

VOWEL QUALITY ASSESSMENT BASED ON ANALYSIS OF DISTINCTIVE FEATURES

Shuping Ran, Bruce Millar and Iain Macleod
Computer Sciences Laboratory
Research School of Information Sciences and Engineering
Australian National University, ACT 0200, Canberra, Australia
shuping@cslab.anu.edu.au

ABSTRACT

This paper presents a novel method for vowel quality assessment which is based on the analysis of distinctive features. A sub-set of the features proposed by Jakobson, Fant and Halle are used. The acoustic features associated with each selected feature are encoded in the weights of a multilayer perceptron. The processing of an individual vowel sound by such feature-detecting perceptrons yields an estimate of that vowel's position in a distinctive feature space. The relative positions of one speaker's monophthongal vowel set within both a two-dimensional and a three-dimensional distinctive feature space are compared with vowel spaces based on articulatory considerations. The extent to which articulatory descriptions of acoustic vowels can be estimated using this method is discussed.

Keyword: Distinctive features; Artificial neural networks, Articulatory features, Vowel quality assessment.

1 INTRODUCTION

The vowels within any dialect or even idiolect are normally phonemically labelled according to their relative positioning and the established phonological structure of vowels in the language. To enable rigorous analysis of vowel sounds, it is necessary to use the skills of a well-trained phonetician who is able to locate any individual vowel within an established vowel reference system such as the cardinal vowel chart of Daniel Jones [3]. The two major dimensions of this chart comprise the open-close dimension which is a relative measure of the openness of the oral cavity, and front-back dimension which is a relative measure of the longitudinal position of the highest point of the tongue. A further distinction, rounded-unrounded, which is a relative measure of roundedness of the lips, is notoriously difficult for even well-trained phoneticians to make independently of the other two dimensions [4].

This paper presents a novel method for vowel quality assessment which is based on the analysis of distinctive features. The features used are a sub-set of those proposed by Jakobson, Fant and Halle [2], namely "compact" (vs "diffuse"), "acute" (vs "grave"), and "flat" (vs "plain") representing the open-close, front-back and rounded-unrounded dimensions of the cardinal vowel chart respectively [7].

2 METHOD

A pilot data corpus comprising five repetitions of [stop][vwl]d utterances by four speakers of Australian English was selected, where [stop] represents the six voiced and unvoiced labial, alveolar, and velar stop consonants of English ([b, p, d, t, g, k]), and [vwl] represents the eleven Australian nominally monophthongal vowels ([i, I, e, æ, a, ɒ, ɔ, ʊ, u, ʌ, ɜ]). The [stop][vwl]d utterances had been manually segmented and labelled previously [7]. Only the pseudo steady-state vowel interval was of interest for this study.

The pseudo steady-state vowel intervals were processed in "frames" of 12.8ms, with adjacent frames having a 6.4ms overlap, by passing them through a Hamming window, then deriving 13 Linear Predictive Cepstral Coefficients (LPCCs) for each frame. The LPCCs were used as input to three 2-layer multilayer perceptrons (MLPs) on a frame-by-frame basis where each MLP was trained to detect the presence of one distinctive feature using its complex interpolative transformation ability [5].

The acoustic features associated with each selected feature were encoded in the weights of a multilayer perceptron. The processing of an individual vowel sound by such feature-detecting perceptrons then yielded estimates of that vowel's position in a distinctive feature space (Figure 1).

3 EXPERIMENT

As an initial assessment of the viability of this approach, the feature detectors were trained using three reference vowels from an existing data corpus. The reference vowels were

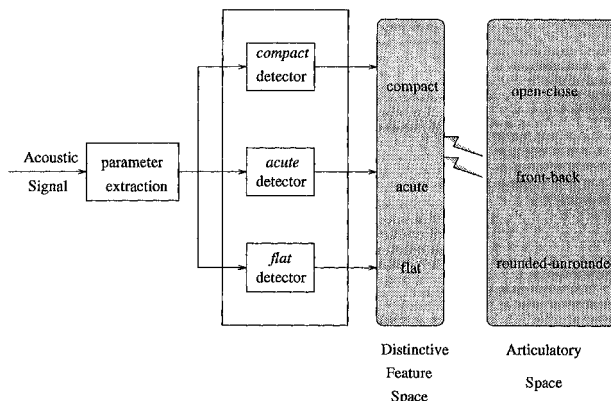


Figure 1: Global view of the design.

