

THE INFLUENCE OF ACCENTS IN AUSTRALIAN ENGLISH VOWELS AND THEIR RELATION TO ARTICULATORY TRACT PARAMETERS

D. R. Dersch

C. Cleirigh

J. Vonwiller

Speech Technology Research Group

Department of Electrical Engineering, The University of Sydney, NSW 2006 Australia
Email: dersch@speech.usyd.edu.au, cleirig@speech.usyd.edu.au, julie@speech.usyd.edu.au

ABSTRACT

In this paper we analyse and compare a low dimensional linguistic representation of vowels with high dimensional prototypical vowel templates derived from a native Australian English speaker. We further perform the same analysis on Lebanese and Vietnamese accented English to investigate how differences due to accents impact on such a representation. In a low dimensional linguistic representation a vowel is characterised by articulatory tract parameters. To simplify the problem, the study is restricted to vowels that, notionally at least, involve a steady state articulation i.e. a stable target configuration of tongue, lips and jaw between preceding and following articulatory transitions. Vowels are represented by the horizontal and vertical position of the part of the tongue involved in the key articulation of a particular vowel, e.g., high or low and front or back. To this is added lip posture, spread or rounded. Prototypical vowel templates are derived as follows. The sound pressure signal is parametrized by 12 mel-frequency cepstrum coefficients. At the centre of each phonetically labelled segment, 180 dimensional phone templates are extracted. For the group of short (/I/, /E/, /A/, /O/, /V/, /U/, /@/) and long vowels (/i:/, /e:/, /a:/, /o:/, /u:/, /@:/) we obtain vowel clusters by averaging over all templates of each vowel class and accent. The speech material is taken from the Australian National Database Of Spoken Language (AN-DOSL). For a comparison of high dimensional vowel clusters derived from speech samples with low dimensional prototypical vowels in the articulatory tract representation we perform a reduction in dimension by a multidimensional scaling transformation in a two dimensional space. Here, a linear transformation maps a high dimensional space on a lower dimensional sub space by optimising the relative distances between data vectors. As an important result we find. i) /@/ and /@:/ are surrounded by the remaining vowels; ii) the overall structure and the relative distances between the prototypical vowels are very similar. Variations in the structure can be explained by the influence of native Australian English, Lebanese Arabic and South Vietnamese accents.

1. INTRODUCTION

Clark and Yallop (pp. 311) [1] trace the use of distinctive phonological features in modern times to Jakobson [2, 3], who drew on earlier phonological concepts of de

Saussure and Hjelmslev. Subsequent important elaborations appeared in Jakobson, Fant and Halle [4] and Jakobson and Halle [5]. Articulatory features – as well as acoustic and perceptual features – have been employed by a wide range of phonological theories at various levels of abstraction ever since. These articulatory parameters were used to divide the phonemes of English into 'sets' where bundles of features could be used to describe phoneme types. For example, for vowels, the mouth was viewed in a 2 dimensional plane based on the position of the part of the tongue involved in the key articulation of a particular vowel. The front to back axis related to the section of the tongue which was raised; and the high low axis related to the height of that part of the tongue in the mouth. To this was added lip posture, spread or rounded. For English, lip rounding is associated with vowels produced with the back part of the tongue raised.

This is the most basic view of the structure of the English vowel system, but if this can be demonstrated to represent some valid assumptions about this system, then it should be possible to compress information about the vowel system into a simpler format such that knowledge of the articulation of the vowels would be sufficient. To investigate this, we determined to compare the 3 dimensional linguistic representation of vowels with the real data MFCC output. We also performed the same analysis on Lebanese and Vietnamese accented English to investigate whether the output of the speech data can indicate the types of influences that produce the accent, for example, the first language phonology, language learning strategy, gender etc. The comparison is between 3 male speakers only.

The study is restricted to vowels that generally involve a steady state articulation – a stable target configuration of tongue, lips and jaw – between preceding and following articulatory transitions [1](pp. 73). That is, we focus on short (eg /I/) and long monophthongs (eg /i:/), and exclude diphthongs (eg /au/) since these are complex vocalic sounds whose articulation requires two target configurations of tongue, lips and jaw (ibid: 73, 102). Monophthongs do involve onglides and offglides because of coarticulation, but the central portion of the vowel is assumed to be constant.

