DIPHTHONG MEASUREMENTS IN SINGAPORE ENGLISH

David Deterding

National Institute of Education
Nanyang Technological University

ABSTRACT - It is often stated that the vowels /eɪ/ and /oʊ/ in Singaporean English Pronunciation (SEP) are characterised by little or no movement, so that they may be regarded as long monophthongs. However, few measurements have previously been made to check such claims. Nine Singaporeans with a range of educational levels were recorded, and the movement of the first formant was measured for their /eɪ/ and /oʊ/ vowels. These measurements were compared with similar recordings of three British university lecturers and also some BBC broadcasters from a standard database. It was found that all the Singaporeans, regardless of educational level, do indeed tend to have more monophthongal /eɪ/ and /oʊ/ than the speakers of British English.

INTRODUCTION

Many scholars have described the characteristics of vowels in Singapore English Pronunciation (SEP), and most have observed that the vowels /oʊ/ (which is traditionally transcribed as /əʊ/ in British dictionaries) and /eɪ/ are often realized as long monophthongs (Platt and Weber, 1980; Tay, 1982; Brown, 1988; Deterding and Hvitfeldt, 1994). However, there has been little work on the measurement of these vowels to provide hard evidence for such claims. Improvements in computer speech analysis software allow us to attempt such measurements.

Based on listening judgements, Platt and Weber (1980:54) claim that nearly all SEP speakers use a long monophthong for /oʊ/, but that the pronunciation of /eɪ/ depends on educational and socio-economic backgrounds, with speakers of higher educational level tending to use more of a diphthong. In this paper, these results will be tested using computer-based formant measurements.

MEASUREMENT OF DIPHTHONGS

It is widely believed that vowel quality is related to the frequencies of formants, particularly the first two formants (F_1 and F_2), so that F_1 is related to the open-close dimension, while F_2 is related to the front-back dimension (though Ladefoged (1993:196) prefers to use the difference between F_2 and F_1 to represent frontness). It is therefore possible to calculate average values of F_1 and F_2 for the monophthongs of a variety of speakers (Cruttenden, 1994:95, quoting values from Deterding, 1990), and thereby derive an approximation to the traditional vowel quadrilateral.

However, we do not really know whether the frequency of these formants actually determines vowel quality, or whether there is just a correlation. In other words, if we find that a speaker produces two tokens of a particular vowel, and the frequency of F_1 is lower for the first token than the second, does this necessarily mean that the first vowel is more close? Actual vowel formant measurements suggest that there is in fact considerable variation in formants without there necessarily being any corresponding variation in vowel quality. Moreover the wide spread of formant values for the vowels from different speakers, even when they are spoken in citation form, must be explained if we are to assume that formant frequency determines vowel quality (Fry, 1979:111).

For diphthongs, the picture is even less clear. We might outline three possibilities for their description (Kent & Read, 1992:103). Firstly, they might be regarded as movement from a starting position (the *onglide*) to a final position (the *offglide*). This is the claim of Fry (1979:114), and it is maybe the underlying assumption of the IPA in using two vowel symbols to represent a diphthong. The problem with this is that a steady-state offglide position is often not achieved in real conversational speech. An alternative approach would be to say that a diphthong consists of an onglide with a trajectory towards the offglide, and the length of the trajectory depends on the duration of the vowel. However, even the onglide is not necessarily achieved in natural, fast speaking rates. The third possibility is to follow Gay (1968), who suggests that rate of change may remain constant even when there is considerable variation in both the onglide and the offglide, and this is particularly true for the /eɪ/ and /oʊ/ diphthongs, where an initial steady-state portion of the vowel may only be evident when the speaking rate is slow.

In this paper, where we are concerned with whether vowels are more or less diphthongal, the suggestions of Gay (1968) will be followed, and the rate of change of formants will be measured. As both /eɪ/ and /oʊ/ are primarily *closing* diphthongs in British English (as opposed to /ɪə/, /eə/ and /ʊə/ which are *centring* diphthongs) (Roach, 1991:21-22), we would expect the principal effect of any diphthongal movement to be on F_1: if there is a more distinct diphthong in British English than SEP, we would expect to find a smaller rate of change for F_1 for SEP.