	<i>front-close</i>	<i>back-open</i>	<i>back-close</i>
	I	ɒ	ʊ
compact	0	1	0
acute	1	0	0
flat	0	0	1

Table 1: Description of reference vowels in terms of features and articulatory description.

chosen according to their relatively extreme positions on the cardinal vowel chart and their stability within Australian English; repetitions of these vowels can thus be used to construct viable models of their relevant features. These reference vowels were [I, ɒ, ʊ], where /I/ was described as a front-close vowel, /ɒ/ as a back-open vowel, and /ʊ/ as a back-close vowel. Table 1 is a description of these vowels in terms of these approximate articulatory descriptions and their corresponding distinctive feature patterns [2].

Each MLP feature detector was trained by comparing its output value to the model values for each vowel-feature combination as shown in table 1 then adjusting its internal weights using the back-propagation algorithm. Training was concluded when the error could not be reduced further. The inputs for this training comprised five repetitions of the three pseudo steady-state reference vowels from all six consonantal contexts. Feature detectors were trained and tested independently for the four speakers but the results of only one typical speaker are presented in this paper due to space limitation.

In order to test the performance of the trained feature detectors, all the 11 vowels in the data corpus were presented as input to the system. The output of each detector was the "activation level" of the single output layer node. This value represents the degree of presence of the feature, on a scale of [0,1], for the current input "frame".

4 RESULTS

The original view of the cardinal vowel system was of a two dimensional space formed by the open-close and front-back articulatory dimensions as described by Daniel Jones [3]. Another view of the cardinal vowel system involves a three dimensional space formed by open-close, front-back and rounded-unrounded dimensions. Daniel Jones' two dimensional cardinal vowels are represented by vowel positions indicated in Figure 5 [4]. In this section, we report the results of this study in two parts which relate to these two and three dimensional spaces.

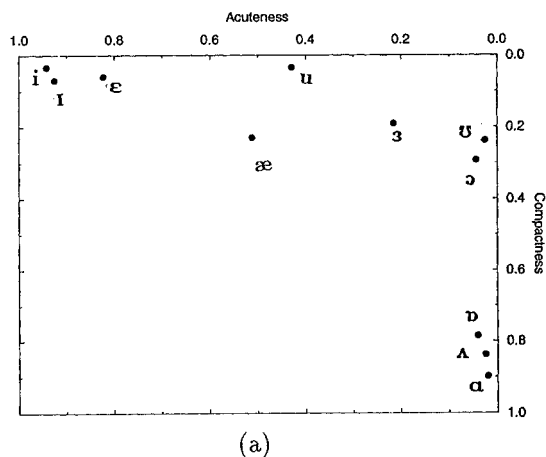
4.1 Two Dimensional Space Formulation

The two dimensional distinctive feature space is generated by combining the output from the *compact* feature detector with that of the *acute* feature detector. *Compactness* ranging from a value of 1 (*compact*) to 0 (*diffuse*) is represented by the vertical axis and *acuteness* ranging from a value of 1 (*acute*) to 0 (*grave*) is represented by the horizontal axis. The position of each vowel in the test set as determined by the feature detector outputs is represented by a point in this two-dimensional space (Figure 2).

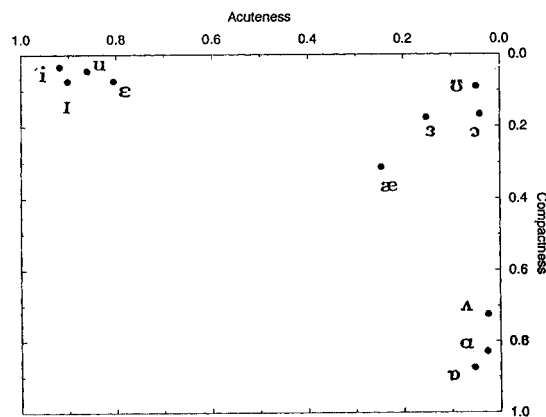
In order to make a comparison with an existing cardinal

vowel description of Australian vowels, we choose Mitchell's cardinal vowel plots for Australian English [6] illustrated in Figure 4 which can be directly related to a distinctive feature space using the axis labels in parentheses. The relative position of the vowels was modified by Bernard [1] who analysed Australian English vowels on basis of acoustic data more recently. The outcome of Bernard's study indicated that the vowel /u/ for the average Australian speaker is more centralised than Mitchell's early study. We therefore ignore differences in the position of /u/ from that in Figure 4.

