REAL-TIME SPEECH PATTERN ELEMENT DISPLAYS FOR INTERACTIVE THERAPY[1]

Evelyn Abberton, Xinghui Hu, Adrian Fourcin

Centre for Speech and Hearing
University College London
4, Stephenson Way
London NW1 2HE

New developments are presented and discussed for the interactive display and separate measurement of frication and nasality in addition to voice quality. Different clinical problems require appropriate combinations of speech pattern elements support different approaches and these are supported by appropriate combinations of speech pattern elements. Particular use has been made of the laryngograph signal so that the displays can be given a sense of immediacy and are highly accurate.

Introduction - the basic problem

Assessment of the intelligibility of a client's speech before, during and at the end of therapy may make use of qualitative (perceptual) and quantitative (physical) descriptors and measures. The latter are of particular value, not only in demonstrating change to the therapist and client, but also in documenting efficacy of treatment - in communicating the evidence! Unity of approach is achieved if the same descriptors can be used in interactive visual feedback during therapy as well as in assessment. However, presentation and quantification of the communicatively relevant phonetic features in an utterance is no simple matter. The transformation between different levels of representation - articulatory-acoustic-auditory-linguistic - is complex and there is no one-to-one correlation between units at any two levels. A particular complexity arises from the need to allow for perceptual normalisation - to map the differing physical, acoustic patterns of speech (and their visual representations) from different speakers - women, men and children - onto the same perceptually relevant linguistic phonetic elements.

Bases for a direct solution

Any visual representation of speech for interactive therapy must have normalisation built into it to allow therapist and client who may have very different sized larynxes and vocal tracts to produce displays that are concordant; a further requirement is that the display must be relatively simple, and correspond to our auditory perception of speech in what users appreciate as a direct and natural way. A well known example is provided by a logarithmic display of speech fundamental frequency against time which corresponds to our perception of pitch changes in speech. If the analysis and presentation are carried out on a cycle-by-cycle basis an additional visual correlate, of rough or smooth phonation quality, is also available. "Pitch" displays of this sort are now well known in speech and language therapy, especially in the Voice Clinic, and show the value of the extraction of linguistically relevant elements from the complex acoustic speech signal rather than presenting "raw" speech as a sound pressure waveform or a spectrogram. (Abberton and Fourcin 1997).

If another element, the amplitude of speech, is taken into account then the links between the segmental composition (voiced and voiceless) of an utterance and its prosodic form are made even clearer in a visual display of "loudness-modulated pitch". The presentation is not stylised

---

but is easily related to perceived rhythm and intonation patterns, since the fundamental frequency contour is thickened for louder voiced portions of the utterance, typically corresponding to syllable centres and major pitch changes.

An advantage of the speech pattern element approach is that it allows the therapist to choose which individual feature or combination of features to work with depending on the problem in hand. The loudness-modulated pitch display, for example, can also have a visually normalised presentation of fricative elements to represent plosive bursts or the notoriously difficult sibilant fricative consonants.

Different clinical problems require different combinations of elements. Lack of frication is a contributory factor in perceived nasality, but the essential aspect of hyper-nasality - excessive nasal air escape - can also be monitored and displayed in a visually simple way related to the overall prosodic structure of an utterance - however short that utterance may be. The fundamental frequency contour is coloured distinctively if nasality is present at any instant.

Examples of particular approaches
Clinical use of quantitative physical methods in speech analysis and display demands robust, reliable real-time processing. We make particular use of the laryngograph output as well as the acoustic signal, together with a nasal accelerometer for the nasality display. The satisfies the requirements of immediacy and accuracy, and provides presentations that are not only acoustically robust but perceptually and linguistically relevant: essential speech contrastive features can be shown and taught as well as the detailed pattern forms important for naturalness. Both perception and production can be worked on and the link between auditory target and articulatory means forged or reinforced.

The figures on the following three pages show particular examples of the application of this approach: PC based displays of combinations of the speech pattern elements relating to voice quality (here with special reference to regularity of phonation), pitch contour control, loudness, frication and nasality.

- ♦ Figure 1 illustrates an aspect of phonation regularity in a VCV context
- ♦ Figure 2 shows the pattern elements of pitch and loudness in irregularity analysis
- ♦ Figure 3 provides an indication of lexical tone control in contrast with
- ♦ Figure 4 which uses the pattern displays for English contrastive intonation
- ♦ Figure 5 uses the displays to give an example of frication quality and differences
- ♦ Figure 6 combines frication with nasality in the display

In all of these examples only black and white printing is used but in practice colour makes the displays much easier to use and to explain to the client. In all cases, the displays are linked to quantitative measurement to assist in the assessment of progress.