2. ARTICULATORY TRACT PARAMETER

Vowels are represented by the horizontal and vertical position of the part of the tongue involved in the key articulation of a particular vowel, e.g., high or low and front or back. To this is added lip posture, spread or rounded. The parameter set for different short vowels are shown in Table 1.

Vowel	tongue posture	lip posture	aperture
/I/	front	spread	high
/E/	front	spread	middle
/A/	front	spread	low
/O/	back	round	middle
/V/	back	spread	low
/U/	back	round	high
/@/	back	spread	middle

Table 1: The three articulatory tract parameters for the seven short vowels /I/, /E/, /A/, /O/, /V/, /U/, /@/

The parameter set for different long vowels are shown in Table 2.

Vowel	tongue posture	lip posture	aperture
/i:/	front	spread	high
/e:/	front	spread	middle
/a:/	back	spread	low
/o:/	back	round	middle
/u:/	middle	round	high
/@:/	back	spread	middle

Table 2: The three articulatory tract parameters for the six long vowels /i:/, /e:/, /a:/, /o:/, /u:/, /@:/

3. SPEECH CODING AND SPEECH VISUALIZATION

3.1. Speech Coding and Phone Representation

The speech material of native and accented Australian English speakers are taken from the Australian National Database Of Spoken Language (ANDOSL), see e.g. [11]. The database comprises a set of phonetically rich, read sentences. For each accent group we use a set of phonetically manually transcribed sentences taken from five different speakers. The sound pressure signal sampled at a rate of 20 kHz is parameterized by 12 mel-frequency cepstrum coefficients by applying a Hamming window of 16 msec duration and 5 msec step size. The mfcc spectrum is pre-emphasised by a filter coefficient of 0.97. At the centre of each phonetically labelled segment, phone templates are extracted. For the group of short vowels (/I/, /E/, /A/, /O/, /V/, /U/, /@/) templates comprising 15 time steps (75 msec) are extracted. For the long vowels (/i:/, /e:/, /a:/, /o:/, /u:/, /@:/) we extracted 30 time steps (150 msec) and averaged two preceding time steps. As a result long and short vowels are represented by 180-dimensional vectors. We obtain vowel cluster by averaging over all tem-

plates of each vowel class. (The vowel /e:/ is undergoing change, - it is a diphthong for older Australians and a long monophthong for younger Australians.

A representation of vowels according to the three articulatory tract parameters tongue posture, lip posture and aperture might be achieved by assigning an integer number to each value, e.g., front => +1, back => -1, spread => +1, round => -1, high => +1, middle => 0, low => -1. Thus each vowel is then coded by a three dimensional vector.

3.2. Multidimensional Scaling

Visualising high dimensional data sets is a challenging problem in data analysis. Vector Quantization and clustering algorithms (see e.g. [8]) may give insights in the structure of high dimensional data sets. Another useful visualisation of a set of N n-dimensional data vectors is *Multi Dimensional Scaling (MDS)*. Here, a linear transformation maps a high dimensional space on a lower dimensional space in such a way, that the relative distances between data vectors are preserved as well as possible. For a closer explanation of MDS see Mardia et. al. [10]. MDS has been successfully applied in investigating a high dimensional space of phone templates [7]. For a better comparison of vowel cluster and vowels in the articulatory tract representation we perform a MDS transformation in a two dimensional space. The results are shown in the next section.

4. RESULTS

4.1. Short vowels

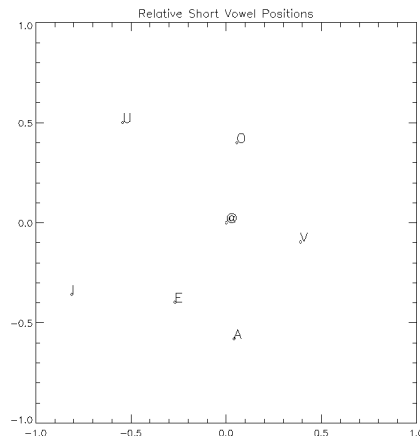


Figure 1: Result of a MDS transformation for the prototypical short vowels.

