COMPARATIVE ANALYSIS OF FORMANTS OF BRITISH, AMERICAN AND AUSTRALIAN ACCENTS

Seyed Ghorshi *Saeed Vaseghi* *Qin Yan*
School of Engineering and Design, Brunel University, London
{Seyed.Ghorshi, Saeed.Vaseghi, Qin Yan}@brunel.ac.uk

10.21437/Interspeech.2006-35

ABSTRACT

This paper compares and quantifies the differences between formants of speech across accents. The cross entropy information measure is used to compare the differences between the formants of the vowels of three major English accents namely British, American and Australian. An improved formant estimation method, based on a linear prediction (LP) model feature analysis and a hidden Markov model (HMM) of formants, 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 mostly conveyed by the first two formants. The third and fourth formants exhibit some significant differences across accents for only a few phonemes most notably the variants of vowel '*r*' in the American (rhotic) accent compared to British (non-rhotic accent). The issue of speaker variability versus accent variability is examined by comparing the cross-entropies of speech models trained on different groups of speakers within and across the accents.
Index Terms: accent, formants, cross entropy, speech recognition.

1. INTRODUCTION

The modelling and measurement of accents is useful in a variety of speech processing applications such as accent identification, accent morphing, multi-accent text to speech synthesis, and speech recognition.

Accent is one of the most fascinating aspects of speech acoustics [1]. The term *accent* may be defined as a distinctive pattern of pronunciation, including lexicon and intonation characteristics, of a community of people who belong to a national, regional or social grouping. It is worthwhile to clarify the similarities and the differences between two closely linked linguistic terms, namely accent and dialect. The term dialect refers to the whole speech pattern, conventions of vocabulary, pronunciation, grammar, and the usage of speech by a community of people [1] while accent refers to a pattern of pronunciation, i.e. the use of vowels or consonants, particular rhythmic forms in intonation, stress patterns and other prosodic features and the abstract (phonological) representations which can be seen as underlying the actual (phonetic) articulation.

An accent is usually associated with a community of people with a common regional, socioeconomic or cultural background. Accents evolve over time influenced mainly by large immigrations and social and cultural trends as well as the mass media. For example, the Australian accent is considered to have been influenced by the waves of mass immigrations to Australian and in particular by London "Cockney" accent, Irish accent and relatively recently by American accent. Similarly, the English Liverpool accent has been influenced by the Irish immigration whereas the Northern Ireland accent has been influenced by the Scottish immigration.

In general, there are two broad approaches to classification of the differences between accents:

- *Historical approach to accent development*. Compares the historical roots of accents and the evolutionary changes in sounds that accents have gone through as various accents merge or diverge. The historical approach compares the rules of pronunciation in accents and how the rules change and evolve over time.
- *Structural, synchronic approach*, first proposed by Trubetzkoy [2] models an accent in a system-oriented fashion in terms of the following systematic differences:
- Differences in phonemic systems.
- Differences in phonotactic (structural) distributions.
- Differences in lexical distributions of words.
- Differences in phonetic (acoustic) realization.

In this work the influences of accents on formants of vowels of speech are investigated.

The databases employed in this work for accent analysis are Australian National Database of Spoken Language (ANDOSL) for Australian English, Wall Street Journal Database Cambridge University (WSJCAM0) for Received Pronunciation British English and Wall Street Journal (WSJ) database for general American English. The subset of ANDSOL of (broad, general and cultivated) Australian accent consists of 18 female and 18 male speakers with a total of 7200 utterances in each category. The subset of WSJ database used for modeling American English contains 36 female and 38 male speakers with 9438 utterances. The subset of WSJCAM0 of British English used contains 40 female and 46 male speakers with 9476 utterances. The style of speech in all databases is read (as opposed to conversational) speech.

The focus of this paper is on the mapping and comparison of the formant space of American, British and Australian accents. The formant models provide a method of assessing the influence of each formant and its trajectory in conveying accent.

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 [3].

1) Influence of the spectral peak due to the glottal vibrations 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 LP analysis).
5) Effects of lips radiation and internal loss on formant bandwidth and frequency.

2.1 Formant Estimation

Formant estimation and classification is described in [4, 5]. Each formant feature vector has 6 parameters [F_k, B_k, I_k, ΔF_k, ΔB_k, ΔI_k]: formant frequency F_k, bandwidth B_k, and intensity I_k together with the slopes of their time trajectories ΔF_k, ΔB_k and ΔI_k. A two-dimensional HMM [4, 5], 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 modeled by a multi-variate mixture Gaussian distribution trained using the EM algorithm. Formants tracks are then obtained using a Viterbi search methods to find the most likely path of formants given HMMs [4, 5]. Figure 1 shows a block diagram illustration of formants estimation procedure. Pre-emphasis is applied to eliminate the pitch effect on the first formant. The average formant frequencies of female speakers of American, British and Australian accents are obtained from HMMs of formants.