DATA

Five Singaporean first-year university students were recorded, reading a story specially designed for the large number of /eɪ/ and /oʊ/ diphthongs, and then talking freely with the author of this paper. The conversation was initially about language usage at home, but in most cases it progressed to a discussion of language usage at university. Three Singaporean office secretaries, one Singaporean university lecturer, and three British university lecturers were also recorded, discussing a picture and then talking with the author about educational experiences. Finally, measurements were also made of BBC broadcasters from the MARSEC database (Roach *et al.*, 1993). All the BBC broadcasters might be described as speakers of RP, and the British university lecturers as speakers of RP or near-RP (Wells, 1982:297).

All the speakers were female. The study was limited to female speakers because there are more female students and office staff on the campus where the study was conducted. It is often suggested that female speech may be more difficult to analyse than male speech. For example, the higher fundamental frequency results in wider spacing of the harmonics, which can make it hard to identify formants clearly in spectrographic analysis (Kent & Read, 1992:157), and female speech tends to be more breathy than male speech, and this may interfere with LPC analysis (Kent & Read, 1992:158). However, it was found in measuring the first formant of speakers of both genders from the MARSEC database that there were occasional problems in measurement for both male and female speakers, and that identification of the location of F_1 was not necessarily harder for females: for both male and female speakers, there were instances when F_1 could not reliably be identified, and so some tokens of the vowels would have to be ignored.

All the Singaporean speakers were ethnically Chinese. Although a significant proportion of Singaporeans are not Chinese, this study only considered Chinese speakers, to limit the variables. Most of the subjects speak Mandarin and one more dialect of Chinese, though one student admitted to speaking Mandarin very poorly. All the students speak English either equally well or better than Mandarin. The three office staff all use English regularly and easily in their work at the university, though their Mandarin is probably better than their English.

The Singaporean university lecturer has a Ph.D. from an Australian university. Despite her linguistic proficiency, she retains a distinct Singaporean flavour to her pronunciation of English. The students all have 'A' levels, and the office staff all have 'O' levels but no 'A' levels. These speakers therefore represent a wide range of educational levels.

The office staff will be referred to with an 'S' prefix: S1, S2 and S3. The university students will be referred to with a 'C' prefix: C1 to C5. The Singaporean lecturer will be referred to with an 'L' prefix: L1. The British lecturers will have a 'B' prefix: B1, B2 and B3. And the BBC broadcasters will be identified from the directory in the MARSEC database where their speech can be found: ASIG, DSIG, and FSIG. In each of these cases from MARSEC, the speaker is the one found at the start of the first file in the directory.

MEASUREMENTS

All the recordings were made on to audio tape. Whenever possible, twenty instances of each vowel were identified for each speaker, though in some cases, particularly for the office staff, there were not always twenty instances to be found. The utterances containing these words were then transferred to a 486 PC and the vowels measured using CSL software from Kay.

The measurements were made from spectrograms with overlaid LPC-based formant tracks. In some cases, the order of the LPC had to be varied: a 12-order analysis was used, but in cases where this resulted in no formant track in the region of F_1, 16-order was attempted instead. In some cases, no reliable measurement could be found, so these tokens were ignored and, wherever possible, other tokens were found.

RESULTS

In Table 1, the average negative rate of change (ROC) for the /eɪ/ and /oʊ/ vowels for each of the speakers, the number of tokens measured (num), and the overall average (ave) for the office staff, the students, the Singaporean lecturer, the British lecturers, and the BBC broadcasters are shown. (In calculating the overall averages, the number of utterances for each speaker was taken into account.) In all cases, the average rate of change (measured in Hz per second) was negative, as is expected for closing diphthongs, so the minus sign has been omitted.

These results indicate that all the Singaporean speakers have a smaller average diphthongal movement for both their /eɪ/ and /oʊ/ diphthongs than all the British speakers. However, there still remains a consistent though relatively small diphthongal movement for the Singaporeans.

		/eɪ/			/oʊ/		
		ROC	num	ave	ROC	num	ave
office staff	S1	114	4	342	600	10	522
	S2	72	9		275	14	
	S3	539	17		713	14	
students	C1	373	20	534	19	18	263
	C2	443	20		317	19	
	C3	316	20		303	20	
	C4	936	20		523	18	
	C5	602	20		155	20	
Sing. lecturer	L1	582	16	582	487	15	487
Brit. lecturers	B1	996	18	1247	823	13	995
	B2	1124	14		1035	13	
	B3	1694	14		1118	14	
BBC	ASIG	1024	20	1606	1261	20	1653
	DSIG	1522	20		1681	20	
	FSIG	2273	20		2018	20	

Table 1. Average negative rate of change (ROC) for F_1 (in Hz/sec), number of tokens measured (num) for each of the fifteen speakers, and overall average rate of change for each class of speaker (ave).