Figure 1 shows the result of a MDS transformation of prototypical phones based on articulatory tract parameters and Figures 2-4 give the mean values of phone clusters for the seven short vowels for native Australian English (AuE), Vietnamese accented Australian English (VAE) and Lebanese accented Australian English (LAE). We found that the four plots exhibited a similar overall structure. In each Figure, /@/ is surrounded by the re-

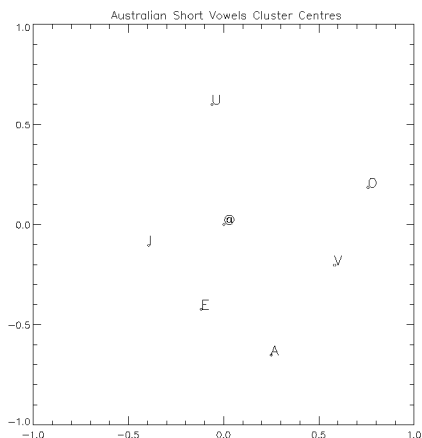


Figure 2: Result of a MDS transformation for short vowel cluster centre taken from native Australian English speaker.

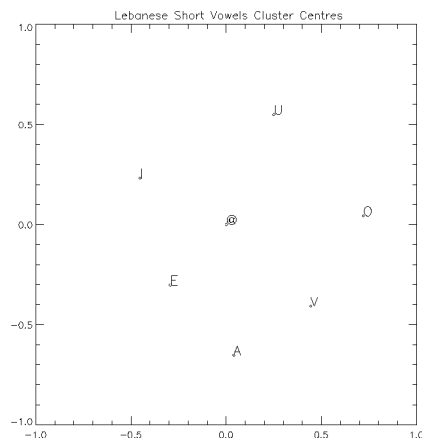


Figure 4: Result of a MDS transformation for short vowel cluster centre taken from Lebanese Arabic accented speaker.

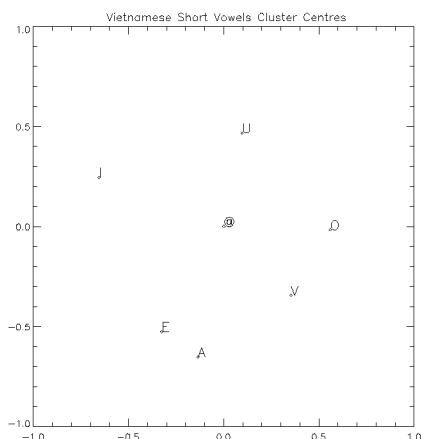


Figure 3: Result of a MDS transformation for short vowel cluster centre taken from South Vietnamese accented speaker.

maintaining six short vowels. The relative next neighbour relationship between the phones is equal in each Figure. The LAE is closer to AuE than VAE. These vowels are represented in all 3 languages, Classic Arabic has a 3 vowel system with long and short vowels, and modern regional Arabic dialects have evolved to include a variety of other vowels. Lebanese Arabic is in this category. Vietnamese has a similar pattern of distribution of vowels but in contrast to the other 2 languages, has rounded and unrounded back vowels [6]. Figure 3 shows that the distance between /E/ and /A/ is quite small which may reflect the fact that Vietnamese has 4 front vowels and /E/ and /A/ are mid and low. For the short vowels then the results indicates that a 3 dimensional linguistic description can be used to represent them.

4.2. Long vowels

Figures 5 – 8 show the results for the long vowels. Figure 5 represents the proto-typical AuE phoneme distribution, Figure 6 the actual speech data AuE, Figure 7 the VAE, and Figure 8 the LAE. There are greater differences here. Duration: The prototypical long vowels are similar in po-

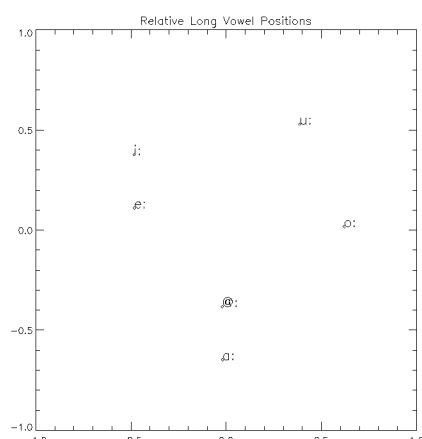


Figure 5: Result of a MDS transformation for the prototypical long vowels.

sition to their prototypical short vowels. However, this is not fully reflected in the speech data of any of the 3 speakers. The Vietnamese speaker comes closest to this pattern. Again the VAE speaker has his vowels well distributed in the vowel space, but not as an exact mirror of the short vowels. /@:/ /u:/ and /o:/ come closest to mirroring /@/ /U/ and /O/ - their short counterparts. Such a well distributed vowel set may reflect the fact the Vietnamese is a tone language and relies on clear vowel distinctions on which to allocate tone. For the LAE speaker, the /e:/ and /@:/ are clustered and separate from all other long vowels. This may represent a distribution based on orthography, as the 2 clustered vowels are both orthographically written as 'e'. This then would make it possible to describe the distribution achieved by the MDS for the LAE speaker as potentially the result of the fact that this speaker learned English from reading.

