ANALYSIS, MODELLING AND SYNTHESIS OF FORMANTS OF BRITISH, AMERICAN AND AUSTRALIAN ACCENTS

Qin Yan Saeed Vaseghi

Department of Electronic and Computer Engineering
Brunel University, Middlesex, UK UB8 3PH
(Qin.Yan, Saeed.Vaseghi)@brunel.ac.uk

ABSTRACT

The formant space of three major English accents namely British, American and Australian are modelled and used for accent conversion. Accent synthesis, through modification of the acoustic parameters of speech, provides a means for assessing the perceptual contribution of each parameter on conveying an accent. An improved method based on a linear prediction (LP) model feature analysis and a 2-D hidden Markov model (HMM) is employed for estimation of formant trajectories of vowels and diphthongs. Comparative analysis of the formant space of the three accents indicates that these accents are partly conveyed by the fronting and backing of vowels. It is found that the first formants of the vowels of British and American English accents are higher than those in Australian accent while Australians have higher second formants in vowels compared to Americans and British. The estimates of the distributions of formants for each accent are used in a speech synthesis system for accent conversion. Perceptual evaluations of accent conversion results illustrate that formants, in particular the second formant, play an important role in conveying accents.

1. INTRODUCTION

Accents are differences in pronunciation by a community of people from a national or regional geographical area, or a social grouping. Accents are affected by differences in the phonetic transcriptions and the acoustic correlates of speech, including formants and their trajectories, pitch trajectory, pitch nucleus and duration parameters [1].

Accent is one of the main factors that impact automatic speech recognition (ASR) and text-to-speech synthesis (TTS). Input speech with a different accent from that of the speech models can result in a significant deterioration in ASR performance, hence accent identification and modelling are essential for robust speech recognition [2]. Similarly accent models are useful for accent morphing in text-to-speech synthesis.

The acoustics of accent are due to the differences in the configurations, positioning, tension and movement of laryngeal and supra-laryngeal anatomical parameters. For example, in [3] Arslan and Hansen point out that generally non-native speakers do not produce the same tongue movement as native speakers, but produce accented sounds based on learned habits of tongue movements of their native language, which implies that their formants move with native language pronunciation. The difference in pitch and pitch trajectories in British and American English accents are analysed and presented in [4]. Recently, Harrington and Watson [5,6] explored 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.

The focus of this paper is on the mapping and synthesis of the formant space of British, American and Broad Australian accents. The synthesis of accents based on formant models provides a method of assessing the influence of each formant and its trajectory in conveying accent The databases employed are ANDOSL for Australian English, WSJCAM0 for British English and WSJ for American English.

2. COMPARISON OF FORMANTS OF BRITISH, AMERICAN, AND AUSTRALIAN ACCENTS

Although automatic formant analysis of speech has received considerable attention and a variety of approaches have been developed, the calculation of accurate formant features from the speech signal is still considered a non-trivial problem. The accuracy of formant tracking using the conventional frame-based LPC analysis is affected by following factors [7]:

1) Influence of pitch on the first formant.
2) Formant movements resulting in the merging of the trajectories of adjacent formants.
3) Rapid formant variation that may occur in consonant vowel transitions or diphthongs.
4) Source-vocal tract interaction (ignored in LPC analysis)
5) Effects of lips radiation and internal loss on formant bandwidth and frequency.

In the next section an improved method is suggested to tackle the first three factors in LPC analysis.

2.1 Formant Estimation

Formant classification is described in [8,9]. Each formant feature vector $[F_k, BW_k, I_k, \Delta F_k, \Delta BW_k, \Delta I_k]$ 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. A 2-D HMM with 3 left–to-right states across time and four left–to-right states across frequency is used to classify formant candidates in each frame among four sequential formant clusters. Given a set of training data, the distribution of each formant vector in each state is modelled by a multi-variate mixture Gaussian distribution trained using the EM algorithm. Formant tracks are obtained using a Viterbi search methods to find the most likely path of formants given the HMMs [8,9]. Figure 1 shows a block diagram illustration of formant estimation procedure. Pre-emphasis is applied to eliminate the pitch effect on the first formant.

Figure 2 shows the histograms of formant distributions of the vowel /IY/ from an Australian speaker. Each peak represents a formant. It can be noted in figure 2-(a) that the glottal formant due to the pitch effect (the first peak) could be mistaken for the first formant (the second peak) when there is no pre-emphasis, while in figure 2-(b) the glottal formant is eliminated by pre-emphasis. In [6], it is observed that rapid formant variation across phoneme boundaries is the dominant factor affecting the accuracy of formant estimation in continuous speech. To reduce these effects, three additional rules are applied as follows:

1) Discard very short phonetic segments.
2) Place a lower limit on the bandwidth of formant candidates and decrease LPC model order to avoid over-modelling
3) Only use formant candidates from the frames in the central (i.e. target) part of phoneme segments.