2.2 Formant Comparison

Figure 2 show the average of first, second, third and fourth formants of Australian, British and American accents. It can be seen that British have higher *F1* than Australian except for vowels /aa/, /ah/, /iy/, /oh/ and /uw/. Americans have a lower *F2* than Australians except for vowels /ah/, /ao/, /iy/ and /uh/. On average, Australian have higher *F3* and *F4* than British and American. British also displayed higher *F3*, *F4* than American except for vowels /ae/, /ah/ and /uw/ in *F3* and *F4* and /iy/ in *F4* only. Male speakers from these accents illustrate a similar set of patterns to females. In phonetics, vowels front and back movements are regarded as correlated

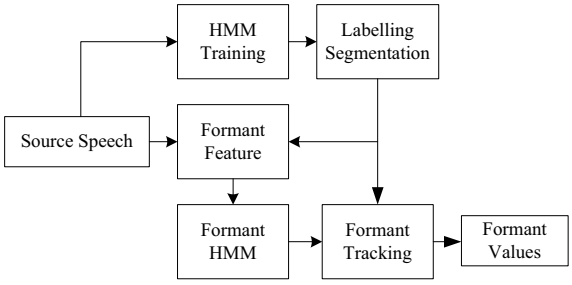

Figure 1: Block Diagram of Formant Estimation.

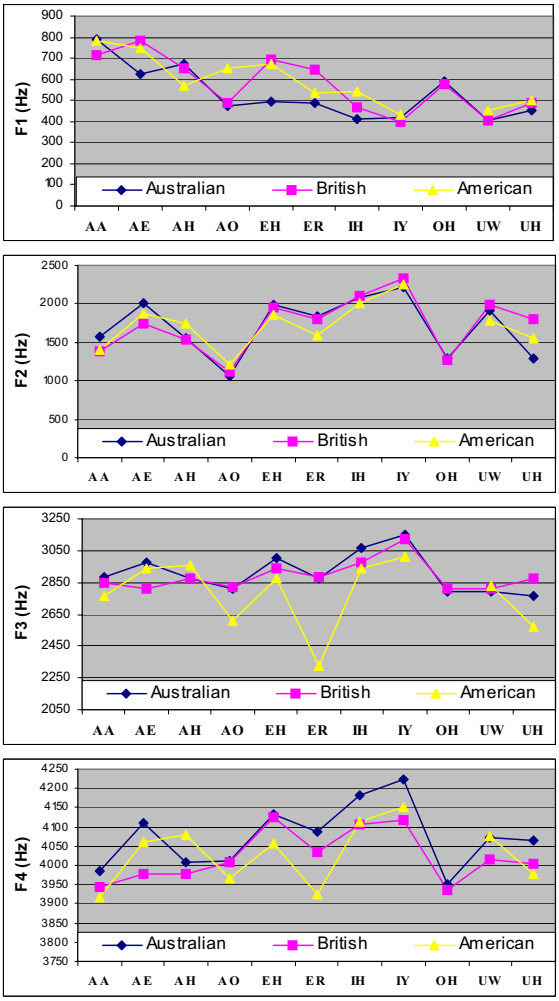

Figure 2: Comparison of Formant of Australian, British and American Accents for Female Speakers.

with *F2* while high and low movements are associated with *F1*. Figures 3 and 4 illustrate the *F1* versus *F2* and *F3* versus *F4* formant spaces of the three accents. It can be noticed that the distances between formants are particularly high for some vowels. For example British and Australian /ao/ have a relatively large distance from American /ao/, American /er/ has a large distance in *F3* and *F4* from British and Australian, the vowels /iy/ and /ih/ in Australian are closer compared to British and American and /er/ and /r/ in American are closer compared to Australian and British.

Figure 3 also shows that /eh/ and /er/ in Australian are raised compared to British and American. Besides, /r/ in American is fronted in Figure 4. It can be concluded that formants play an important role in conveying the difference between English accents.

3. CROSS ENTROPY ACCENT METRIC

A suitable choice for an accent metric should be able to measure the systematic differences in the pronunciations across different accents and also remove the effect of the differences due to the

speakers' characteristics. A measure of the differences in the pronunciation patterns of words in two accents may be defined by measuring the changes due to insertions, deletions or substitutions of phonemes in each word as well as the changes in the phonetic realization of phonemes and the effect of accent in stress and intonation characteristics of syllables and phrases. Even at the relatively simple level of the differences in the phonemic pronunciation and acoustic-phonetic realizations of words in different accents, an accent metric must be able to quantify the effects of a whole set of changes ranging from relatively subtle differences in acoustic realization of a phoneme to more obvious changes due to substitution, deletion and insertion of phonemes.

