

Speech Pattern Element Analysis and Display – “Lx Speech Studio”

Evelyn Abberton^{1,2}, Adrian Fourcin^{1,2}, Xinghui Hu²
Colin Bootle² and David Miller²

Department of Phonetics and Linguistics University College London¹
Laryngograph Ltd²
evelyn@phonetics.ucl.ac.uk lx@laryngograph.com

Abstract

Both teaching and research in Speech Sciences differ radically from comparable work in the physical sciences because of the need to link different levels of representation – and analysis. Spectrographic and waveform presentations are useful first tools but the vital phonetic information which they encapsulate is not readily accessible. The present work is aimed at meeting three speech science teaching objectives:–

- to show how quite simple phonetically relevant features can be derived in real-time from combinations of acoustic and physiological data
- provide a basic PC workstation giving a platform for the analysis, display and measurement of single and combined speech pattern element sets which are directly related to: pitch; loudness; regularity; friction; and timbre
- discuss particular examples (taken from English, French, Chinese, Sekgalagadi and speech pathology in the demonstrations).

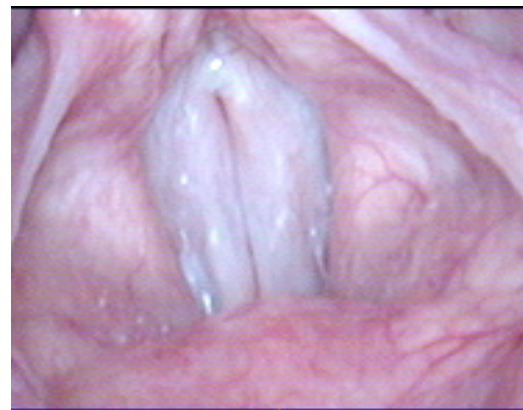
Special attention is given to establishing links between auditory, articulatory and physical levels of description and to provide for normalisation in display and quantification.

A Physiological aspect of Voice

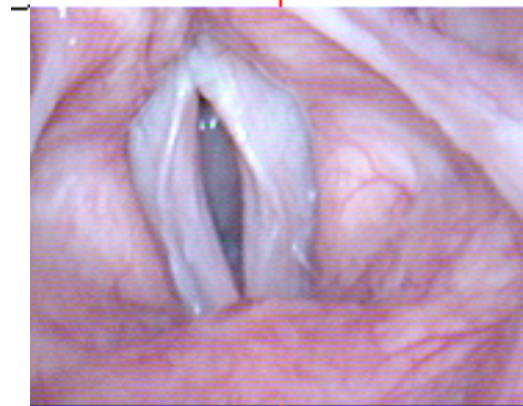
In all languages voiced sounds carry the greatest functional loads. A clear understanding of voice production is essential to all phonetic discussions of voice and voiceless activity. There is no good substitute for hands-on experience and modern PC processing and simple hardware now make this quite feasible.

The Laryngograph is an especially valuable tool for both teaching and research in the Phonetics Laboratory and in the Voice Clinic. Its utility, however, can only be properly appreciated, when it is used with a clear recognition of what it can – and cannot – provide.

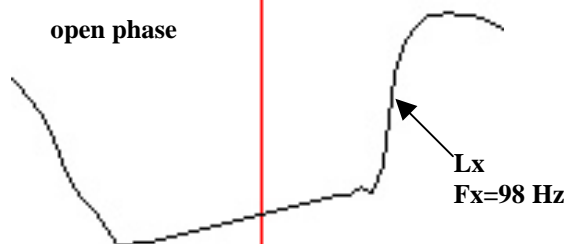
The two figures on the right show particular synchronous instances of the relations between the Laryngograph waveform, Lx, and views of the vibrating vocal folds. Figure 1 shows the vocal folds almost at their greatest contact in a stroboscopic view which has been triggered by the Lx waveform itself – the trigger instant is shown by the vertical line. Figure 2 shows the maximum opening of the speaker’s vocal folds with an evident, sustaining mucosal wave and a new trigger.



contact phase



open phase



Figures 1 & 2 Laryngograph triggered strobe views
Fig. 1 above shows the near peak of vocal fold
contact and 2, below, the middle of the open phase.

