ON COMPARING ACOUSTIC AND ELECTROPALATOGRAPHIC REPRESENTATIONS OF ENGLISH OBSTRUENTS

Helen Fraser (1) & Gerry Docherty (2)

(1) Department of Linguistics, University of New England, NSW, Australia.
(2) Department of Speech, University of Newcastle-upon-Tyne.

1. INTRODUCTION

In this paper[1] we report on a study in which we have been performing a detailed comparison of acoustic and electropalatographic (EPG) representations of the production of some obstruents in English. The availability of a robust multi-channel analysis system [2] has meant that for the first time it is relatively straightforward to carry out a comparison of EPG and acoustic (and other) representations of speech. Most previous work on articulatory-to-acoustic mapping has (understandably) been focused on mapping overall vocal tract shapes on to vowel quality (e.g [3], [4]) and on analog models of stop (e.g. [5]) and fricative (e.g. [6]) production.

By its very nature EPG in itself allows analysis of only a very limited aspect of the activity of the vocal tract in the production of speech. An EPG/acoustic study, however, allows us to address certain questions pertaining to the relationship between linguo-palatal contact patterns and the corresponding speech waveform; questions such as to what extent do we find occlusion during (acoustically-defined) stops; or close approximation during (acoustically-defined) fricatives? To what extent is the evolution of a lingual gesture reflected in the acoustic signal (e.g. as vowel transitions)?

In this paper, we briefly describe the material which we have been analysing, and present some results and (in outline) some theoretical and methodological questions which we think our data raises for phonetic theory and practice.

---

[1]Earlier versions of this paper have been presented at the Leeds Experimental Phonetics Symposium, the Colloquium of the British Association of Academic Phoneticians, and the 2nd Speech Production Seminar. A fuller account of this work has been to submitted to Journal of Phonetics [1]. The second author wishes to acknowledge support from the University of Newcastle-upon-Tyne Small Grants Research Sub-Committee.

2. METHOD

The instrumentation used in our study[2] is the multi-channel system located in the Speech Research Lab, Reading University, and described in comprehensive detail in [2]. This system provides simultaneous digitisation of six channels; speech signal, laryngograph signal, oral airflow, nasal airflow, oral pressure, and the EPG data. In the present study oral air pressure was not recorded. The output from the contacts on the electropalate was sampled at 200 Hz and the speech waveform at 20 KHz. The digitised articulatory, aerodynamic and acoustic parameters were aligned on the screen of a PC for subsequent analysis.

A single speaker (FG) was recorded producing four repetitions of the following two sentences.

'Stats by such a student will surely be trashed after tests';

'The lad showed us a strategy for trimming his moustache surreptitiously'.

These sentences were designed to contain a large number of alveolar and palato-alveolar consonants and consonant sequences in (as often as possible) a non-high vowel context. There is no doubt that in one sense the amount of data is very small. However, given the data explosion which results from a 5 msec update rate of the EPG contact patterns it was felt that in these initial stages of our study it would be advisable to perform a detailed analysis of a restricted data set rather than to overstretch the database size at the risk of a less detailed analysis.

Analysis of the data was carried out in the following way. The EPG and acoustic representations were examined independently of one another. Significant events in the acoustic waveform were identified (e.g. onset and offset of stop closure, of fricatives, of intervals of voicing, of vowels and other approximants) and a segmentation was performed. The criteria used for segmentation were as follows.

V-stop boundary: This was marked at the point at which there was a *coincidence of* amplitude drop and change in the form of the waveform to be more sinusoidal.

Stop-V boundary: the stop release and aspiration (if any) were collapsed into a single acoustic category delimited by the onset of noise (corresponding to the burst) following a stop closure

[2]The data reported in this paper was recorded as part of the ACCOR project database. We would like to thank Bill Hardcastle, Fiona Gibbon, Katerina Nikolaidis and Linda Shockey for their assistance.

and the first peak of the laryngographic signal corresponding to the vowel.

Stop-fricative boundary: marked at the point where the noise in the signal corresponding to the fricative began.

Fricative-V/V-fricative boundary: marked at the end/beginning of the interval of noise in the signal corresponding to the fricative.

