Direct Displays of Phonetic Dimensions

Evelyn Abberton, UCL
Adrian Fourcin, Laryngograph Ltd.

A major concern in teaching Phonetic Science is to establish relationships between different levels of representation of speech: production, acoustics and perception. Pedagogically, it is essential not to overcomplicate or oversimplify acoustic information. For example, the rich segmental and prosodic information in classical spectrographic displays is not easy to extract; on the other hand, stylised, symbolic representations of voice quality and intonation can often be less than informative.

In this presentation we illustrate a hardware and software system (Speech Studio from Laryngograph Ltd.) that allows the user to analyse, measure and display phonetic dimensions in normal and pathological speech in ways that are based not simply on arithmetical or computer processing convenience, but which are explicitly related to characteristics of human auditory processing, and are directly based on speech production: as well as microphone signals, information is gathered from (superficially applied) larynx and nasal sensors (Fourcin & Abberton (2007). The corresponding visual displays are informative in their own right, and also valuable for biofeedback in the teaching and learning of new phonetic contrasts and patterns – both segmental and prosodic. Speech Studio is a digital system interfaced to a laptop computer. Conventional spectrograms and speech and EGG (Laryngograph) waveforms are available, but also Speech Pattern Element time and frequency displays which take account of auditory processing in these domains. Dimensions available include fundamental frequency, which may be supplemented by amplitude of the speech signal, for the study of correlates of pitch, loudness and rhythm – tones and intonation. Fundamental frequency is displayed on a logarithmic not linear scale since this is how we perceive pitch. The fundamental frequency traces are derived directly from the vibrating vocal folds with no smoothing, so that the presence of voiceless segments is clearly visible as are the micro-intonation due to voiced obstruents and the irregular vibration related to phonation type. In addition, a representation of voiceless frication is available, showing aspiration and the different frequency spread of intensity among sibilants. This element is lowered visually on the vertical frequency scale to represent the auditory integration of this high frequency information with the lower frequency elements of the speech. The presence of sections of nasalised speech can be shown by a change of colour of the fundamental frequency-amplitude trace.

Speech Studio analyses are used by phoneticians and speech scientists to examine the segmental and prosodic features of both well-studied and less well investigated languages. See, for example, Lindsey et al (1992), Heselwood (1996). Speech and Language Therapists are major users of Phonetics as a tool in the description and diagnosis of pathologies. Evidence-based practice demands that clinicians demonstrate results, and in the field of dysphonia and voice quality Speech Studio analyses of voice pitch range and phonation type are becoming standard. In addition, analyses related to temporal, pitch irregularity, and to vocal fold contact phase variability also offer new quantitative insights into aspects of perceived voice quality – an area of normal and clinical phonetic study notorious for descriptive vagueness and ambiguity.

Pitch, loudness and friction are three important phonetic perceptual dimensions and it is very helpful in teaching and research to be able to display their physical correlates simply. The spectrogram,although it is a standard tool in the phonetcs laboratory, is not very helpful in this regard because the physical correlates of these percepts are not readily separable from the mass of acoustic information displayed.

The display shown in Figure 2 makes use of the same recorded utterance that was basic to the spectrographic analysis of Figure 1. The fricative sounds appear in the high frequency part of the spectrogram; in the pattern display of Figure 2 they have been transposed to the region just above the fundamental frequency range in order to provide a visually compact representation. [ʃ] and [s] are not voiced and they appear clearly demarcated from the voiced intonation contours shown below. These voice contours have their midlines set by the fundamental frequency of phonation, measured on a cycle by cycle basis, whilst their thickness is set by the peak acoustic amplitude measured for each individual vocal fold cycle. Stress on the work "see" is associated with an increase in loudness – shown by an increase in thickness of the line – and , after the typical precursive rise. for a voiced sound, a prominent fall.

Figure 3 illustrates the effect of using a contrastive intonation with the stress on "can". Here the stress on the initial voiceless aspirated consonant is indicated by the burst and the pitch contour for the voiced sound starts without a precursive rise.

