vocal fatigue.

In Figure 6 notice the third formant is more stable. The balance of intensity is spread out, and the reading seems visually clearer and less cluttered. The higher, less retracted, tongue position in this sample will have the effect of 'medialising' the vowels, see Kayes (1).

Acoustically it may be advantageous to medialise the vowels in this way. A common problem in singing and speaking is that 'back' vowels appear to lack intensity and brightness in comparison with the 'front' vowels. Medialisation can be used to achieve equally bright vowels across the spectrum. Using this strategy the vowels are distinguished by altering the middle of the tongue rather than the front or back of the tongue.

It is interesting to note that in terms of vocal effort the second sample was easier to produce, with less effort in breath and at vocal fold level—the balance between flow and resistance was perceptually more even.

Tone Onsets: Glottal Onset & Glottal Attack

Figure 7 shows the vowel /iː/ or 'EE' is repeated four times, twice with a glottal onset, and twice with a hard glottal attack. The aim is to explore the relationship between breath and phonation in one type of onset.

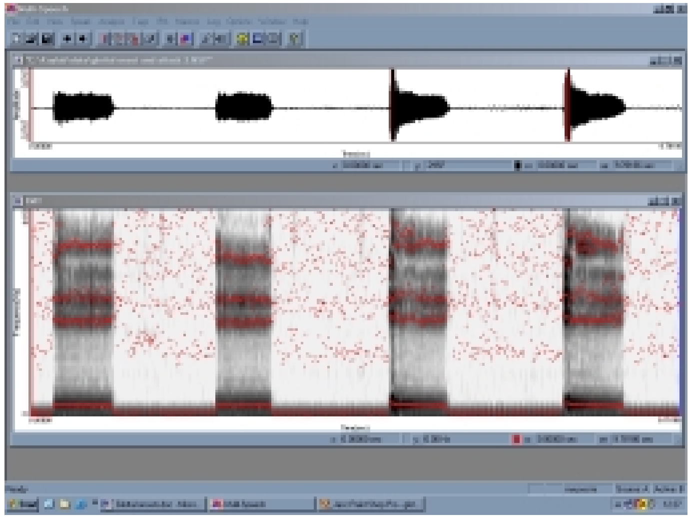

Figure 7: Glottal onset and glottal attack

At the beginning of the first two samples the signal starts with a clean vertical line from the bottom to the top of the spectrographic screen, and the rest of the pattern is even. In the latter two samples the vertical line at the start of the sound is much darker, more uneven, and lasts longer, disturbing at least a quarter of the reading.

This can be seen more clearly in the enlarged display of samples two and four: Figures 8 and 9 respectively. Sample two shows a clean start that settles quickly into a stable reading.

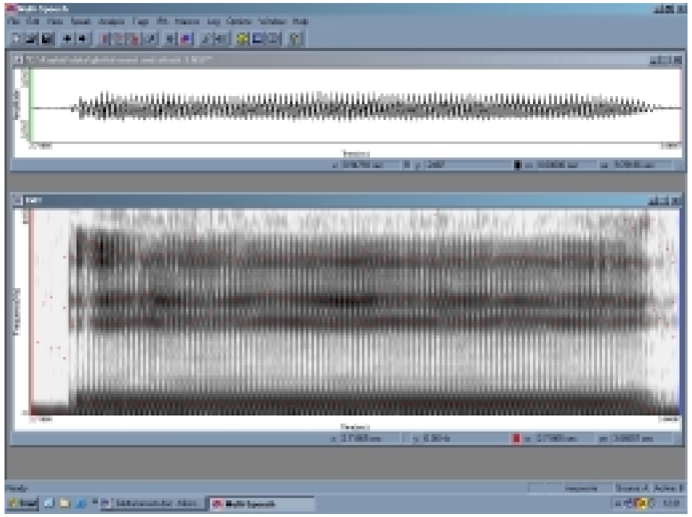
Figure 8: Sample two: glottal onset

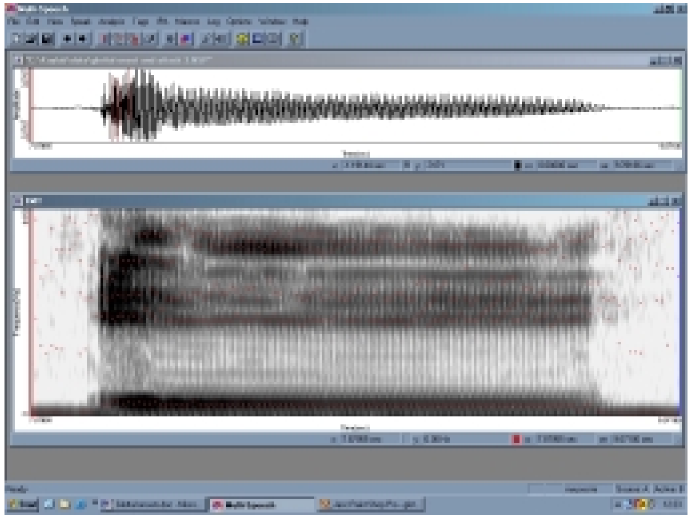
Figure 9: Sample four: glottal attack

Sample four begins with a much stronger attack that lasts more than 160 milliseconds. An excess of air pressure has built up behind the glottal closure, and the resulting explosion takes almost half the sample to control.