These, of course are not the only acoustic segmentation criteria which could have been applied. However, they have all been used commonly in the past in acoustic investigations of phonetic realisation based on analysis of the time waveform. And of course, by adopting this procedure, we have followed the common practice of choosing what appears to be a reasonable set of criteria in the particular case and applying them consistently.

In analysing the EPG waveform, we wrote a detailed shorthand description of the contact patterns through time. The major feature which we focused on in our description of the EPG patterns was the minimum width of the open channel in rows 1-4 of the palate (i.e. in the alveolar zone) at any given time. In a pilot study this had been found to be a reasonable indicator of the tip-blade gesture. In addition, we recorded variation in the total number of contacted electrodes on the entire palate, and in the total number of contacted electrodes on each of rows 1-4 of the palate individually.

3. RESULTS

Figure 1 shows examples of our analysed data.

Our overall finding is that there is a surprising degree of inconsistency in the correlation of EPG and acoustic representations. The most reliable feature is that the loss of the total occlusion for a stop normally coincides closely with the onset of the release burst in the acoustic signal. However, in much of our data, it is difficult to identify precise lingual gestures or patterns of contact which consistently correlate with features in the acoustic signal. This finding does not appear to be an artefact of the technique used or of the acoustic segmentation criteria which have been adopted (i.e. it is not the case that a constant displacement factor or a different set of segmentation criteria would lead to a better match). The EPG data paints a picture of gradually changing lingual configuration, in sharp contrast to the acoustic signal which is characterised by abrupt changes from one particular signal type to another (e.g. fricative-to-vowel, stop-to-fricative, etc.).

The following more specific details of the data illustrate the above general finding (for further details, see [1]).

(a) acoustic closure phases do not always correspond to complete occlusion in the EPG data (e.g. Fig 1a frames 206ff, Fig 1(b) frame 438).

(b) there is an unreliable correlation between channel width and occurrence of noisy acoustic signal. In particular the points at which the acoustic signal changes from fricative to resonant and vice versa do not always correspond to any change in degree of stricture as indexed by our EPG parameters (e.g. Fig 1a frames 170 and 188, Fig 1b frames 469 and 497).

(c) a further reflection of the aforementioned unreliable correlation is to be found in the instances when we find a surprisingly long interval with very little change if any being registered in the EPG contact patterns corresponding to several unmistakably significant changes in the characteristics of the acoustic signal (e.g. Fig 1a ff218-226, Fig 1c ff916-948).

4. DISCUSSION

On the face of it, our data look a little problematic. Any expectations we might have entertained about events matching up (based on conventional phonetic theory) have not been met with any degree of consistency, and there does not appear to be any 'simple' explanation based on measurement procedures etc. It does not on first examination seem that there is a simple convertibility between EPG and acoustic representations of speech.

However, a little thought suggests that it is the assumption that the two representations should match up that is problematic, rather than the fact that they do not. In other words, there is a strong sense in which these findings are not too surprising. It is well known that the vocal organs are in quasi-constant motion during the production of an utterance, and of course, the acoustic signal is the product of more than just the width of the linguo-palatal channel.

We doubt that anyone would seriously dispute this statement in itself, but it is worth bearing in mind that its implications -- non-convertibility of representations of speech -- are not always followed as thoroughly as they might be; *viz* the fact that it is practically a truism in Linguistics that fricatives are produced with a narrow channel. Our experiment would therefore seem to be useful in highlighting a point of methodological significance -- as a warning that individual techniques should not be used in isolation, but in combination, at least until far more research has made their convertibility more obvious.

This important but fairly simple methodological point is related to a more complex theoretical issue arising from our data which we consider worthy of further attention, that of the description of phonetic categories. The problem of mismatching categories arises with respect to our insistence on using the same vocabulary to describe the two representations. Terms like STOP, FRICATIVE, VOCOID, ASPIRATION are commonly used in both acoustic and EPG studies of speech. The problems raised by our analysis remind us that acoustic and articulatory representations are not simply alternative descriptions of a single unitary object. The acoustic description represents the outcome of a combination of complex articulatory processes -- of which an EPG description represents only one aspect. Certain intersections of these (generally) gradual articulatory processes produce the (generally) discrete acoustic categories. The different descriptions of speech create different categories and require different vocabularies.

