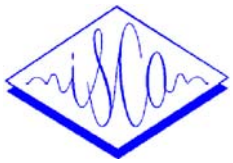

INTERSPEECH 2004 - ICSLP
8th International Conference on Spoken Language Processing
ICC Jeju, Jeju Island, Korea
October 4-8, 2004

MODELLING AND RANKING OF DIFFERENCES ACROSS FORMANTS OF BRITISH, AUSTRALIAN AND AMERICAN ACCENTS

*Qin Yan Saeed Vaseghi Dimitrios Rentzos Ching-Hsiang Ho**

Department of Electronic and Computer Engineering
Brunel University, UK UB8 3PH
*Fortune Institute of Technology, Kaohsiung, Taiwan
{Saeed.Vaseghi, Qin.Yan, Dimitrios.Rentzos}@brunel.ac.uk *ch.ho@center.fjtc.edu.tw

10.21437/Interspeech.2004-207

ABSTRACT

The differences between formants of British, Australian and American English accents are analysed and ranked. An improved formant model based on linear prediction (LP) feature analysis and a two-dimensional(2D) hidden Markov model (HMM) of formants is employed for estimation of the formant frequencies and bandwidths of vowels and diphthongs. Comparative analysis of the formant trajectories, the formant target points and the bandwidth of the spectral resonance at formants of British, Australian and American accents are presented. British vowels and diphthongs have smaller formant bandwidth than Australian. A method for ranking the contribution of different formants in conveying an accent is proposed whereby formants are ranked according to the normalized distances between the formants across accents. The first two formants are considered more sensitive to accents than other formants.

1. INTRODUCTION

The differences across accents are affected by the following factors: (a) differences in the phonetic transcriptions of words, (b) differences in the acoustic production of phonemes and (c) differences in intonation [1]. The differences in the acoustic production of phonemes across accents are affected by the differences in the formant trajectory and duration of phonemes.

This paper presents a comparative investigation of the differences in the formant trajectories of vowels and diphthongs of British, Australian and American accents. Previous work on comparative analysis of the formants of accents of English include the work of Harrington and Watson [2,3] on the differences of formants between subclasses of Australian English: Broad Australian English, General Australian English and Cultivated Australian English and between New Zealand and Australian English. Arslan and Hansen investigated the difference between the formants of native and non-native speakers of English [4].

This paper presents statistical methods based on two-dimensional hidden Markov models for modeling the probability distribution of formants. The probability models are subsequently used for estimation of formant trajectories.

Experimental results presented in this paper illustrate the differences across the accents between the frequency and bandwidth of resonance at formants. A formula is introduced for ranking the influence of each formant in conveying the difference across accents.

The databases employed for accent analysis are ANDOSL for broad Australian English, WSJCAM0 for British English (Received Pronunciation) and WSJ for American English as listed in Table 1.

This paper is organized as follows. In section 2 formant estimation and modeling is introduced. Section 3 compares the formant trajectories of British, Australian and American accents. Section 4 presents a comparison of the bandwidth of the spectral resonance at formants across accents. Section 5 presents a method for ranking the contribution of formants in conveying accents and finally section 6 concludes this paper.

2. FORMANT ESTIMATION

Although automatic formant analysis of speech has received considerable attention and a variety of approaches have been developed, the estimation of accurate formant features from the speech signal remains a non-trivial problem. Formant classification is described in [5,6]. Each formant feature vector [F_k, BW_k, I_k, ΔF_k, ΔBW_k, ΔI_k] has 6 parameters: formant frequency F_k, bandwidth BW_k, and intensity I_k together with the slopes of their time trajectories ΔF_k, ΔBW_k, and ΔI_k.

Database Name (accent)	No. Speakers (female/male)	No. Sentences
ANDOSL (Australian)	18/18	7200
WSJ (American)	36/38	9438
WSJCAM0(British)	40/46	9476

Table 1 Databases configuration

Parameters	Values
Sample Frequency	10kHz
Frame size	25 ms
Frame rate	10 ms
No. of states per speech HMMs	3
MFCC features for speech HMMs	39
No. of Gaussian Mixtures per state in speech HMMs	20
LP order	13
No. of states per Formant HMMs	5 or 6
No. of Gaussian components per state in formant HMMs	4
Bandwidth threshold in discarding LPC poles	800 Hz

Table 2 Configuration of analysis of acoustic correlates of accents

A 2D HMM with 3 left–to-right states across time and five left–to-right states across frequency is used to classify formant candidates in each frame among five sequential formant clusters. The HMM configuration parameters are listed in Table 2.

