ANALYSIS OF FRICATIVES USING MULTIPLE CENTRES OF GRAVITY

Alan A. Wrench
CSTR, University of Edinburgh, Edinburgh EH1 1HN, UK.

ABSTRACT
An algorithm for describing speech spectra in terms of multiple centres of gravity is compared to traditional methods for parameterising fricative spectra. LPC peak-picking analysis and single centre of gravity measures are compared with Multiple Centroid Analysis (MCA) and the strengths and weaknesses of this newer approach are discussed.

INTRODUCTION

Traditional Spectral Parameters
It has been known for many years that fricative spectra exhibit formants and anti-formants. It seems likely from the wealth of research done on vowel perception that formants play an important part in our perception of fricative quality and accordingly, fricative formant frequencies have been studied in an effort to correlate them with articulatory parameters. Formant estimates have been recorded manually by visually peak-picking the fricative spectrum [1][2]. Despite being time consuming and subjective, hand labelling Most researchers have held back from using automatic formant analysis procedures for parameterising fricatives. Heinz and Stevens [3] fitted pole-zero model by hand to spectra. However, it is well known that automatic estimation of pole-zero parameters cannot be solved directly and while there do exist useful automatic procedures [4] for estimating these parameters they have tended to be ignored by researchers in this field most likely because the methods are not implemented in speech analysis packages. Although the fricative spectrum contains zeros associated with the source constriction or cavities posterior to a source of frication, several researchers have used LPC (which assumes an all-pole model) to approximate the spectral contour and have either identified peaks by eye or applied automatic peak picking in order to identify the fricative formant structure [5]. Due to the varying number of formants (and anti-formants) it is not possible to identify an ideal order for the LPC analysis of voiceless fricatives. A low order can produce biased or unresolved formant estimates and poor estimation of bandwidth. A higher order reduces this bias but is susceptible to producing spurious peaks.

The lack of theoretical basis for using an all-pole model combined with practical problems outlined above has lead many researchers to use the simpler calculation of the centre of gravity as a "general property detector" [6] for parameterising fricative and other speech spectra. The underlying assumption of this approach to fricative analysis is that the spectrum can be modelled as a single normal distribution which may reflect the dominant front cavity formant. The mean (1st moment) of this distribution is termed the centroid and popularly referred to as the centre of gravity. Despite this model being underspecified for fricatives which exhibit two or more formants, the centre of gravity has often been used to measure relative changes between fricative productions [7][8]. By extending the parameters to include higher moments of this estimated normal distribution (corresponding to skewness and kurtosis), the ability of this model to discriminate between fricatives is enhanced [9][10].

A natural extension to this model to cope with fricative spectra containing more than one formant was proposed by Jassem [11]. To do this the spectrum was split into two or three partitions. He investigated several criteria for automatically determining the partitions but found that fixed partitions were sufficient. The purpose, however, was not to relate each centre of gravity to a formant but to use them as an abstract set of parameters which might be fed into a statistical classifier in order to distinguish between fricative phonemes.

Overview of MCA
Aware of benefits of centroid analysis to the estimation of fricative spectra, Crowe looked for a method of optimally fitting multiple centroids to a speech spectrum intending it as a means of estimating vowel formants. He successfully generalised the centre of gravity calculation using a global least-squares error criterion to determine the optimal partitioning of the spectrum and thus provided a principled method for determining multiple centroids of a single multi-modal spectral distribution [12].

In broad terms the algorithm works as follows. Taking, as an example, dual centroid analysis: It operates by evaluating centroids for every possible partitioning the spectrum into two and choosing the pair of centroids that result in best overall fit. For each possible position of the boundary, the centre of gravity is calculated for the part of the spectrum lying in each partition.

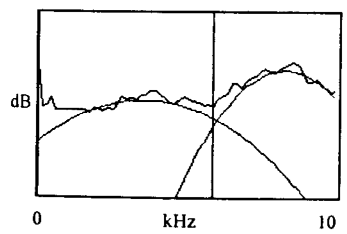

Figure 1. Optimum boundary indicated by vertical line and centres of gravity calculated for each partition (shown as the parabolic apex).

Each centroid is estimated as the frequency that gives the minimum squared error value. The two minimum error values are summed and stored for each possible boundary position. This process is repeated for all possible partitionings. The output of the analysis is the centroid pair corresponding to the partitioning with the lowest minimum error score.