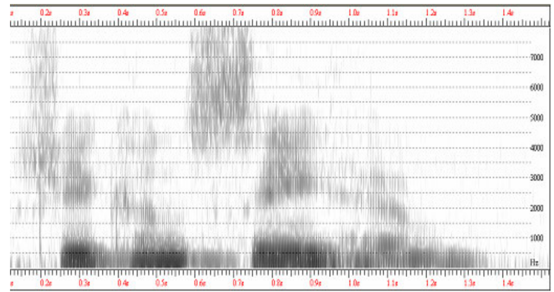

Figure 1 Broad band spectrogram
She can `see him

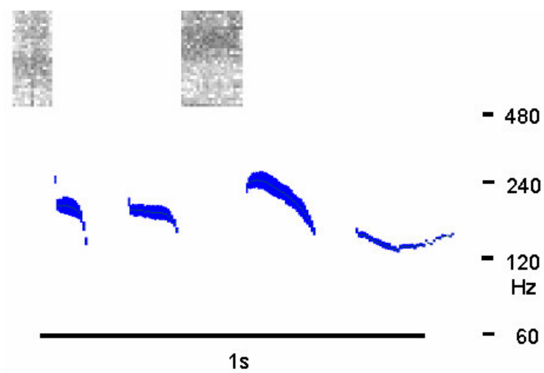

Figure 2 pitch loudness friction
She can `see him

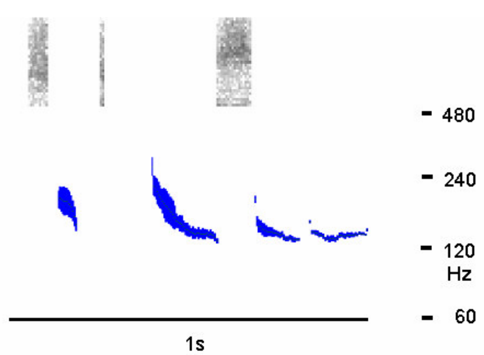

Figure 3 pitch loudness friction
She `can see him

Nasality is a further perceptually prominent dimension that is also not easy to visualise in the spectrographic analysis of a speech sound sequence. Figure 4 illustrates a particular way of linking this percept to the previous types of display by the use of a simple colour change in the intonation contours that are linked to the voiced parts of sequence. In all these examples whilst the figures are compelling in their simplicity, the full value of the processing is only properly appreciated when live interactive displays are used. We find that the student can make corrections "on the fly" by making use of the continuously stored picture on the screen of the computer.

The hardware is contained in a small USB box and runs on a laptop computer. Our hearing mechanisms are far more powerful than even the most advanced current processors and the secret of these displays lies in the use of the laryngograph® to provide accurate real-time signals. This availability of laryngeal information also makes its feasible to show aspects of the laryngeal articulations that can distinguish segmental contrasts. Figure 5 shows a normally aspirated pulmonic egressive [pʰ] then, immediately below, an ejective, where the upward compressive movement of the larynx is associated with pre-phonatory vocal fold closure, marked by an upward spike, and a following baseline rise. Finally, in this figure, implosive consonant articulation requires a decompression and this is linked to the downward movement of the larynx as a whole and shown here by the rather large downward baseline swoop of the laryngograph trace. In all these traces the thick green trace is the Lx signal.