The 2D HMM-based formant classifier [5,6] may associate two or more formant candidates (i.e. LP model pole frequencies) $F_{i(t)}$, with the same formant b, in these cases formant estimation is achieved through minimization of a weighted mean square error objective function

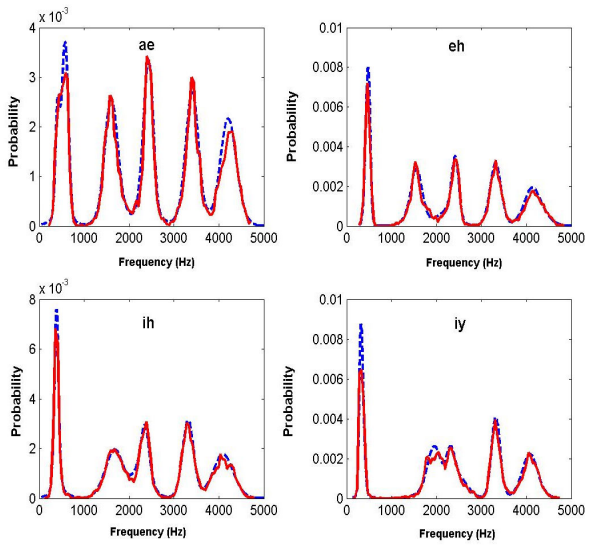

Figure 1: Comparison of histograms and HMMs of Formants from an American male speaker. Bold dashed line: Gaussian curves modelled by HMMs. Thin solid line: Histograms X axis: Frequency (Hz); Y axis: probability.

$$\hat{F}_b(t) = \min_{F_b(t)} \sum_{i=1}^{I_b(t)} w_{bi}(t) \left[\frac{(F_{i(t)} - F_b(t))^2}{BW_i(t)^2} \right]$$

where t denotes the frame index, b is the formant index, $I_{b(t)}$ is the total number of the formant candidates classified as formant b. The squared error function is weighted by a perceptual weight $1/(BW_i)^2$ where BW_i is the formant bandwidth, and $w_{bi}(t) = P(F_i|\lambda_b)$ is a probabilistic weight where λ_b is the Gaussian mixtures models of b^{th} formant state of a phoneme-dependent HMM of formants. The success of this method can be seen in Figure 1 showing the close match between the histograms of formant candidates of a phoneme and the corresponding Gaussian models of HMM states for the vowels *ae, eh, ih* and *iy*. It can also be seen that the peaks of estimated Gaussian curves and histograms, which occur at formant frequencies, coincide.

3. COMPARISON OF FORMANT TRAJECTORIES OF BRITISH, AUSTRALIAN AND AMERICAN ACCENTS

In order to obtain an average formant trajectory model for each phoneme, the estimated formant trajectories of the different examples of each phoneme are linearly time-warped to ensure the same duration. Then all the time-normalized trajectories for each phoneme are averaged.

During the realization of each phoneme, the acoustic configuration of vocal tract traverses a time trajectory aiming to reach a *target* point in the formant space. Using the method described in [5,6], the acoustic *target* point of a vowel is marked as a single time point (a "static" spectral slice [7]) between the acoustics onset and offset of the vowel. For high front vowels, the target is marked at the point where F2 reaches a peak; for back vowels, the target is marked at the point where F2 reaches a trough;

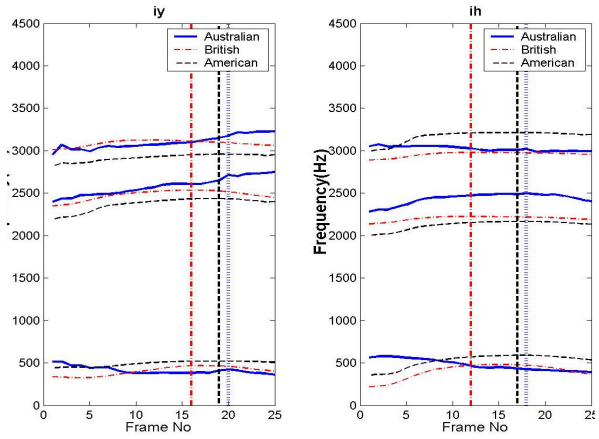

Figure 2: Average Formant Trajectories of *iy ih* of British, Australian and American after alignment (female)
X-axis is the normalized time. Y-axis is frequency (Hz).

for open vowels, the target is marked at the point where F1 reaches a maximum. When there is no acoustic evidence of any kind for a target point, the target is marked at the vowel's temporal mid-point. In rising diphthongs, two targets are marked using the same method as for monophthongs.

The normalized formant trajectories of some vowels are shown in Figure 2. Vertical lines are superimposed on these trajectories to mark the average time at which the vowel targets occur. It is noticeable that American vowels have considerably delayed target points compared with those of British vowels. This is particularly evident in *iy*. However, for *ih* the target is a good deal closer to the vowel's temporal mid-point.