The centroid can be thought of as fitting a Gaussian (normal) distribution to the power spectral distribution. The variance of this normal distribution can be thought of as an estimate of bandwidth and if the spectral distribution within a single partition contains a single formant then the centroid and associated variance represent the formant frequency and bandwidth. Multiple centroid analysis can be achieved more efficiently by placing constraints on how the spectrum may be partitioned and by bark scaling the spectrum prior to analysis. These measures improve the speed and accuracy of formant estimation when MCA is applied to vowel analysis [13].

Approach of the study
In this study we will take sustained examples of three voiceless fricatives and compare the results of analysing the speech using:

i) LPC analysis
ii) Single centre of gravity
iii) Multiple Centroid Analysis

Voiceless obstruents were chosen because they are known to be differentiated by their spectra alone.

The three speech segments were selected to contrast two allophones of /s/ (lip-rounded and non lip-rounded) with a non lip-rounded allophone of /ʃ/. These examples provide three distinct spectra which highlight the different behaviour of the three analysis methods.

METHODOLOGY
Recordings were made by a male speaker in an office environment using a Shure SM10A close-talking microphone and 16-bit soundcard sampling at 20kHz. Background noise levels were measured to be at least 25dB below the signal at all frequencies above 1khz.

The frame size used for all analyses was 6.4ms with a shift of 2 ms. In order to reduce the spectral variance, which can obscure the underlying resonant structure, a 100ms period of analysis was used. In the case of LPC analysis, the autocorrelation function was accumulated for each frame and averaged over a 100ms segment taken from the centre of the fricative. The LPC coefficients and associated spectrum were then calculated from this averaged function. Both the centre of gravity and MCA were based on time averaged spectra. An FFT was performed on each frame over the same 100ms portion of the fricative and the resulting power spectra were accumulated and averaged. The centre of gravity measure and MCA were calculated from the averaged power spectral distribution lying above 1000Hz. Note that the MCA algorithm was applied directly to this spectrum and Bark scaling was not employed.

RESULTS

Figure 2a), b) and c) show the LPC (order 4) and FFT log spectra for /ʃ/, /s̨/, and /s̨/ respectively, extracted from the nonsense words 'eeshee', 'eessee', and 'oossoo'. The centre of gravity is calculated from the linear power spectrum. In order to represent the 1st and 2nd moments the centre of gravity is shown as the log of the corresponding normal distribution. Figure 3a), b) and c) show the same FFT spectra with LPC (order 6) superimposed and two centres of gravityestimated using MCA (order 2). Figure 4 a), b) and c) show the same FFT spectra with LPC (order 8) and three centres of gravity.

Using the general rule of thumb that two poles are required for each peak with two extra poles to model the spectral gradient, the order of the LPC analysis was selected to try to equate the maximum number of peaks with the number of centroids shown in the same figure.

DISCUSSION

We can see that the two-centroid analysis in Figures 3b and 3c model the formant structure well. The peaks of the LPC spectrum by visual comparison are biased. Increasing the order of the LPC analysis from 6 to 8 redresses this discrepancy in performance. In figure 3a the lower centroid matches the principal resonance well but, with no clearly defined second formant the upper centroid does not correspond to any feature. LPC, by contrast, models the spectrum in figure 3a by a single peak.

In formant tracking of vowels it is generally true that a fixed number of formants will exist within a given frequency range and having a fixed number of centroids is an advantage in that it permits merged formants to be resolved. In the case of fricative analysis, where the number of peaks varies as the length and shape of the cavities from source to lips changes, this advantage turns to disadvantage. This is clearly demonstrated in figure 4. where in each spectrum two centroids are associated with a single peak.

It is possible that, as abstract acoustic parameters, the 1st and 2nd moments of a pair of centroids may correlate well

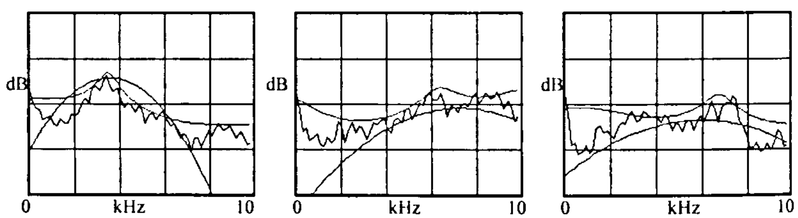

Figure 2. a) /ʃ/, b) /s̨/ and c) rounded /s̨/ 100ms segments analysed by LPC (order 4) and MCA (order 1) i.e. a single centre of gravity

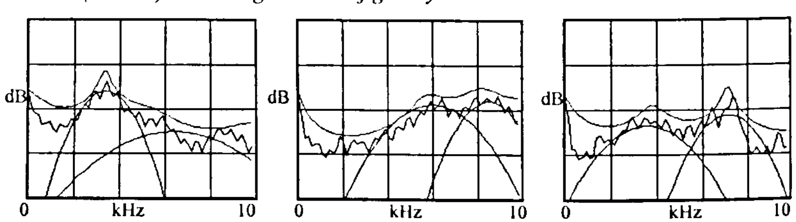

Figure 3. a) /ʃ/, b) /s̨/ and c) rounded /s̨/ 100ms segments analysed by LPC (order 6) and MCA (order 2) ie. optimised fit of two centres of gravity.