The idea behind this is to make use of the steadiest part (target) of formants in each vowel. In Figure 2-(a)(b)(c) the hump around 1700Hz is easily mistaken for the 2^{nd} formant although /IY/ does not have any formant in that frequency range. After improvement the hump disappears in Figure 2-(d) and the second and third formant become clearer. Formant frequencies are obtained eventually by averaging the central part of formant trajectory.

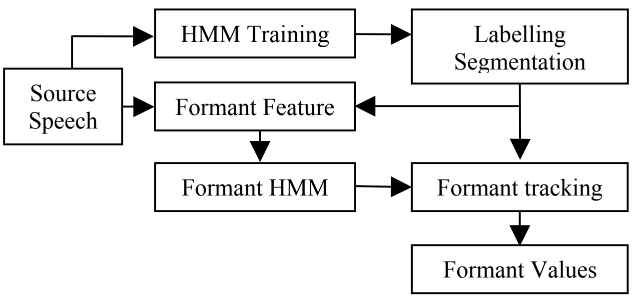

Figure 1: Block Diagram of Formant Estimation

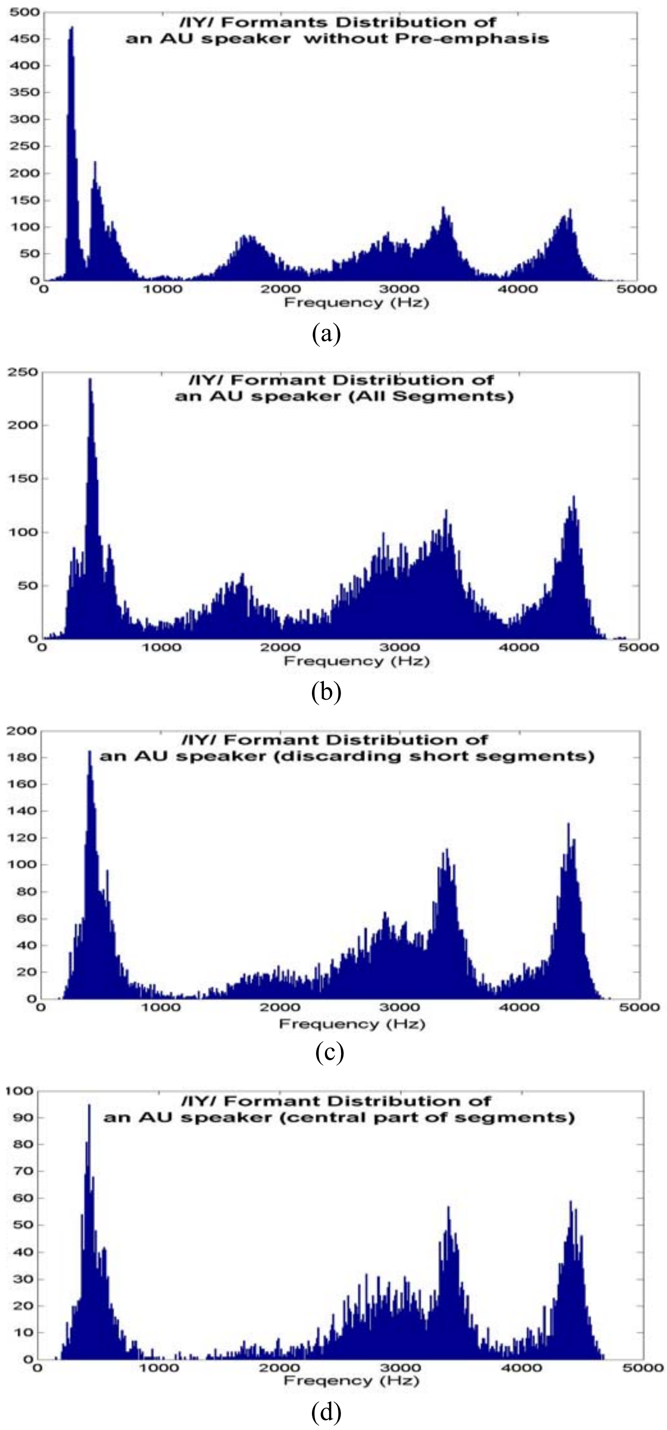

Figure 2: Histograms of Formant Distributions of /IY/ from an Australian Speaker

(a) Without Pre-emphasis,
(b) With Pre-emphasis,
(c) Discarding short segments with limited bandwidth and LP order
(d) Take the central part of segments with limited bandwidth and LP order