3.1 Cross Entropy of Accents

Cross entropy is a measure of the difference between two probability distributions [6]. There are a number of different definitions of cross entropy. The definition used here is also known as Kullback-Leibler distance. Given the probability models P_1(x) and P_2(x) of a formant, or a phoneme, or some other speech feature or unit in two different accents, measures of their differences are the cross entropy of accents defined as:

$$CE(P_1,P_2) = \int_{-\infty}^{\infty} P_1(x)\log_2\frac{P_1(x)}{P_2(x)}dx \qquad (1)$$

Note that the integral of $P(x)$ log $P(x)$ is also known as *the differential entropy*. The cross entropy is a non-negative function. It has a value of zero for two identical distributions and it increases with the increasing dissimilarity between two distributions [6, 7]. The cross entropies between two different left-right N-state HMMs of speech with M-dimensional (formant) features is computed as the sum of cross-entropies of their respective states obtained as

$$CE(P_1,P_2) = \sum_{s=1}^{N}\sum_{i=1}^{M}\int_{-\infty}^{\infty} P_1(x_i \mid s)\log_2\frac{P_1(x_i \mid s)}{P_2(x_i \mid s)}dx_i \qquad (2)$$

where $p(x_i|s)$ is the probability distribution of the i^{th} mixture of speech in state s. Cross entropy is asymmetric $CE(P_1,P_2)\neq CE(P_2,P_1)$. A symmetric cross entropy measure can be defined as

$$CE_{sym}(P_1,P_2) = \big(CE(P_1,P_2)+CE(P_2,P_1)\big)/2 \qquad (3)$$

In the following the cross entropy distance refers to the symmetric measure and the subscript *sym* will be dropped. The total distance between two accents can be defined as

$$AccDist = \sum_{i=1}^{N_U} P_i CE\big(P_1(i),P_2(i)\big) \qquad (4)$$

where N_u is the number of speech units and P_i the probability of the i^{th} speech unit. The cross-entropy distance can be used for a wide range of purposes including:

(a) To calculate the differences between two accents or the voices of two speakers.

(b) To cluster phonemes, speakers or accents.

(c) To rank voice or accent features.

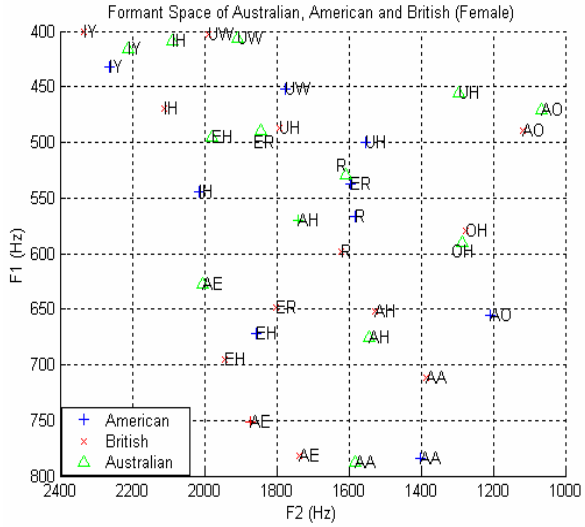

Figure 3: *F1/F2* space of Australian, British and American.

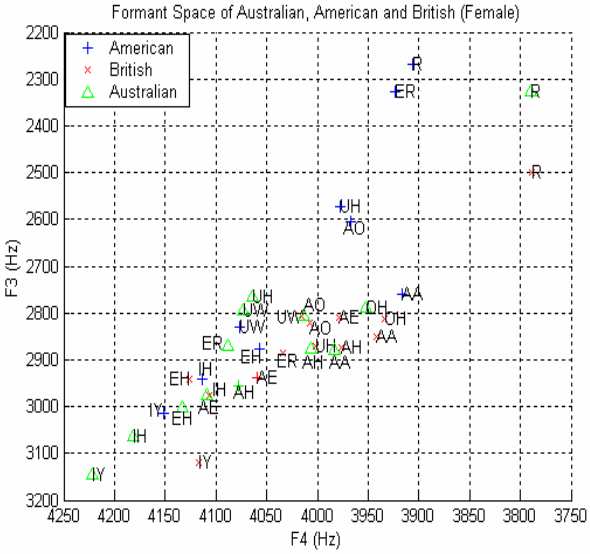

Figure 4: *F3/F4* space of Australian, British and American.

4. CROSS ENTROPY QUANTIFIACTION OF ACCENTS OF ENGLISH

In this section we describe experimental results in application of cross entropy for quantification of the influence of accents on the formants' of vowels. The plots in Figure 5 illustrate the result of measurements of inter-accent and intra-accent cross entropies of speech models. Eighteen speakers were used to obtain each set of models for each group in each accent. The result clearly shows that in all cases the inter-accent model differences are significantly greater than the intra-accent model differences. Furthermore, the results show that in all cases the differences between Australian and British are less than the distance between American and British (or Australian).