DISCUSSION

For all the speakers, there was a great deal of variation in the rate of change measurements for the individual tokens. For example, for the /oʊ/ of speaker C3, the values range between –1493 Hz/sec and +222 Hz/sec, and this kind of variation was typical for all the speakers. It is not clear if this variation arises from the influence of neighbouring consonants, or if it arises from genuine differences in the production of the diphthongs by the speaker, or if it occurs because of limitations in the analysis and use of formant frequencies to describe the quality of vowels. The effect of neighbouring consonants would seem inadequate to explain all of the variation, because for this speaker, C3, the same word 'no', spoken in a phrase on its own, is found with rates of change varying between –1493 Hz/sec and +222 Hz/sec. Does this mean that sometimes this speaker actually produces /oʊ/ as an opening diphthong? Impressionistic listening suggests that this is not the case. Further work is needed to investigate exactly how diphthongs can best be measured acoustically.

Although the results show that diphthongal movement is consistently smaller for Singaporean than British speakers for both /eɪ/ and /oʊ/, all the Singaporeans still do have some movement with these vowels. It is not clear if this is due to the inherent movement found in all vowels, or whether /eɪ/ and /oʊ/ should still be classified as diphthongs in SEP, but with less movement than in British English

The results suggest that there is no clear difference in the pronunciation of these two diphthongs between Singaporeans of different educational levels. For /oʊ/, the students actually have less diphthongal movement than the less well-educated office staff; and for /eɪ/, although there is greater movement for the students than for the office staff and also for the lecturer compared to the students, the latter difference is too small to allow us to conclude that the diphthongal movement of this vowel depends on educational levels.

Although there undoubtedly are large differences between the English of well-educated and less well-educated Singaporeans, this apparently does not have much influence on the pronunciation of these two diphthongs. The use of a relatively monophthongal realization of these two diphthongs by all Singaporeans might be regarded as a distinctive characteristic of the local speech. And the fact that even well-educated Singaporeans do not adopt a British model for the pronunciation of these sounds suggests that, pedagogically, there is little point in trying to force students to pronounce these two sounds as diphthongs.

ACKNOWLEDGEMENTS

This work was partly supported by an NTU/NIE project entitled 'An Instrumental Study of the Sounds and Prosodies of Current Singapore English Pronunciation'.

REFERENCES

Brown, A. (1988). "Vowel differences between Received Pronunciation and the English of Malaysia and Singapore: Which ones really matter?", in Foley, J. (ed.). *New Englishes - The Case of Singapore*, 129-148. (Singapore University Press: Singapore).

Cruttenden, A. (1994). *Gimson's Pronunciation of English* (5th Edition). (Edward Arnold: London).

Deterding, D. (1990). *Speaker Normalisation for Automatic Speech Recognition.* Ph.D. Thesis, Cambridge University.

Deterding, D. & Hvitfeldt, R. (1994). "The features of Singapore English Pronunciation: Implications for teachers". *Teaching & Learning,* 1994 (1), 98-107.

Fry, D. B. (1979). *The Physics of Speech.* (Cambridge University Press: Cambridge).

Gay, T. (1968). "Effect of speaking rate on diphthong formant movement". *Journal of the Acoustical Society of America.* 44 (6), 1570-1573.

Kent, R. D. & Read, C. (1992). *The Acoustic Analysis of Speech.* (Singular Publishing Group: San Diego).

Platt, J. & Weber, H. (1980). *English in Singapore and Malaysia.* (Oxford University Press: Oxford).

Roach, P. (1991). *English Phonetics and Phonology* (2nd Edition). (Cambridge University Press: Cambridge).

Roach, P., Knowles, G., Varadi, T & Arnfield, S. (1993). "MARSEC: A Machine-readable Spoken English Corpus", *Journal of the International Phonetic Association,* 23 (2), 47-58.

Tay, M. W. J. (1982). "The phonology of Educated Singapore English", *English World-Wide*, 3 (2) 135-145.

Tongue, R. K. (1979). *The English of Singapore and Malaysia* (2nd Edition). (Eastern Universities Press: Singapore).

Wells, J. C. (1982). *Accents of English 2: The British Isles.* (Cambridge University Press: Cambridge).