The relative next neighbour relationship between the phones is equal in each Figure, except for the native Australian English speaker where the relative distance between /u:/ and /e:/ appears to be much smaller than that between /i:/ and /e:/.

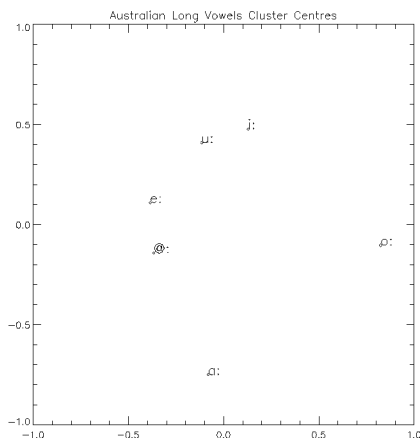


Figure 6: Result of a MDS transformation for long vowel cluster centre taken from native Australian English speaker.

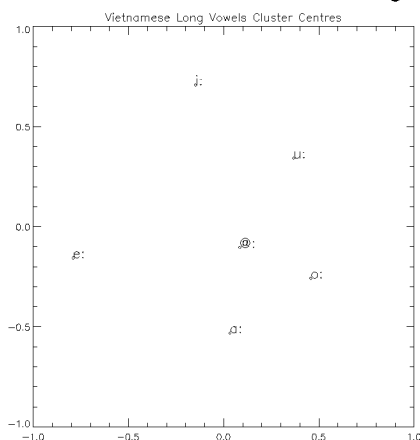


Figure 7: Result of a MDS transformation for long vowel cluster centre taken from South Vietnamese accented speaker.

tances in the 180 dimensional phone template space shows, that the difference in the distance is only in the range of five percent. That means, that the low dimensional MDS transformation corrupts the true distances for these vowels.

5. REFERENCES

1. J. Clark, C. Yallop, An introduction to phonetics and phonology. Oxford: Basil Blackwell, 1990.
2. R. Jakobson, Observations sur le classement phonologique des consonnes. Proceedings of the Third International Congress of Phonetic Sciences pp. 34-41, Ghent, 1939.
3. R. Jakobson, On the identification of phonemic entities. Travaux du Cercle Linguistique de Copenhagen 5 pp. 205-13, 1949.
4. R. Jakobson, C. G. M. Fant and M. Halle, Preliminaries to speech analysis: the distinctive features and their correlates. Cambridge, Mass.: MIT Press, 1952 (MIT Acoustics Laboratory Technical Report 13.)
5. R. Jakobson and M. Halle, Fundamentals of language. The Hague: Mouton, 1956.
6. D. H. Nguyen, Vietnamese. Bernard Comrie (ed) The Major Languages of the World Routledge London, pp. 49-68, 1990.
7. D. R. Dersch, Neural network approaches to speech recognition: A Generalized Radial Basis Function Network for speaker-independent phone classification. *Proceedings of the 7th Australian International Conference on Speech Science and Technology (SST-96)*, Adelaide, December 10-12, 1996, pp. 263-268, Canberra, 1996.
8. R. M. Gray, "Vector quantization", IEEE ASSP Magazine, vol. 1, pp. 4-29, 1984.
9. J. Ingram and J. Pittam, Auditory and acoustic correlates of perceived accent change: Vietnamese schoolchildren acquiring Australian English. *Jour. of Phonetics*, pp. 127-143, 1987.
10. Mardia K. V., Kent J. T. and Bibby J. M. (1979) *Multivariate Analysis* Academic Press, London.
11. B. Millar, J. Vonwiller J. Harrington, and P. Dermody, "The Australian National Database of Spoken Language", ICASSP-94, pp. 67-100, Adelaide, 1994.
12. M. J. Munro, Production of English vowels by native speakers of arabic: Acoustic Measurements and accent-ness ratings. *Language and Speech*, pp. 39-66, 1993.