The closeness of Australian and British accents in comparisons to American accent is also supported by cross-accent speech recognition results. The speech recognition results

for varying accents of models and test data, shown in tables 1 and 2, are obtained from phoneme-dependent HMMs trained on 39 dimensional cepstral features including delta and delta delta cepstrum. The results show that on average cross accent speech recognition between Australian and British yields about 25% less error than between Australian and American or British and American. These results are consistent with the results in Figure 5 which shows that formants of British and Australian accents are closer to each other than to those of American. The results of Figure 5 also reveal that the distance of models trained on different speaker groups is much higher across accents than within accents.

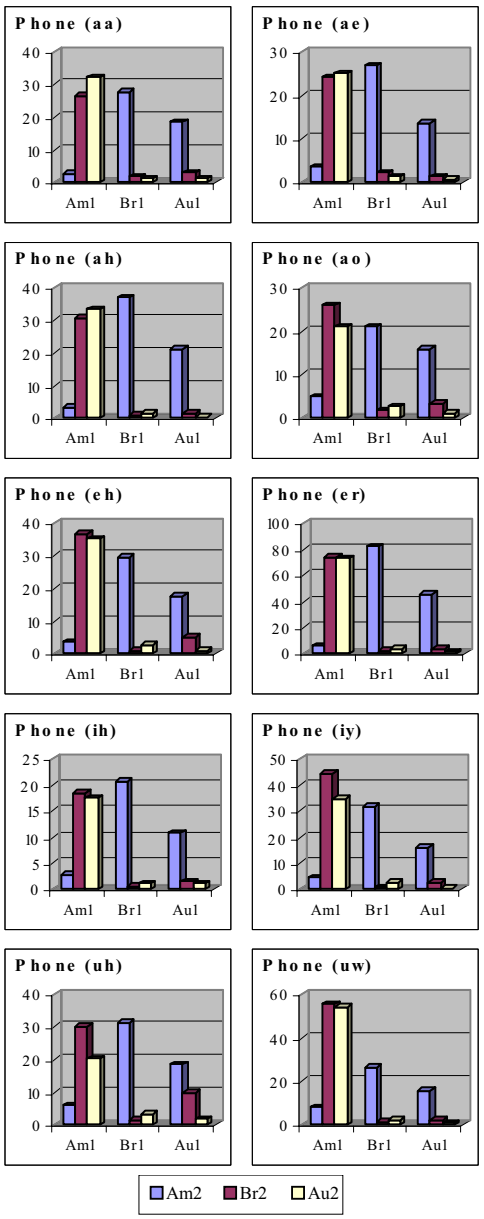

Figure 5: Plots of inter-accent and intra-accent cross entropies of a number of phonemes of American, British and Australian accents. Note each colour-keyed column shows the cross entropy of a group of one speech accent from another indicated on the horizontal axis.

MODEL / INPUT	Br	Am	Au
Br	**30.1**	53.7	42.3
Am	51.3	**33.6**	53.0
Au	41.8	51.6	**29.0**

Table 1: The effect of accent on the (%) error rate of automatic speech recognition accuracy (Female Speakers).

MODEL / INPUT	Br	Am	Au
Br	**33.1**	53.4	43.4
Am	51.3	**34.8**	51.9
Au	45.4	51.1	**31.9**

Table 2: The effect of accent on the (%) error rate of automatic speech recognition accuracy (Male Speakers).

5. CONCLUSIONS

The formant space of three major English accents namely British, Australian and American are compared. A 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. Results show that the formants of the vowels play an important role in conveying the difference between English accents. Furthermore the cross entropy is applied for quantification of the effect of accents on formants. The cross entropy is used to investigate the effect of accent and speaker variability by measuring the differences on models trained on speaker groups within accents and across accents. It is clear that the accent variability is much greater than speaker variability.

6. REFERENCES

[1] J. C. Wells, *"Accents of English,"* Volume: 1,2 Cambridge University Press, 1982.

[2] N. S. Trubetzkoy (1931), "*Phonologie et geographie linguistique" Travaux du Cercle Linguistique de Prague* 4.pp.228-234

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

[4]Ho Ching-Hsiang, "Speaker Modeling for Voice Conversion", PhD thesis, School of Engineering and Design, Brunel University (2001).

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

[6] J. E. Shore and R. W. Johnson,"Properties of cross-entropy minimization," *IEEE Trans. Inform. Theory*, vol. IT-27, pp.472-482, July. 1981.

[7] E.T. Jaynes, "On the rationale of maximum entropy methods," *Proc. IEEE,* vol. 70, pp. 939-952, Sep. 1982.