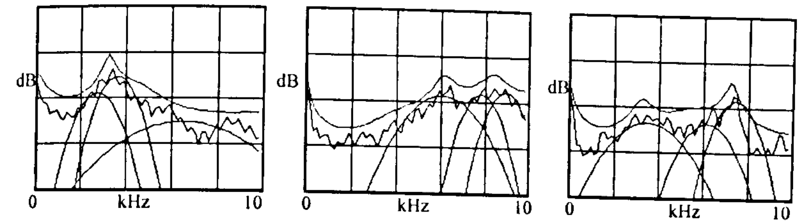

Figure 4. a) /ʃ/, b) /s̨/ and c) rounded /s̨/ 100ms segments analysed by LPC (order 8) and MCA (order 3) i.e. optimised fit of three centres of gravity

with articulatory parameters just as they have done for single centroid [9]. A rigourous comparative study is required to determine whether such parameters hold any advantage over the first 4 spectral moments of a traditional single distribution model or alternatively the set of centroids provided by fixed partitions as advocated by Jassem.

CONCLUSION

Unless MCA can be modified so that it automatically identifies the optimum number of centroids as well as their position, it is less useful than peak-picking LPC for the purpose of automatically identifying fricative formant structure. Multiple centroids may, however, be suitable as abstract parameters for fricative identification.

REFERENCES

1] Nguyen, N. and Hoole, P. (1993) Frequency variations of the lowest main spectral peak in sibilant clusters., *Proc. Eurospeech 93*, Vol. 1, pp 81-84.

[2] Mcgowan, R.S. and Nittrouer, S. (1988) Differences in fricative production between children and adults: Evidence from an acoustic analysis of /sh/ and /s/., *J. Acoust. Soc. Am.*, Vol 83, pp 229-236.

[3]Heinz, J.M. and Stevens, K.N. (1961) On the properties of voiceless fricative consonants, *J. Acoust. Soc. Am.*, Vol. 33 No. 5, pp 589-596

[4] Stieglitz, K. (1977) pp On the simultaneous Estimation of poles and zeros in speech analysis, IEEE Trans. ASSP, Vol. 25, No. 3, pp 223-233

[5] Soli, S.D. (1981) Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation, *J. Acoust. Soc. Am.* Vol. 70, pp. 976-984.

[6]Zue, V.W. (1991) From signals to symbols to meaning: On machine understanding of spoken language, Proc. Int. Conf. Phonetics, Speech, Vol. 1, pp 74-83

[7] Baum, S.R. and McNutt, J.C. (1990) An acoustic analysis of frontal misarticulation of /s/ in children. *Journal of Phonetics*, Vol. 18, pp 51-63.

[8] Nittrouer, S., Studdert-kennedy, M. and McGowan, R.S. (1989) The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal Speech and Hearing Research*, vol. 32, pp. 120-132.

[9] Forrest, K., Weismer, G., Milenkovic, P. and Dougall, R.N. (1988) Statistical analysis of word-initial voiceless obstruents: Preliminary data. *J. Acoust. Soc. Am.*, vol. 84, pp. 115-123.

[10] Nittrouer, S. (1995) Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *J. Acoust. Soc. Am.*, Vol. 97, pp. 520-530.

[11] Jassem, W. (1979) Classification of fricative spectra using statistical discriminant functions. In Lindblom, B. and Ohman, S. (eds.) *Frontiers of Speech Communication Research*, (London: Academic Press), pp. 77-91.

[12]Crowe, A. and Jack, M.A. (1987) Globally optimising formant tracker using generalised centroids, *Electronic Letters*, Vol. 23, No. 19, pp 1019-1020

[13] Wrench, A.A., Watson, J.M.M., Soutar, D.S., Robertson, A.G., And Laver, J. (1994) Fast formant estimation of childrens speech. *Proc ICSLP94*, Vol. 3, pp. 1651-1654.