This exercise can be helpful in eliminating breathy tone; it can also be used to address hard glottal attack by establishing appropriate effort levels in breath and vocal folds.

Speech, Nasal Speech & Nasal Twang

Many of our clients are musical theatre and pop singers. They are not always able to differentiate between hypernasality and 'twang' (the 'ring' of the voice or the singer's formant), which is required for heightened effect in these vocal genres. We use the following exercise to separate the resonance function of the nasal port from that of the aryepiglottic sphincter (epilarynx or laryngeal collar).

Figure 10 shows three different vocal sets that are held and linked. An oral tone is changed to a nasal tone, then twang is introduced as an additional factor.

Oral to nasal: in the first segment the nasal port is closed as in normal vowel production for /i:/. In the second segment, the vocal set is held, but the nasal port is partially opened (air and sound are coming out of both mouth and nose). Note the appearance of a new band of harmonics and a new formant between 500Hz and 1,000Hz.

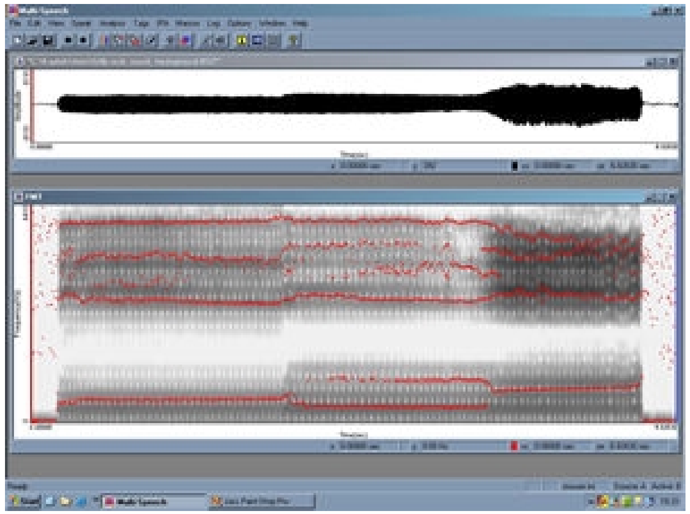

Figure 10: oral tone, nasal tone and nasal tone with twang

Nasal to 'nasal twang': in the third segment, the vocal set is held with the nasal port half-open, and twang is added. The aim is to make the upper part of the reading turn substantially darker keeping the lower part the same. This shows the increase in amplitude comes from the tightening or narrowing of the laryngeal collar and not from pressed phonation. Pressed phonation (pressing the true vocal folds together to achieve a brighter or louder tone) would show as a major darkening of the signal in the lower portion of the screen (below 500Hz) with very little change above, resulting in a more unbalanced reading.

Figure 11 shows a more advanced vocal task is to move between oral to nasal twang. Here the vocal set for twang is held, and the nasal port is opened and closed. The aim is to keep the blackened area in the upper half of the spectrographic screen constant while moving between oral to nasal resonance. Everything else should remain stable. Note the alteration of energy in the lower half of the screen as the resonation in the nasal cavity is added and removed.

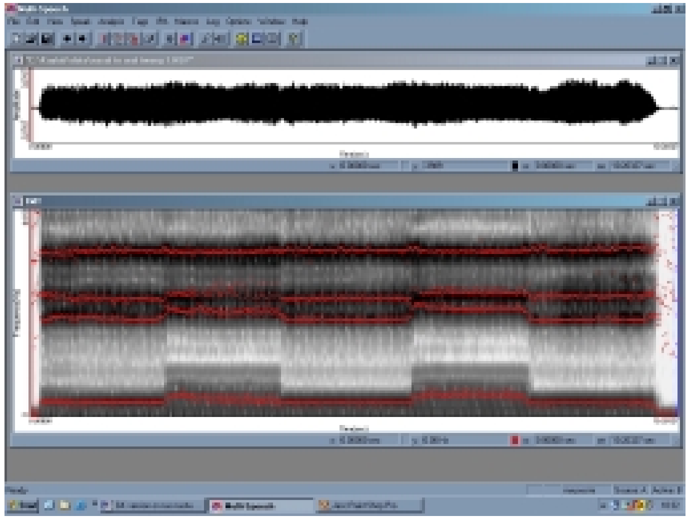

Figure 11: Oral to nasal twang

Demonstrating a Song in Four Different Vocal Sets

Figure 12 shows ‘Happy Birthday’ in D♭ as sung by bass JF using four distinct voice qualities: sob, nasal twang, opera and speech. This task enables the user to check their accuracy in creating differentiated vocal sets.