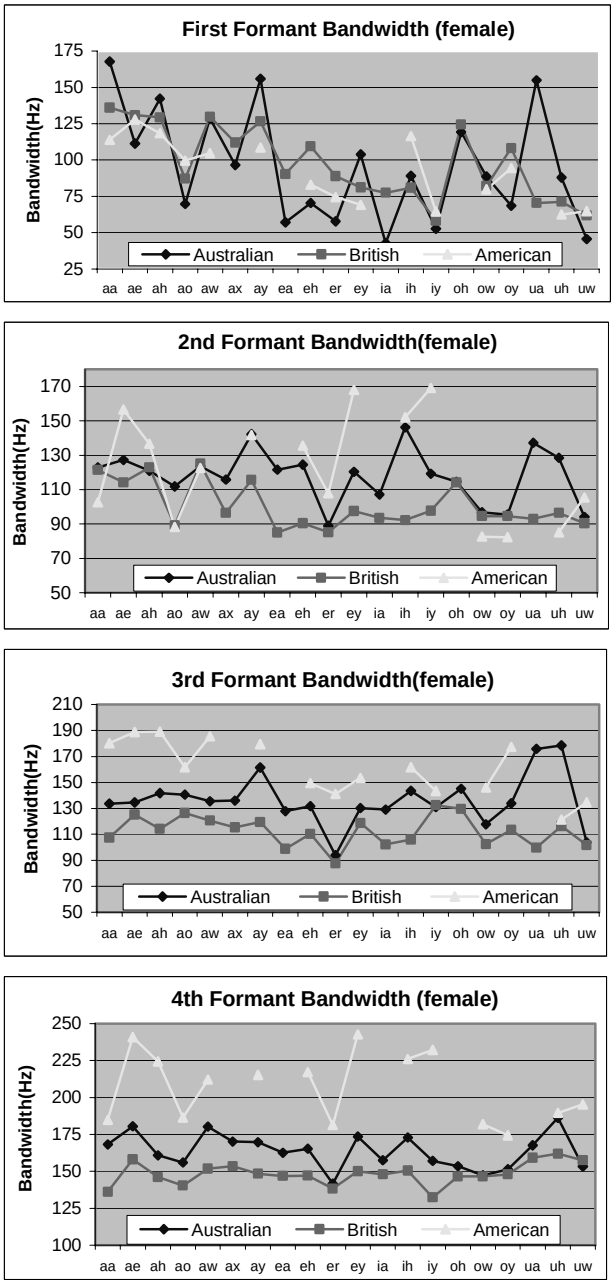

Figure 3: Comparison of the formant bandwidth of British, Australian and American accents (female)

4. COMPARISON OF FORMANT BANDWIDTH OF BRITISH, AUSTRALIAN AND AMERICAN ACCENTS

Average formant bandwidths of vowels and diphthongs are obtained from the weighted average of the mean of the

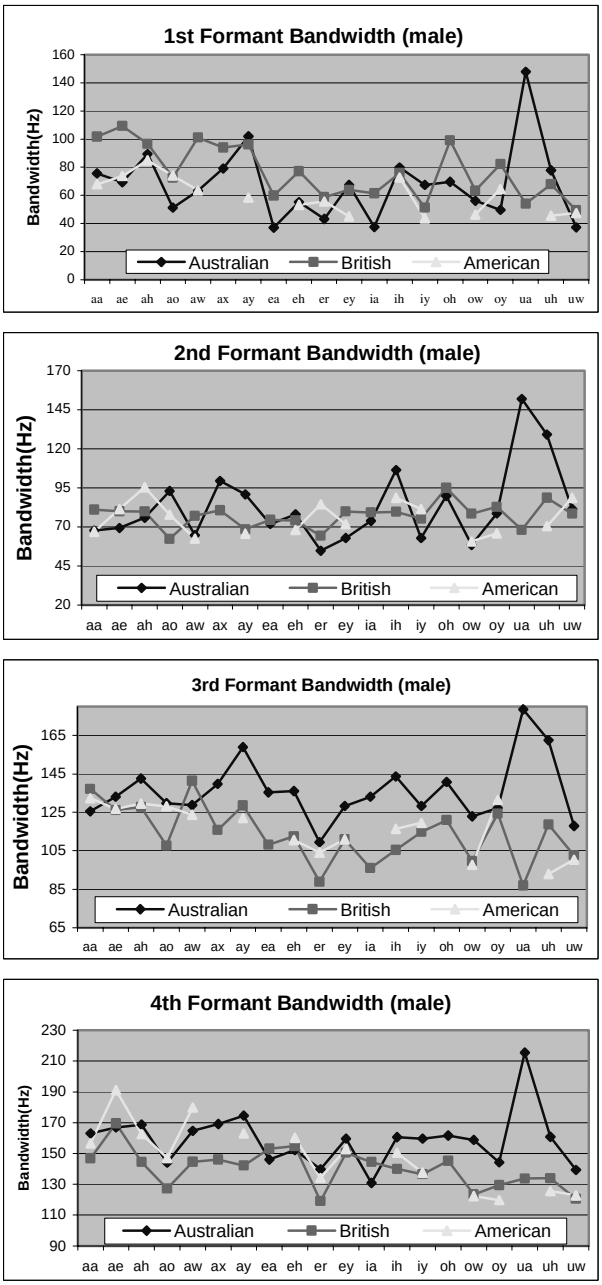

Figure 4: Comparison of Average formant bandwidth of vowels in British, Australian and American Accents (male)

