

A COMPARATIVE ANALYSIS OF UK AND US ENGLISH ACCENTS IN RECOGNITION AND SYNTHESIS

Qin Yan, Saeed Vaseghi

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

ABSTRACT

In this paper, we present a comparative study of the acoustic speech features of two major English accents: British English and American English. Experiments examined the deterioration in speech recognition resulting from the mismatch between English accents of the input speech and the speech models. Mismatch in accents can increase the error rates by more than 100%. Hence a detailed study of the acoustic correlates of accent using intonation pattern and pitch characteristics was performed. Accents differences are acoustic manifestations of differences in duration, pitch and intonation pattern and of course the differences in phonetic transcriptions. Particularly, British speakers possess much steeper pitch rise and fall pattern and lower average pitch in most of vowels. Finally a possible means to convert English accents is suggested based on above analysis.

1. INTRODUCTION

In the recent years, there have been significant advances in speech recognition systems resulting in reduction in the error rate. Two of the most important remaining obstacles to reliable high performance speech recognition systems are noise and speaker variations. An important aspect of speaker variation is accent. However, current speech recognisers are trained on a specific national accent group (e.g. UK or US English accents), and may have a significant deterioration in performance when processing accents unseen in the training data. An understanding of the causes and acoustic properties of English accents can

also be quite useful in several areas such as speech synthesis and voice conversion.

In [3] J.C. Wells described the term accent as a pattern of pronunciation used by a speaker for whom English is the native language or more generally, by the community or social grouping to which he or she belongs. Linguistically, accent variation does not only lie in phonetic characteristics but also in the prosody.

There has been considerable research conducted on understanding the causes and the acoustics correlates of native English accent. A study in [3] examined a variety of native English accents from a linguistics point of view. Recently more focused studies have been made on acoustic characteristics of English accents. In [4] a method is described to decrease the recognition error rate by automatically generating the accent dictionary through comparison of standard transcription with decoded phone sequence. In [1], rather than using phonetic symbols, different regional accents are synthesized by an accent-independent keyword lexicon. During synthesis, input text is first transcribed as keyword lexicon. Until post-lexical processes, accent dependent allophonic rules were applied to deal with such features as /t//d/ topping in US English, or r-linking in British English. The advantage of this method is that it avoids applying different phonetic symbols to represent various accents. In addition, [2] established a voice conversion system between British and US English accents by HMM-based spectral mapping with set rules for mapping two different phone sets. However, it still has some residual of original source accent characteristics in the converted result.

In this paper, experiments began with cross accent recognition to quantify the accent effects between British accent (BrA) and American accent (GenAm) on speech

recognition. A further detailed acoustics feature study of English accent using duration, intonation and frequency characteristics was performed.

2. CROSS ACCENT RECOGNITION

At first, a set of simple experiments was carried out to quantify the effect of accents on the speech recognisers with accent specific dictionaries. The model training and recogniser used here are based on HTK [9]. British accent speech recogniser was trained on Continuous Speech Recognition Corpus (WSJCAM0). American accent speech recogniser was trained on WSJ. Test sets used are WSJ si_dt_05 si_et_05 and WSJCAM si_dt5b, each containing 5k words. Both recognisers employ 3-state left-to-right HMMs. The features used in experiments were 39 MFCCs with energy and their differentiation and acceleration.

Accent	British model	American model
British input	12.8	29.3
American input	30.6	8.8
Average	21.7	19.1

Table 1: % word error rate of cross accents speech recognition between British and American accent

Table 1 shows that for this database the American English achieves 31% less error than the British English in matched accent conditions. Mismatched accent of the speaker and the recognition system deteriorates the performance. The result was getting worse by 139% for recognizing British English with American models and 232% for recognizing American English with British models. The results are based on word models compiled from triphone HMMs with three states per model and 20 mixtures Gaussians per state.

The next section examines the acoustics features of both English accents in an attempt to identify where the main difference lies in addition to the variation in pronunciation.

3. ANALYSIS OF ACOUSTIC FEATURES OF US AND UK ENGLISH ACCENTS

3.1 Duration

Figure 1 shows that the vowel durations at the start and the end of sentences in BrA is shorter than that in GenAm. This could be due to the following reason.

British speakers always tend to pronounce last syllable fast. It is the case especially for consonants. However, Americans tend to realize more acoustically complete pronunciation.

Table 2 gives the comparison of two database in speaking rate. The speaker rate of Wsjcam0 is 7.8% higher than that of Wsj. This is in accordance with comparison in phone duration in Figure 1.