For example, our data provide graphic illustration of the fact that the notion of degree of stricture (which is the foundation of one of the major descriptive phonetic dimensions, not to mention of the features underlying many phonological feature systems) is not a very powerful one when applied to EPG data. We have found a good deal of overlap in the contact patterns corresponding to fricatives and stops, and fricatives and vowels. From a theoretical point of view, as phoneticians we have to think hard about the descriptive categories which we commonly use to describe the differences between (for example) fricatives and vocoids. The commonly used terms -- stop, fricative, vocoid -- may well be inappropriate for looking at EPG data, and possibly not even appropriate for acoustic data. They belong at the abstract level of a phonetic transcription, but they could be positively damaging if unthinkingly applied to data such as that which we have been looking at.

In summary, whilst at first glance the data appears to be problematic (we do not find as high a degree of correspondence between the two representations as we might have expected), instead, it may be that it is the theory/vocabulary that we are using to frame our expectations and observations that is inadequate. Our result suggest that it is always necessary to constrain the use of phonetic vocabulary in relation to the aims, methods and contexts of any particular experimental situation.

5. CONCLUDING COMMENTS

So much of what we know about speech is based on an acoustic (normally spectral) representation (due to the dominant position of spectrography over the years). The ability to combine acoustic and articulatory (such as EPG) information into a single analysis promises to give us a new perspective on speech production. But what our study shows is that as ever more sophisticated instrumentation is brought to bear on the study of speech, a good deal of effort must be invested to ensure that the theory (which after all is the frame within which the instrumental observations are interpreted) is allowed to evolve in tandem. We hope that our continuing study of the fine detail of acoustic, aerodynamic and articulatory characteristics of obstruent production will allow some of these theoretical questions to be addressed.

6. REFERENCES

[1] Docherty, G.J. & Fraser, H. (1990) 'On the relationship of acoustic and electropalatographic representations of speech' Paper submitted to J. Phonetics.

[2] Hardcastle, W.J., Jones, W., Knight, C., Trudgeon, A., & Calder, G. (1989) 'New developments in electropalatography: a state of the art report' *Clinical Linguistics & Phonetics*, 3, 59-69.

[3] Fant, G. (1960) *Acoustic Theory of Speech Production* Mouton: The Hague

[4] Ladefoged, P., & Harshman, R. (1979) 'Formant frequencies and movements of the tongue.' in Lindblom, B., & Ohman, S. (eds) *Frontiers of Speech Communication Research*, Academic Press: London, pp. 25-34

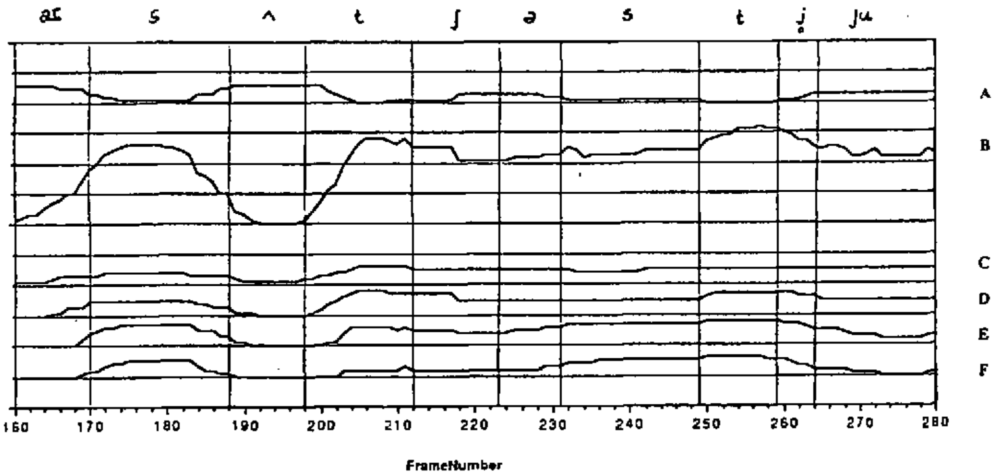
[5] Rothenburg, M. (1973) *The Breath-Stream dynamics of Simple-Released-Plosive Production*. Bibliotecha Phonetica, No. 6, Karger: Basel.