Pattern Elements in Intonation Contrasts

The set of five graphs on the right, at the top of the page, present different, but complementary aspects of two English intonation patterns, a fall and a rise. The broad band spectrogram at the top gives an excellent indication of total durations and spectral structure and contains the details of vocal fold frequency in the fine spacing between its vertical striations. The two waveforms, for the acoustic speech waveform, Sp, and the corresponding Laryngograph waveform, Lx, similarly contain detailed information which is of great potential perceptual importance. For both of these examples, however, it is not always obvious to visual inspection how this information is organised and how it varies coherently in the utterances.

The Qx traces immediately on the right give an indication of voice quality variation. They have been derived directly from the Lx waveform, by plotting the ratio of the width of the Lx contact phase (for each individual larynx cycle at 70% of its peak value) against time. For the normal fall there is typically a reduction and for the rise an increase in closed phase quotient and these effects are shown quite clearly. The speaker's main auditory monitoring is directed towards pitch control, however, and this is very clear for normal voice in the Fx traces, although Fx here is measured without smoothing every larynx period.

The last set of traces shows the result of combining two perceptually important physical speech pattern elements, intensity [Ax] and frequency [Fx], so as to obtain a representation which is of greater phonetic value. Once more there is no period to period smoothing and both the broad structures of intonation control are readily visible together with micro-structures which can be of phonetically contrastive importance.

Creaky voice is an especially important component of contrastive voice quality and the use of period by period processing makes it feasible to show its presence, and nature. The set of three traces on the right are for an English creaky voice falling intonation pattern.

Qx, at the top, give a clear representation of the way in which, midway through the utterance, successive vocal fold periods vary in contact duration. This effect is just as important as the variability in period [instantaneous vocal fold frequency] which is basic to the classic description of "diplophonic" pitch which is given to creaky voice. The middle trace, for Fx, shows this pitch related frequency ripple on the main intonation contour. For both the Fx and the Qx traces the final characteristic rather violent break-up is also important to the perception of this voice quality. The last trace of Ax and Fx to give another view of this voice characteristic which provides a way of assessing the joint occurrence of pitch and loudness related physical components – in real-time.