First, notice the red formants in the lower screen, which ‘sit’ lower or higher on the screen for each vocal set. When the formants move higher or lower as a group, this indicates a change in the height of the larynx.

The first quarter of this reading is in sob quality made with a lowered larynx, ‘thin’ vocal folds and a tilted thyroid. The formants are lower as a group, and the majority of signal strength is towards the bottom of the screen.

The second quarter is in nasal twang. The vocal folds remain at their thinner fold setting but that there is increased formant energy up to 3,000 Hz from aryepiglottic constriction. There is slightly more concentration of energy in the upper harmonics, and the larynx is higher (the formants are higher on the screen).

The third quarter is in full opera quality—there is a ‘thick fold’ reading, and the singer’s formant (the twang factor from the previous reading) is maintained with strong readings in the middle of the screen. The formants move down as a group as the larynx is lowered from the higher position required to access twang quality.

The fourth quarter is in straight speech quality (modal register), with ‘thick folds’, no laryngeal tilt and a resting larynx height. It is interesting to note that, even as a bass, the resting height of the subject’s larynx is fairly high.

A more detailed description of the qualities described above and their physiological characteristics is given in Estill (2), Harris *et al* (3) and Kayes (4).

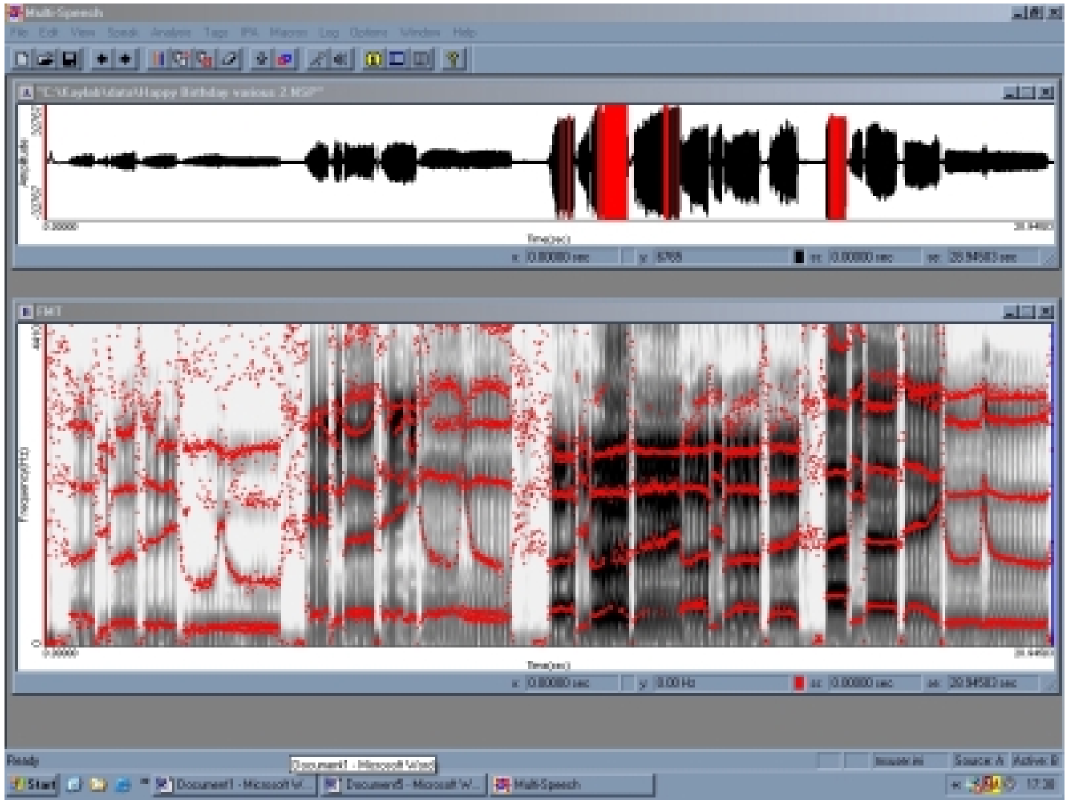

Figure 12: ‘Happy Birthday’ sung using four distinct voice qualities: sob, nasal twang, opera and speech

References

1. Kayes G. (2000). *Singing and The Actor*, A & C Black, London, 101-105.
2. Estill J. (1981). An analysis of the spectra of four voice qualities: Speech, Sob, Twang and Opera. In V Lawrence (Ed). *Transcripts of the 10th Symposium: Care of the Professional Voice*, The Juiliard School, New York City, 1982, 31-50.
3. Harris, Harris, Rubin and Howard (1998). *The Voice Clinic Handbook*, Whurr, 1998, 77, 172-5, 181.
4. Kayes G. (2000). *Singing and The Actor*, A & C Black, London, Chapter 12.

Contact Details

Address: Vocal Process Ltd
137 Sunderland Road
LONDON
SE23 2PX
United Kingdom

Telephone: +44 (0)20 8291 1758

Email: info@vocalprocess.net

Web Site: http://www.vocalprocess.net