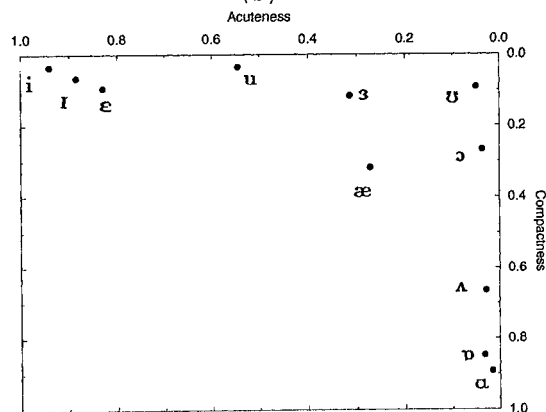
Observing Figure 2, we can see that clusters of vowels represented in this generated feature space generally preserve their relative position in the different stop consonant



(a)



(b)



(c)

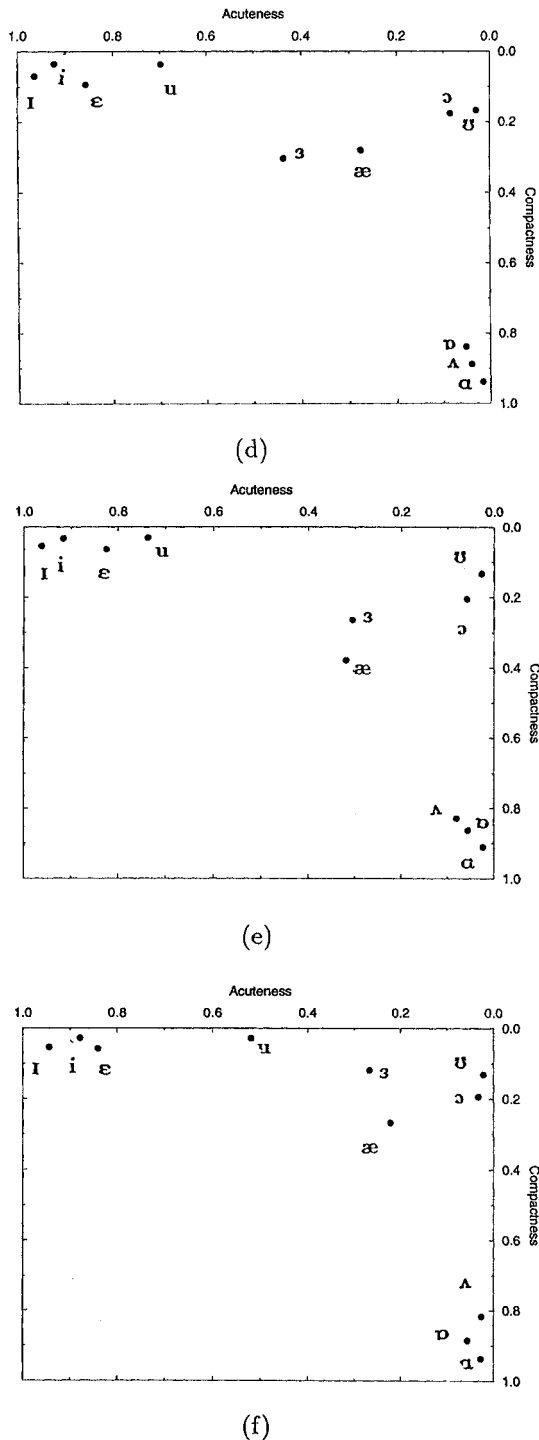


Figure 2: *IM's 11 pseudo steady-state vowels on Acuteness versus Compactness plane in the context of six stop consonants: (a) [b]; (b) [d]; (c) [g]; (d) [p]; (e) [t]; (f) [k].*

contexts. Absolute positions of individual vowels do change according to the context, with /ɜ/ being the most variable. The intermediate position of /æ/ on both dimensions is well maintained in five of the six consonant contexts. The lack of a front-open model is most evident in the absolute placement of /æ/ and /ε/.