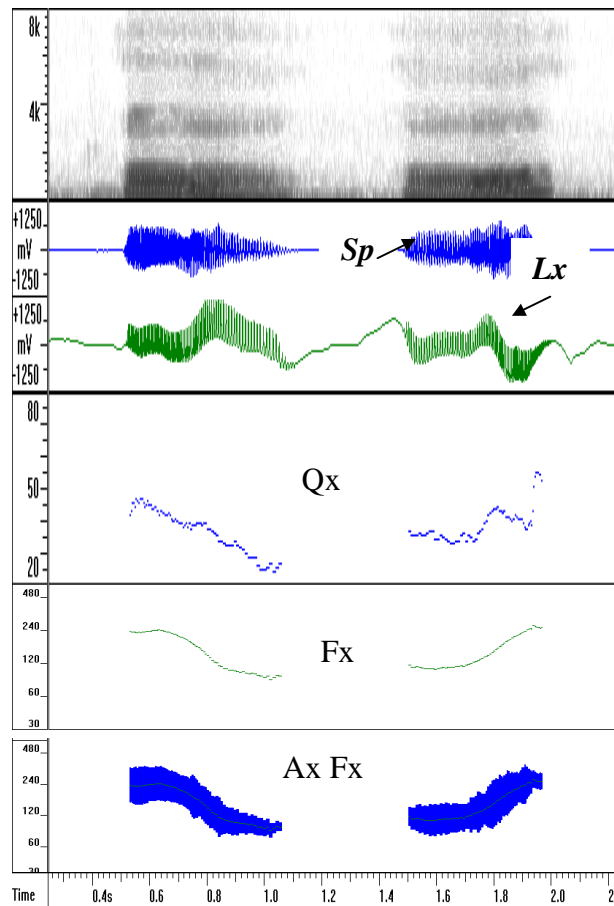


Figure 3 Fall and Rise , elements and patterns
 Qx % of Tx period; Fx in Hz

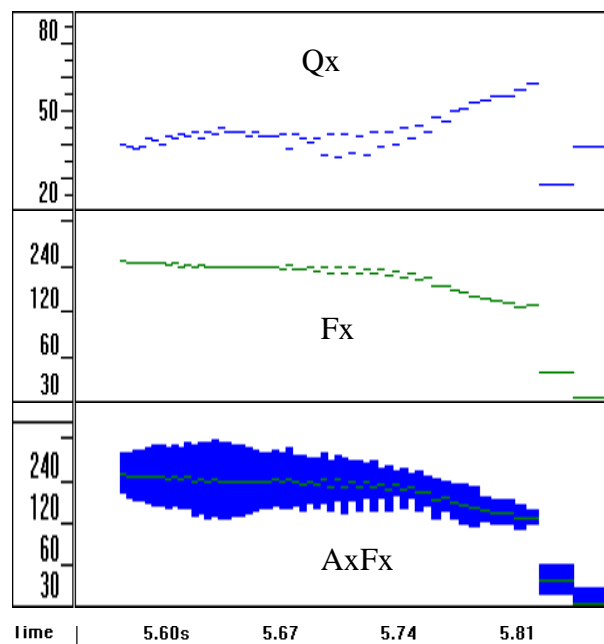


Figure 4 Aspects of creaky voice

Plosives, Nasals, Fricatives and Affricates

The eight panels on the right each contain an example of the condensed application of the sets of analyses shown on the preceding page in figures 3 and 4. The detailed information which they make available is essential to an in depth understanding of the nature of particular speech contrasts but these data families are unsuitable for the purposes of teaching and real-time therapy. Concentrating attention only on the main auditory pattern aspects of the utterances provides a better immediate insight into phonetically important features.

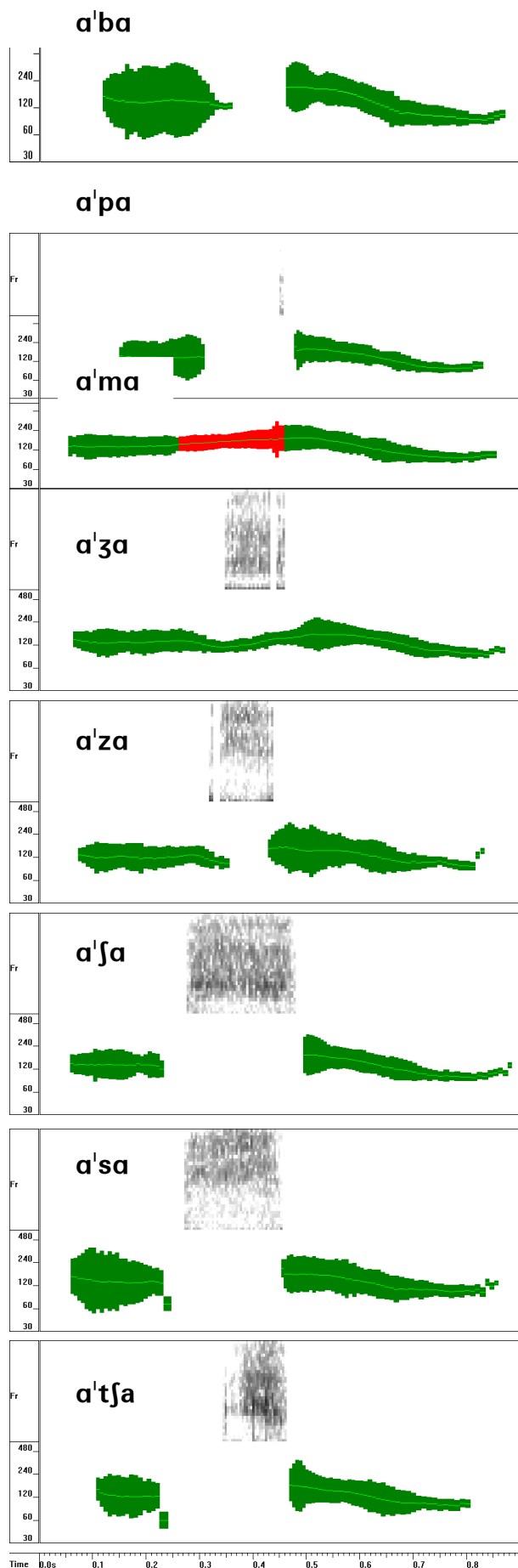
The utterance [a'ba] in the top panel shows the components of pitch and loudness represented against time by the position of the centre of each trace vertically for instantaneous voice frequency and by its vertical width for peak acoustic amplitude. It can now be seen easily that although the stress is on the final vowel, the first sound is of greater amplitude. The transition into the closure phase is accompanied by a gradual reduction in voicing amplitude, of a quite different type from that found at the final termination of voicing.