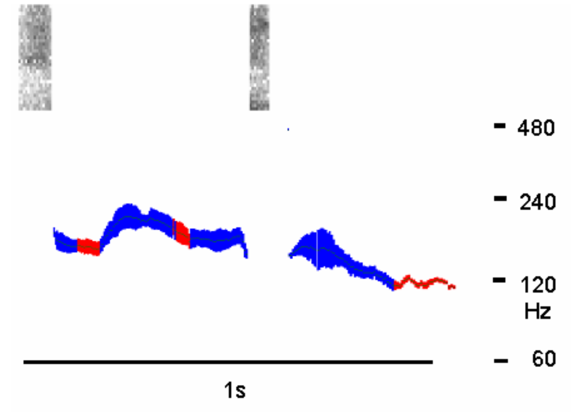

Figure 4
pitch loudness friction nasality
Some ˈreally 'nice ˎwine

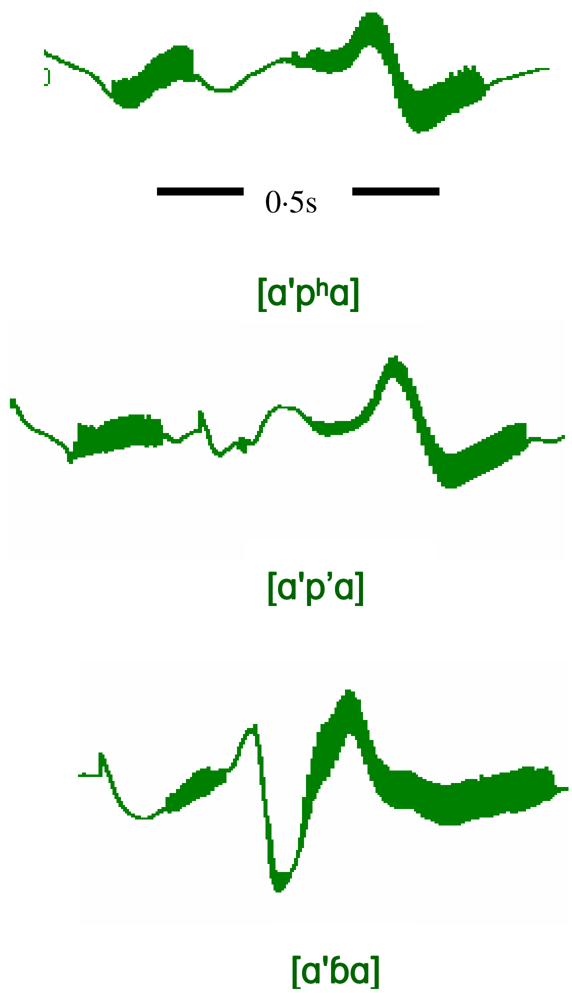

Figure 5
contrastive laryngeal gestures

Phonetic Dimensions in the Voice Clinic

Extracts are shown from the case notes for two patients,

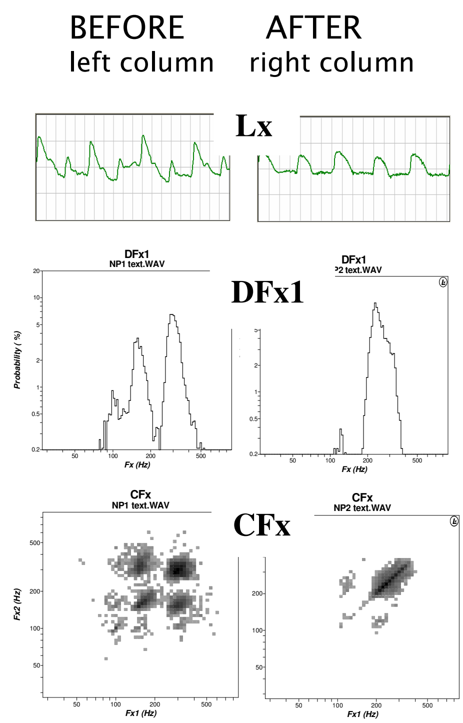

Figure 6 Surgical intervention

Figure 6 relates to remedial surgery performed by Julian McGlashan FRCS for a woman with vocal fold scarring. Recordings of “Arthur the Rat” are routinely made before and after surgical intervention and the top laryngograph waveforms are taken from the vowel [ɑ] in the word Arthur. The DFx1 distribution on the left shows how the vocal fold vibrational irregularity affects running speech with three modes of vibration before surgery and only one, at a normal frequency, after treatment. The pre-treatment voice is dominated by an atypically high fundamental frequency and also by extreme roughness. The CFx scatterplots are produced by the simple cross comparison of successive vocal fold closures for the whole of each recording, The normal voice has a well defined simple diagonal line corresponding to the small frequency changes that ordinarily occur between cycles and this is now approximated in the post treatment condition shown in the right CFx plot. These analyses are based on the use of the perceptual correlates of pitch that are discussed above but, since they are based on physically defined aspects of voice, they also make it possible to derive quantitative measures of improvement in voicing regularity – that have now been shown in other work to be linked to established clinical perceptual scales (GRBAS).

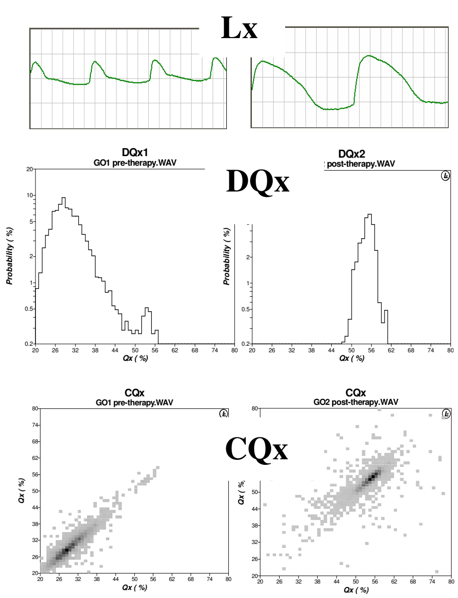

Figure 7 Speech Therapy

Figure 7 comes from work by Lesley Cavalli MRCSLT and shows the results of quite a brief therapy session with a young male puberphonic adult. The Lx waveforms now are both regular but prior to treatment there is a tendency to breathy voice and, of course, a tendency to produce an abnormally high pitch. DQx is a distribution of the closed phase percentages in the whole of the voice recording and is a simple way of showing the range of control available to a speaker for this important aspect of voice quality. The scale is from 20% to 80% in the analyses and, on the left, the mean Qx value is only ~20% whilst after therapy it is ~55%. This change in closed phase control has been accompanied by a marked change in mean frequency of phonation. It is also accompanied by a small change in the regularity cycle to cycle closure regularity , CQx, and a large change in CFx.

References

Fourcin,A & Abberton,E (2007) Hearing and Phonetic Criteria in Voice Measurement: Clinical Applications. *Logopedics Phoniatrics Vocology LPV*, In Press.

Lindsey,G, Hayward,K & Haruna,A. (1992) Hausa Glottalic Consonants: a Laryngographic Study. *Bulletin, SOAS, University of London,* vol. LV Part 3, pp. 511-527.

Heselwood, B (1996) Glottal States and Emphasis in Baghdadi and Cairene Arabic. *Paper 53, Centre for Middle Eastern & Islamic Studies, Durham University*, pp. 20-44.