4.2 Three Dimensional Space Formulation

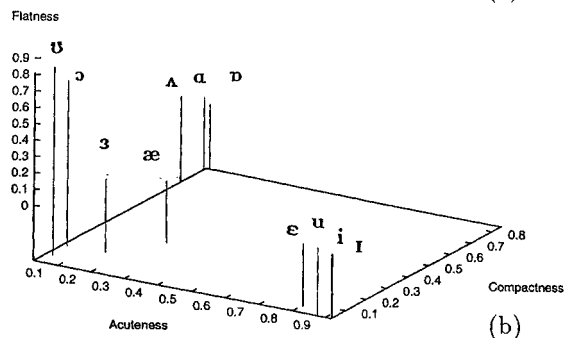
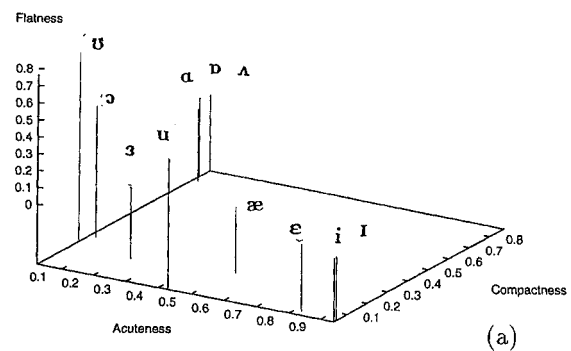
The three dimensional space is formed by using the two dimensional space as a base plane and adding a vertical axis representing the output of the *flat* feature detector yielding the dimension of "flatness", ranging from 1 (*flat*) to 0 (*plain*).

The position of each vowel in the test set as determined by the three feature detector outputs is represented by the top of a line rising vertically from the base plane (Figure 3). The length of the lines representing flatness (or in articulatory terms—roundedness) for each vowel may be compared with the roundedness estimates for cardinal vowels (Figure 5) proposed by Ladefoged [4].

The relative high flatness of /ʊ/ and /ɔ/ (with mean flatness equal to 0.68) and the relative low flatness of /i/, /ɪ/, and /ε/ (with mean flatness equal to 0.035) correlate well in general terms with their proximity (see Figure 4) to the cardinal vowels 7 and 2 respectively (see Figure 5). The flatness of /ɑ/, /ɒ/, and /ʌ/ (with mean flatness equal to 0.085) would be expected to take intermediate values due to their proximity to the cardinal vowel 5. The much lower flatness measured here is due to the fact that the reference vowel /ɒ/ was taken as unrounded (*plain*) according to Mitchell [6]. A better model for flatness is expected to be created if only /ɪ/ and /ʊ/ are taken as reference vowels.

5 DISCUSSION AND CONCLUSION

This pilot study has shown that multilayer perceptrons operating on a cepstral representation of the vowel space on a speaker dependent basis can encode the rudimentary characteristics of the distinctive features of all vowels when trained with a subset of 'extreme' vowels. Although broad inter-vowel relationships are preserved in the generated feature domain, to gain a more precise measurement which is comparable to the articulatory phonetic vowel quality assessment by well-trained phoneticians a more refined set of reference vowels and reference vowel descriptors in the feature