2.2 Formant Comparison

Figure 3 shows the average of the first, second, third and fourth formants of Australian, British and American accents. Except for the vowels /aa/ /ah/ /oh/, other Australian vowels have lower F1 than the British. American vowels display higher F2 than Australian except for /er/. On average, 2nd formants of vowels in Australian are 11% higher than those of British and 8% higher than those of American. In formants F3 and F4, Australians consistently displays higher formant frequencies than British. American speakers also have higher F3 and F4 than British speakers except for /er/. The 2nd formant is the most dynamic. It has the widest frequency range up to 2KHz. Male speakers from these accents illustrate a similar set of patterns to females. In phonetics, vowels front or back movements are regarded as correlated with F2 while high and low movements are associated with F1. Figure 4 5 illustrate formant spaces of the three accents. It can be noticed that compared to British and American:

1) /ae/ and /eh/ in Australian are raised
2) /uw/ and /aa/ in Australian are fronted
3) /iy/ and /ih/ in Australian are closer

Besides, the American accent has the lowest /ao/ compared to British and Australian. It can be concluded that formants play central role in conveying different English accents. In particular, the second formant is considered to be most affected by accents.

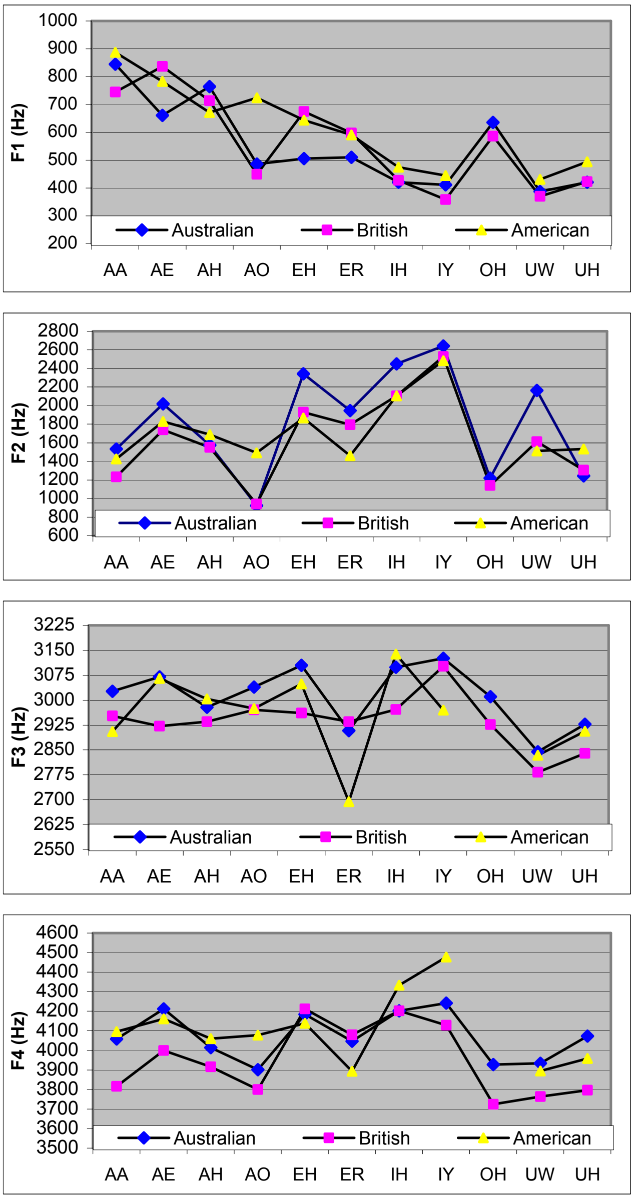

Figure 3: Comparison of Formants of Australian, British and American (female)

3. ACCENT CONVERSION

To evaluate the importance of the impact of formants in conveying an accent, a set of experiments are conducted