[a'pa] is, in contrast, with an evident plosive release, different voicing onset and different temporal and intensity organisation of the initial vowel. [a'ma] in the ordinary use of this system has the nasal component in red as opposed to green for voice, but the pattern is quite distinct. It is interesting to see the small velic closure "click" at the end of the nasal. Although simple these pattern convey a great deal of phonetically interesting information.

[a'za] shows the simultaneous presence of voicing and frication – with a typical reduction in voicing intensity during the frication. In the next example [a'za] voicing does stop during the frication although the percept of a voiced fricative is clear.

The difference between the energy – frequency distributions is clearly shown both in these two examples and for the voiceless cognate fricatives [a'ʃa] and [a'za]. In each case, for these voiceless fricatives the prior cessation of voicing before the establishment of frication is clear, as is the abrupt onset of voicing after frication.

For the last item, in this small set of examples, a combination of some of the features of the previous utterances is shown in the pattern display for [a'tʃa]. An abrupt cessation of voicing in the initial vowel, accompanied by the drop in Fx which can accompany the preparation for a front articulated voiceless fricative, precedes the marked closure interval which characterises a voiceless stop. The release of the stop is followed by a clear interval of aspiration which leads into the brief establishment of the palato-alveolar fricative. Finally there is a clear and sudden onset of voicing for the transition from the voiceless setting of the vocal folds to the establishment of normal voicing.



Longer term pattern structures

The overall composition of conversation and discourse is dependent not only on the nature of the individual phonetic elements of which the sounds of speech are made but also on quite complex layers of successive levels of phonetic control. Structures of the gross temporal organisation of voicing, intonation, pausing; of the finer details of their use in contrastive prosody and in the delineation of frication and nasality are amongst the important targets for quantification.

Only a few examples are given here, but even for these some radically important issues have to be addressed. Perhaps the most important concerns the degree of reliability which can be associated with an analysis and the sample of speech on which it is based. The top figure on the right is for the distribution of vocal fold periods, measured with 1MHz time sampling. These "normal" Dx distributions are based on an 18 minute data sample (produced to reduce fatigue by concatenating nine 2m samples). For this total duration of data stationarity is achieved not only for the representation of tone groups but also for measures of larynx frequency range and regularity. When sample lengths of 2m are used the measures of range and regularity are still defined to within 2 to 3% of those for the long samples – but the detailed shapes of the distributions is different because the tone group population is unrepresentative.

The second example uses the Dx analysis for the examination of a pathological conversational voice sample of 2m duration. Dx1, which includes every larynx period, is evidently very different from the normal and this can be understood from a stroboscopic examination of the speaker's damaged vocal folds. The second order, Dx2, distribution gives a clear indication of the very poor auditory nature of this voice since its modal peaks show where, in Fx, the speech is associated with a clear voice pitch. Continuous speech phonetogram analysis is a useful extension of this type of connected speech (Sp & Lx) representation since it gives the range of intensities and larynx frequencies which typify a speaker's output. A further development of this approach which is of special phonetic interest is shown in the Qx distribution. Here, the contact phase ratios shown in figures 3 and 4 have been applied to a long sample of the Lx signal (the 18m sample has been used but the results are essentially identical for this parameter when 2m samples are used). The speaker is a young woman of ~35y and she shows a reduction in contact phase ratio, which is associated with a breathy voice quality, which progresses with increase in vocal fold frequency. This is a typical feature for the speech of normal English women (and may be more general). The pathological voice speaker has much lower Qx values and a three island structure dominated by the higher frequency mode. The final histogram shows a distribution of the total times allotted to frication, voicing, silence and nasality in the very long sample but the proportions are the same even for 2m data samples from the same speaker.