Good and bad voice qualities with falling intonation in two CVC contexts: interactive pattern and spectrographic displays
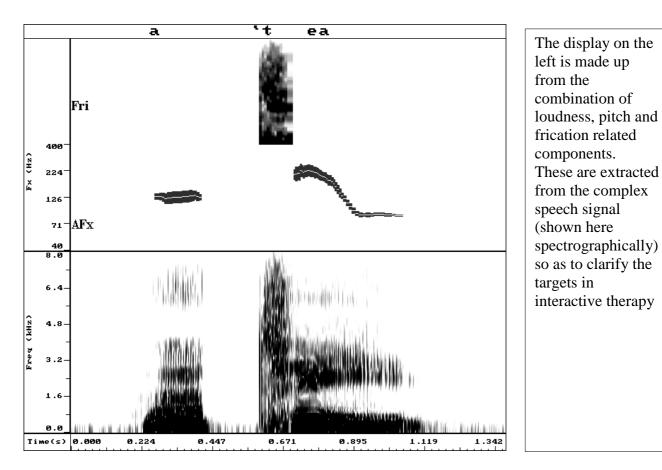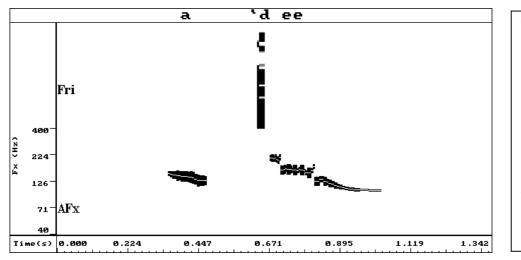

The display on the left is made up from the combination of loudness, pitch and frication related components. These are extracted from the complex speech signal (shown here spectrographically) so as to clarify the targets in interactive therapy

Figure 1 *Pattern Element above, Spectrogram below; good voice quality*

The voiced intervocalic consonant here has a smaller burst and a longer closure duration; simple patterns make these easy to see – as well as the evident creak.

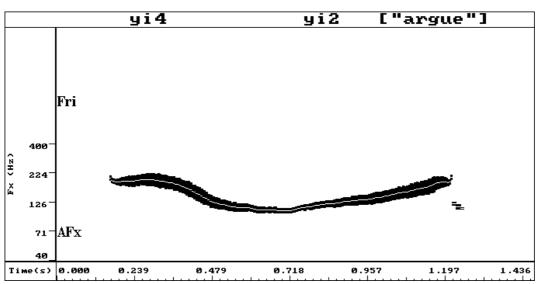**Figure 2** *Pattern Element Interactive Display; broken creak voice quality*

Figure 3. The two pattern elements of loudness and pitch against time for a citation form two syllable Chinese word are shown. The utterance relates to lexical meaning rather than emotional expression; the change in fundamental frequency is slow and the dynamic range of intensity is small.
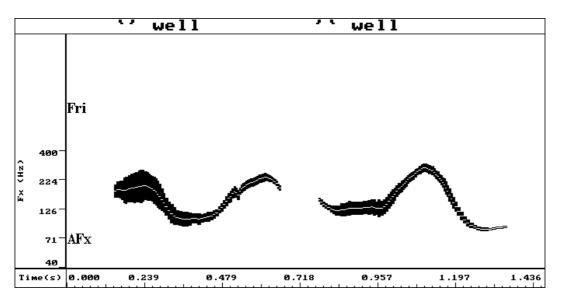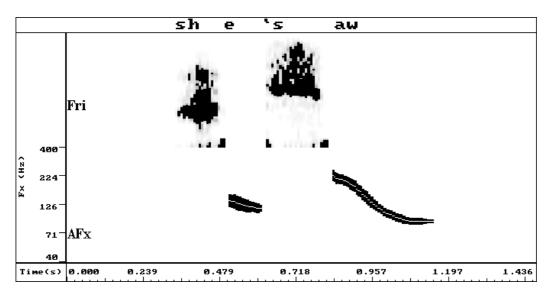

Figure 4 These two pattern displays are for contrastive fall-rise and rise-fall intonation forms in English. Since the aim here is expressive the pitch and loudness element changes are larger and more rapid than in the Chinese example.

Pattern Element examples of Frication and Nasality

Figure 5 *The use of a perceptual model as the structural basis of the interactive displays makes it possible not only to simplify the presentation of therapy training material for voice work but also to use normalisation so as to simplify in an acceptable way the introduction of contrastive fricative information.*
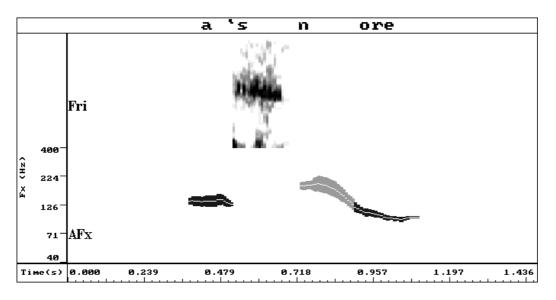

Figure 6 *Here the additional pattern element of nasality has been introduced. The figure shows the presence of nasality in the time course of the utterance by a lighter shading - coarticulation has made this shading spread into the final vowel from the nasal consonant.*

In Conclusion

This very brief overview has concentrated on a limited number of the basic dimensions of speech contrasts. Many others are of daily clinical importance - voice quality and timbre for instance - but this type of display is potentially available for all. Measurement can also be tied in to each of the categories. In summary, a list of possible practical advantages of pattern element processing might include:-

♦ the clarification that comes from the reduction of a complex display to one which is adapted to a well defined therapy task

- ♦ the real-time feedback of essential aspects of production helps in the training of perception and can substantially improve self-monitoring
- ♦ the provision of outcome measures is directly relatable to therapeutic work when quantification uses the same measures as the therapy itself

References

Abberton, E. and Fourcin, A., 1997, Electrolaryngography. In Instrumental Clinical Phonetics, edited by Martin J. Ball and Chris Code (London: Whurr), 119-148.

Acknowledgement