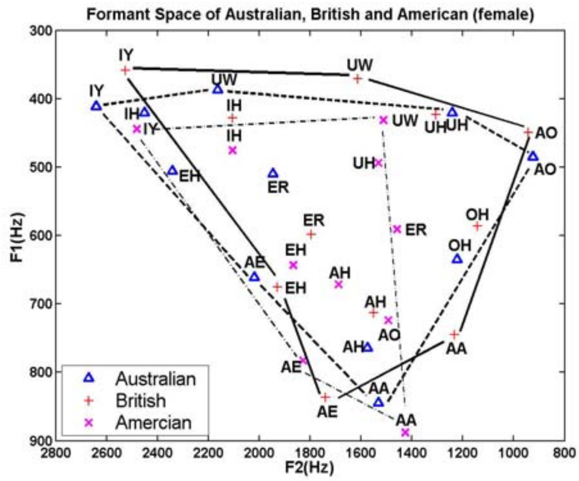

Figure 4: Formant Space of Australian, British and American (female)

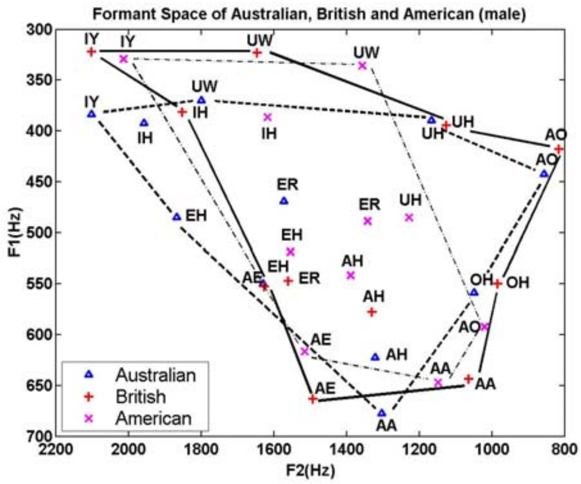

Figure 5: Formant Space of Australian, British and American (male)

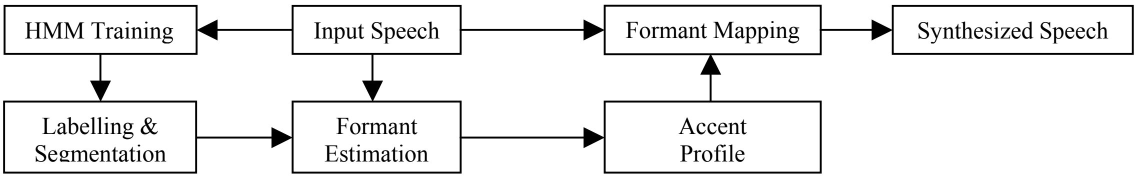

Figure 6: Diagram of Speech Accent Synthesis based on Formant Modification.

to convert the accent of recorded speech through appropriate modifications of formants. Perceptual tests are performed to evaluate the influence of each formant on accent. In one set of experiments formant frequencies of vowels in the utterance from Australian speakers are adjusted via LPC spectral frequency mapping to the average formant frequencies of British speakers according to formant estimation above. Figure 6 displays the diagram of synthesis experiments. The accent profile block is a complete description of acoustic correlates of accent parameters such as formant and their trajectories, pitch and its trajectory and duration. Results show that formant shifting exerts an important impact on accent. Particularly by shifting the first and second formant, it is agreed among the listeners that converted utterance carries a certain amount of British accent. On the other hand, Australian accent still can be heard since only formants positions have been modified.

4. DISCUSSIONS AND CONCLUSION

This paper explores the formant space of three major English accents: British, American and Australian. Accuracy of formant estimation based on traditional LPC analysis increases after excluding the formant boundary and pitch effects. Results shows that Australian vowels have lower first formants and higher second formants compared to British and American. Also the second formant has the widest frequency range. Perceptual experiment results demonstrate that formants play an important role in accent. The 2nd formant is regarded as the most dynamic and influential formant for conveying accents.

For further improvements, other correlates of accent such as pitch nucleus, pitch and trajectories and duration will be included in accent morphing to achieve better results.

5. ACKNOWLEDGEMENTS

We thank Dimitros Rentzos for the formant mapping programs and Ching-Hsiang Ho for his help in formant models and the UK's EPSRC for funding project no GR/M98036.

6. REFERENCE

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

[2] Humphries J., "Accent Modelling and Adaptation in Automatic Speech recognition", PhD Thesis, Cambridge University Engineering Department (1997)

[3] Arslan L. M., Hansen H., "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)

[4] Yan Q., Vaseghi S., "A Comparative Analysis of UK and US English Accents In Recognition and Synthesis", *ICASSP,* Orlando (2002)

[5] 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)

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

[7] 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)

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

[9] Acero A, "Formant Analysis and Synthesis Using Hidden Markov Models", *Proc. of the Eurospeech Conference*, Budapest (1999)