The results of these comparisons are shown in Figure 1. Note that results are only presented for models common to both system phones sets.

Speak Rate (no/sec)	Phone	Word
Wsjcam0	9.77	3.04
Wsj	10.39	2.82

Table 2 : Speak rate in Phone and word from Wsjcam0 and Wsj

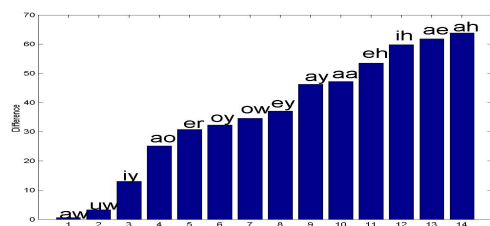


Figure 1: Difference of Vowel duration of GenAm and BrA at the utterance starts and ends

3.2 Pitch Characteristics

Table 3 and 4 list average pitch values and numbers of speakers from both databases. Figure 2 displays the difference of average vowel pitch frequency of male speakers of two accents while Figure 3 shows the corresponding comparison of female speakers. Even BrA has lower average pitch than GenAm in the whole phone set, for the common vowels, their average pitch in BrA is still much more lower than corresponding part in GenAm. It is interesting to note that for most of vowels, British speakers give lower pitch than American counterparts. For British female speakers, its 118% lower than American female in average while it drops down to 7.7% when comparing with British male and American male in the common set vowels. In accordance with [4], diphthongs such as *ay uw er*, display more difference than other vowels. Furthermore, average pitch frequency of the last word of sentences from male speakers of both accents

also clearly demonstrate similar results that British speakers are generally speaking lower than their counterpart. Besides, it can be noted that British male speakers gain high average pitch in three vowels : *uh*, *ih* and *ae*.

Speaker No.	Male	Female
Wsjcam0	112	93
Wsj	37	41

Table 3: Number of speakers Wsjcam0 and Wsj

Avg Pitch	Male	Female
Wsjcam0	115.8 Hz	196.2 Hz
Wsj	127.8 Hz	208.9 Hz
Difference	9.4%	5.7%

Table 4: Average pitch of Wsjcam0 and Wsj

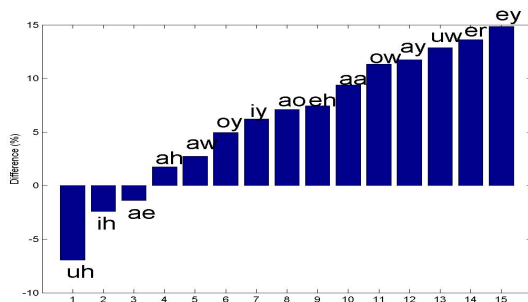


Figure 2: Difference of average pitch value of vowels of GenAm and BrA (male speakers)

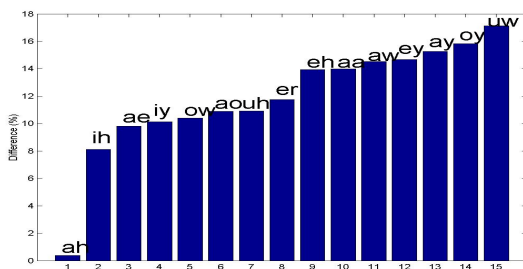


Figure 3: Difference of average pitch value of vowels GenAm and BrA (female speakers)

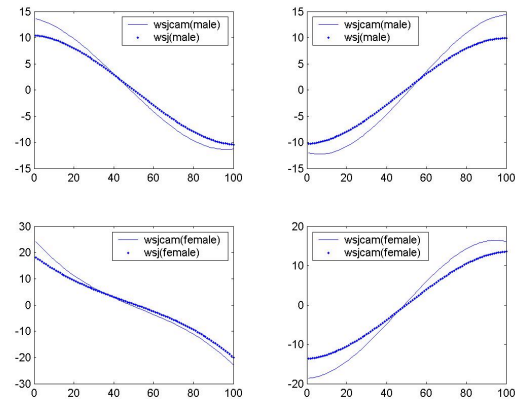


Figure 4(a)

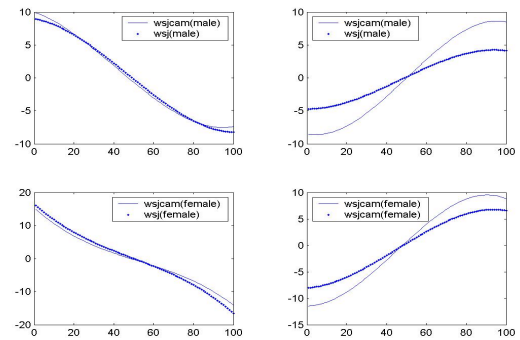


Figure 4(b)