Gaussian pdfs in the states of 2D HMMs associated with each formant. The component the Gaussian mixture model with the largest variance is not used in the summation of the weighted means from 2D-HMM, as this component is associated with the formant candidates that cannot be clearly associated with one formant.

Accent Pair	Formant Ranking Order			
	1	2	3	4
British & Australian	1st	2nd	4th	3rd
British & American	2nd	1st	3rd	4th
Australian & American	2nd	1st	3rd	4th

Table 3: Ranking of formant correlates of British, Australian and American accents. Importance is ranked from 1 (high) to 4 (low).

Figure 3 and 4 illustrate a comparison of average formant bandwidths of British, Australian and American accents. For the 2nd, 3rd and 4th formants, Australian has consistently larger average formant bandwidths than British. This pattern is much clearer in female speakers than male speakers. There is no explicit pattern for 1st formant bandwidth across accents. Furthermore, American female has higher average formant bandwidths than British and Australian female while it is not the case in male speakers. In addition, the average formant bandwidths of vowels and diphthongs of the female speakers are constantly higher than those of male speakers. This complies with the results by Childers[8].

5. RANKING OF FORMANT CORRELATES

In [6] it is shown that the second formant of speech usually has the largest frequency range compared to other formants. In order to assess the importance of each formant on conveying an accent *A* compared to a reference accent *B* the formants are ranked according to some distance measure. A simple formulae for ranking the formant of accents *A* with reference to the formants of accent *B* is proposed as

$$\underset{i}{Rank}\left(\sum_{v=1}^{V}\left[\frac{F_{vi}^{A}-F_{vi}^{B}}{0.5\left(F_{vi}^{A}+F_{vi}^{B}\right)}\right]^{2}\right) \qquad (1)$$

Where $Rank(\cdot)$ can be a sorting function that sorts the formants in increasing or decreasing importance, F_{vi}^{A} is the average gender-dependent i^{th} formant of the vowel v from accent A, V is the number of vowels. Average formants of the vowels of each pair of accents are used in Eq. (1) to obtain a ranking of formants influence in conveying accents. This formula is applied in other accent pairs as well (such as broad Australian and British, American and broad Australian). The experimental results rank the 2nd and the 1st formants as the most important two formants for accents, which are also consistent with perceptual evaluation in [6].

6. CONCLUSTION

This paper described a comparative analysis of the formant trajectories and bandwidth of British, Australian and American accents. A method of ranking formants based on normalized distances between formants are introduced. The 2nd formant is considered most important to conveying the difference between accents.

In future work, more efforts will be made on further investigation of the influence of the formant bandwidth to accents and refinement of the formant ranking formulae including the formant bandwidth.

7. ACKNOWLEDGEMENTS

We wish to thank the UK's EPSRC for funding project no GR/M98036.

8. REFERENCE

[1] Wells J.C., *Accents of English*, Cambridge University Press, (1982).

[2] Watson C., Harrington J., Evans Z., "An Acoustic Comparison between New Zealand and Australian English Vowels", *Australian Journal of Linguistics* (1996)

[3] Harrington J., Cox F., Evans Z., "An Acoustic Phonetic Study of Broad, General, and Cultivated Australian English Vowels", *Australian Journal of Linguistics* 17: 155-184 (1997)

[4] Arslan L., Hansen J., "A Study of Temporal Features and Frequency Characteristics in American English Foreign Accent", *Journal of Acoustic Society of America*, vol. 102(1), p. 28-40, (1997)

[5] Ho Ching-Hsiang, "Speaker Modelling for Voice Conversion", PHD thesis, Department of Electronic and Computer Engineering, Brunel University (2001)

[6] Yan Q., Vaseghi S. "Analysis, Modelling and Synthesis of Formant Spaces of British, American and Australian Accents", ICASSP pp. 712-715 (2003)

[7] Harrington. J., Cassidy, S. "Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in Australian English", Language and Speech 37 p357-373 (1994)

[8] Childers D.G., Wu K., "Gender Recognition From Speech. Part II: Fine Analysis". *Journal of Acoustic Society of America*, vol. 90, p.1841-1856, (1991)