[6] Shadle, C. (1990) *'Articulatory-acoustic relationships in fricative consonants.'* in Hardcastle, W.J. & Marchal, A. (eds) *Speech Production and Speech Modelling*, Kluwer:Dordrecht, pp 187-209

FIGURE 1: Three examples of the acoustic/EPG representations of the data analysed. The transcription at the top of each example (and the accompanying segmentation lines) is based on the segmentation of the time waveform (absent from the figures due to space constraints). TRACE A shows the minimum channel width in the anterior 4 rows of the palate measured in number of uncontacted electrodes on a scale of 0-6. TRACE B shows the total number of contacted electrodes on the entire artificial palate on scale of 0-50. TRACES C-F show the total number of contacted electrodes in rows 1, 2, 3, and 4 respectively of the front of the artificial palate on a scale of 0-8.

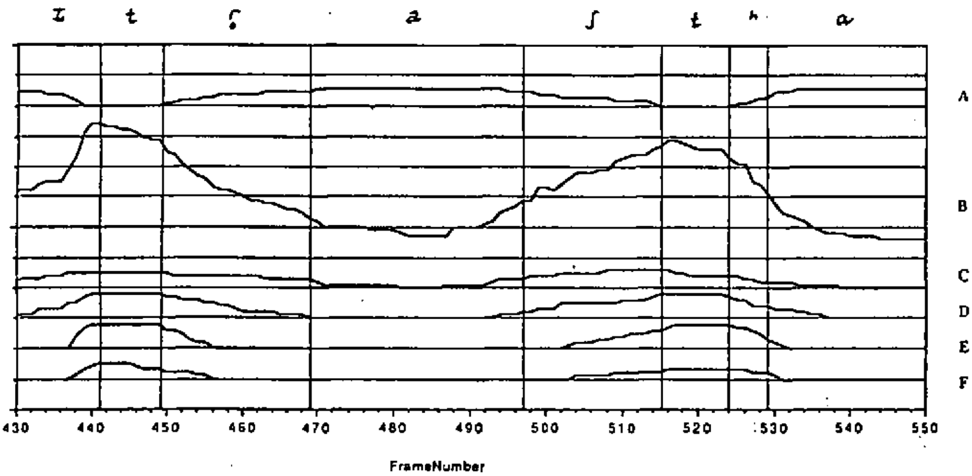
ACOUSTIC AND EPG REPRESENTATION OF OBSTRUENTS

Sentence 1, Repetition 3



(A)

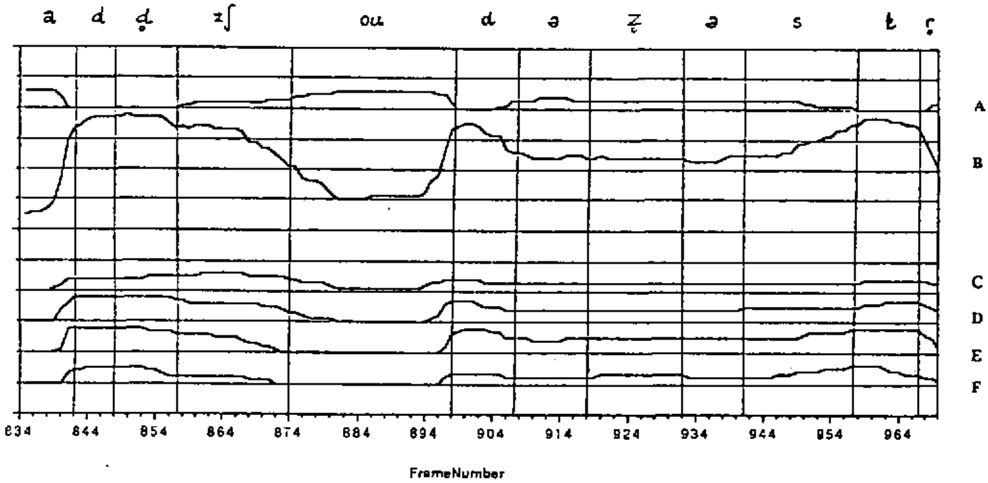
Sentence 1, Repetition 3



(B)

ACOUSTIC AND EPG REPRESENTATION OF OBSTRUENTS

Sentence 2, Repetition 1



(C)