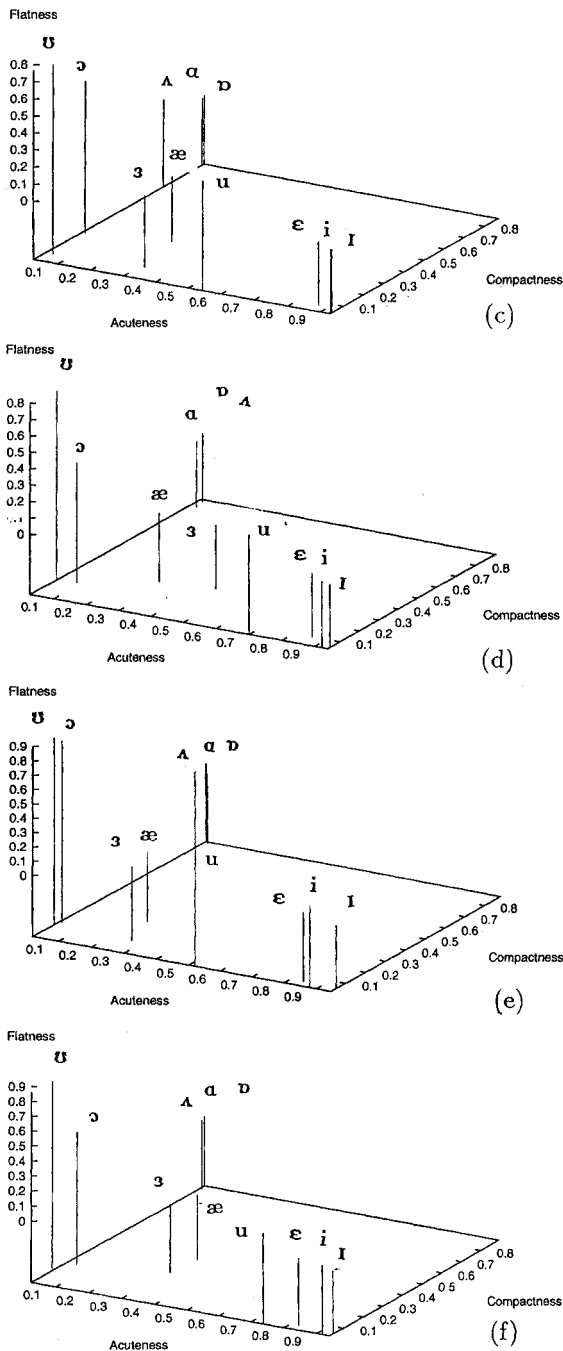


Figure 3: IM's 11 pseudo steady-state vowels in an Acuteness versus Compactness versus Flatness space in the context of six stop consonants: (a) [b]; (b) [d]; (c) [g]; (d) [p]; (e) [t]; (f) [k].

space are required. Two further developments are therefore proposed.

It is proposed that the speaker should produce the four articulatorily defined cardinal vowels for use by the system as references for training instead of using three reference vowels from the existing data corpus. This should have at least two effects: first, naturally extreme vowels should be placed more accurately as they will not then be so close to the boundaries of the "trained" feature space, and second, the front-open corner of the articulatory vowel space will

have a reference model which was missing from the experiments reported in this paper.

It is further proposed that the natural vowels of the speaker be carefully placed on a cardinal vowel chart by a suitably trained phonetician so that a quantitative assessment of the technique proposed in this paper can be made.

If both of these proposed refinements can be achieved then it is likely that the method presented in this paper can be effectively tested as a tool to provide an objective measurement of phonetic quality for vowel quality assessment on acoustic data alone.

References

- [1] Bernard, J. R. (1989), "Quantitative aspects of the sounds of Australian English", in Collins, P. and Blair, D. (Eds.), *Australian English: The Language of A New Society* (University of Queensland Press, Australia), pp. 87-204.
- [2] Jakobson, R. Fant, C. G. M. and Halle, M. (1961), "Preliminaries to speech analysis, The distinctive features and their correlates", Technical Report No.13 of the M.I.T. Acoustics Laboratory, (The M.I.T. Press).
- [3] Jones, D. (1956), *An Outline of English Phonetics* (W. Heffer & Sons Ltd., Cambridge, MA).
- [4] Ladefoged, P. (1975), *Three Areas of Experimental Phonetics* (Fourth edition), (Oxford University Press, London).
- [5] Lippmann, R. P. (1987), "An introduction to computing with neural nets", *IEEE Trans. on Acoustics, Speech and Signal Processing*, 4(2), pp. 4-22.
- [6] Mitchell, A. G. (1946), *The Pronunciation of English in Australia* (Halstead Press, Sydney).
- [7] Ran, S. (1994), *Speech Knowledge Modelling for Speech Recognition: A Study Based on Distinctive Features*, PhD thesis, The Australian National University.

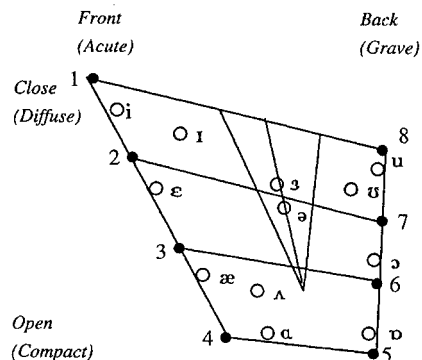


Figure 4: Two dimensional cardinal vowel system ([6], p. 63. Numbers are used to represent the cardinal vowels in order to avoid confusions between those Australian English vowels and cardinal vowels which use the same phonetic symbols.)

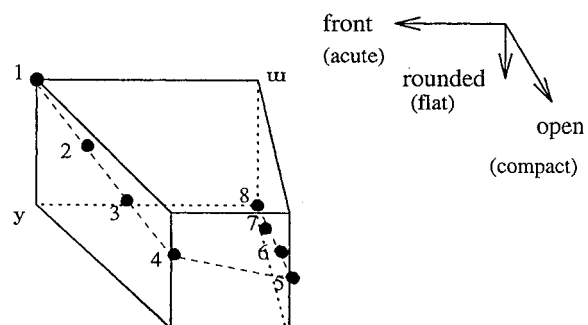


Figure 5: Three dimensional cardinal vowel system ([4], p. 140. Phonetic symbols for cardinal vowels are replaced by cardinal vowel numbers for compatibility with Figure 4.)