Figure 4 (a): Average of Rise and Fall patterns from British and American speakers

Figure 4 (b): Average of Rise and Fall patterns of last word of the sentences

Xlabel: uniform duration (1.812ms), Ylabel: frequency

3.3 Prosody

Prosody is usually made up of *Intonation-groups*, *Pitch Event* and *Pitch Accent*. *Intonation-groups* are composed of a sequence of pitch events within phrase. *Pitch Event* is a combination of a pitch rise and fall. *Pitch accent*, either a pitch rise or a pitch fall, is the most elementary unit of intonation.

In [6], a rise fall connection (RFC) model was applied to model the pitch contour by Legendre polynomial function [a1, a2, a3], where a1, a2, a3, called discrete Legendre Polynomial Coefficients, were related to the average contour, average contour slope and average trend of the slope within that pitch accent. Rise and fall are detected according to f0 contour. Based on this, experiments were made on computing the average pattern of pitch accents

(Fall and Rise only in this case) to explore the numerical difference of both accents in intonation. Figure 4(a) illustrates the average of rise and fall patterns from both male and female speakers. It is noticeable that British speakers intend to have steeper rise and fall than American speakers. Particularly, for rise pattern, their difference in pitch change rate reaches 34% in average while fall pattern only gives 21% difference. In addition, it is also noticeable that pitch range narrows towards the end of an utterance as [8].

Further to the results that American speaker tends to speak lower in final words of sentences. Figure 4(b) indicates that BrA Rise pattern in the last words is much more steeper than that of GenAm with pitch change rate of 48% and 32% respectively.

In contrast, the fall pattern is almost same in either figure. Then British speakers possess much steeper pitch accent than American speakers.

5. DISCUSSIONS AND CONCLUSION

We have presented a detailed study of acoustic features about two major English accents: BrA and GenAm. In addition to the significant difference in phonetics, the slope of Rise and Fall accent also exhibits great difference. British speakers tend to speak with lower pitch but higher pitch change rate, especially in the rise accent. Future experiments are to be extended to other context-dependent pitch pattern analysis besides utterance end.

In general, accent conversion/synthesis could be simplified into two aspects: phonetics and acoustics. Beep dictionary and CMU dictionary explicitly display the phonetics difference between two accents in terms of phone substitute, delete and insert. In this paper, we began the exploration of acoustics difference between two accents in the view of duration, pitch and intonation pattern.

Therefore, the accent synthesis is planned to carry on by two steps for future experiments.

1) Pronunciation modelling by transcribing GenAm by BrA phones to map phonetic difference of two accents [4] or vice versa.

2) Prosody modification [7] [8]. By applying Tilt model base on decision-tree HMM, tilt parameters are changed according to above analysis. The advantage of Tilt model lies in its continuous tilt parameters, which better describe the intonation pattern than RFC models or FUJISAKI models [7]. A new pitch contour is then synthesized after changing tilt parameters according above study.

6 ACKNOWLEDGEMENTS

This research has been supported by Department of Computing and Electronic Engineering, Brunel University, UK. We thank Ching-Hsiang Ho for the program of detecting the pitch accents.

7. REFERENCE

- [1] Susan Fitt, Stephen Isard, *Synthesis of Regional English Using A Keyword Lexicon*. Proceedings Eurospeech 99, Vol. 2, pp. 823-6.
- [2] Ching-Hsiang Ho, Saeed Vaseghi, Aimin Chen, *Voice Conversion between UK and US Accented English*, Eurospeech 99.
- [3] J.C. Wells, *Accents of English*, volume:1,2, Cambridge University Press, 1982
- [4] Jason John Humphries, *Accent Modelling and Adaptation in Automatic Speech recognition*, PhD Thesis, Cambridge University Engineering Department
- [5] Alan Cruttenden, *Intonation*, Second Edition 1997
- [6] Ching-Hsiang Ho, *Speaker Modelling for Voice Conversion*, PHD thesis, Department of Computing and Electronic Engineering, Brunel University
- [7] Thierry Dutoit, *Introduction to text-to-speech synthesis*, Kluwer (1997)
- [8] Paul Taylor, *Analysis and Synthesis of Intonation using Tilt Model*, Journal of the Acoustical Society of America. Vol 107 3, pp. 1697-1714.
- [9] Steve Young, Dan Kershaw, Julian Odell, Dave Ollason, Valtcho Valtchev, Phil Woodland, *The HTK Book. V2.2*