VOWEL CHANGE IN RECEIVED PRONUNCIATION:

EVIDENCE FROM THE QUEEN'S ENGLISH

Jonathan Harrington, Sallyanne Palethorpe,
and Catherine Watson
Macquarie Centre for Cognitive Science and
Speech Hearing and Language Research Centre,
Macquarie University, Sydney, Australia.

ABSTRACT

It is well established that accent changes in general originate from younger members of the community whose speech includes more innovative forms of pronunciation. However there is a paucity of studies that have examined experimentally whether a person's vowel space changes with time in the same direction as that of the wider community. In order to examine this further, we analysed nine of the Christmas broadcasts made by Queen Elisabeth II spanning three time periods (the 1950s; the late 1960s/early 70s; the 1980s). An analysis of the monophthongal formant space showed that the first formant frequency was generally higher for open vowels, and lower for mid-high vowels in the 1960s and 1980s data than in the 1950s data, which we interpret as an expansion of phonetic height. The second formant frequency showed a more modest compression in later, compared with earlier years.

INTRODUCTION

It is well established from a variety of different studies that accent changes in general, and vowel changes in particular, originate from younger members of the community whose speech characteristics include more innovative forms of pronunciation (e.g., Eckert, 1988; Labov, 1994; Wells, 1982). The idea that older speakers' vowels are scarcely (Labov, 1994), if at all (Wells, 1982), affected by changes in a community forms the basis for 'apparent time' studies that document accent change through a comparison between the speech of young and older members from the same community, gender, and socioeconomic groups. However, partly because of the inherent difficulty of obtaining recordings from the same subjects over a long time period, there is a paucity of 'real time' studies that have examined experimentally whether a person's vowel space changes with time. There are two acoustic studies that are relevant. Firstly, an acoustic comparison of the vowels produced by the same speakers of Received Pronunciation (RP) in 1964 and 1983, Bauer (1985) found that the same speaker's vowels had shifted in the direction of the community vowel changes. Secondly, Yarger-Dror (1994) analysed acoustically thirteen speakers of Montreal French in 1971 and 1984 and found that the speakers ' continue to advance towards a newer phonology well into middle age'.

In order to examine this further, we decided to analyse acoustically the vowels from the Queen's Christmas message over a period of 35 years. The Queen's accent has some characteristics of what Gimson (1966) describes as 'conservative RP", a form 'used by the older generation and traditionally, by certain professional groups' and has only some, but not all, of the features of what Wells (1982) refers to as Upper Crust, or U-RP. These broadcasts are well suited to the aim of investigating vowel changes within the same person over a number of years. The broadcasts are stylistically very similar, if not identical in most respects. They have been recorded annually since 1952, and the recording quality of all the materials is very high. Finally, assuming that the Queen is a conservative speaker of RP, we can be reasonably confident that any vowel changes would provide robust evidence that age is not immune to community changes.

METHOD

We digitised nine Christmas broadcasts: an 'early' group, the 50s (1952, 1954, 1957), a 'mid' group, the 60s (1967, 1968, 1972), and a 'late' group, the 80s (1983, 1985, 1988). Each passage was transcribed orthographically and for all passages, we identified major prosodic boundaries and marked the acoustic vowel targets of all accented words from spectrograms in the EMU speech database analysis system (Harrington & Cassidy, 1999). Formant and fundamental frequency data were calculated automatically using the ESPS system and this data was checked and manually corrected if needed for all the pitch-accented vowels.

The accented vowels were pooled across each of the three time periods (50s, 60s and 80s) and the first two formant frequencies were analysed. Those tokens that were outside 2.45 standard deviations from the mean were considered to be outliers and were removed from the final analysis. Also removed were tokens of the vowels [ɪ and ɛ] that preceded a velarised realisation of /l/ (e.g. 'still', 'fell') because of the substantial anticipatory influence of this consonant on these vowel targets (Lehiste, 1964). [u] was also affected by the preceding [j] context (e.g. 'new,' 'few') but since [u] fronting was one of the main areas of investigation we relabelled these tokens as [ju] to differentiate them from [u] preceding by other contexts. Finally, 2337 monophthongal vowels remained, with a distribution and vowel type shown in Table 1.

Year Group	50s	60s	80s		50s	60s	80s
Lax Vowels				Tense Vowels			
ɪ [HID]	103	105	123	i [HEED]	82	85	82
ɛ [HEAD]	138	139	156	ɜ [HEARD]	37	28	33
æ [HAD]	96	94	77	ɑ [HARD]	35	26	42
ʌ [MUD]	80	80	83	ɔ [HOARD]	78	74	82
ɒ [NOD]	65	65	64	u [FOOD]	24	12	17
ʊ [HOOD]	8	6	21	ju [FEW]	31	32	37

Table 1. Distribution and number of vowel type across year

RESULTS

The mean values for the first two formants of the monophthongs in the three different year groups are shown in Table 2. The lax vowels showed considerable changes from the 1950s to the 1980s data in terms of a 'vertical' expansion of the vowel space brought about by the raising of F1 for [æ] and [ʌ] and the lowering of F1 for [ɪ] and [ʊ]. The overall effect of this expansion is that the [ɪ ɛ æ] spaces overlap extensively in the 50s data but much less so in the 60s and 80s data. Similarly, while [ʊ] overlapped considerably with [ɒ] and even with [ʌ] in the 50s data, the separation of these vowels was greater in the 80s data. Table 2 also shows that the height relationship between [ɛ] and [ɒ] increases from the 50s to 80s data due to a slight lowering of F1 for [ɛ] and a raising of F1 for [ɒ]. The relationship between the F1 of [æ] and [ʌ] also changes over the time period with the two vowel spaces being closer to each other in the 60s and 80s data than in the 50s data.

The trend for a 'vertical expansion' of the vowel space from the 50s to the 80s can also be seen in F1 for the tense vowels: F1 for [i ɔ u ju] is lower in the 60s/80s than in the 50s and F1 for [ɑ] is higher in the 60s/80s than in the 50s. There is also a raising of F2 in [ju] and [u] in later years and a slight lowering of F2 in [i], resulting in a decreased distance between the vowel spaces for [ju/u] and [i]. There was also a progressive lowering of F1 and F2 for [ɜ] and because of the opposite direction of the F2 changes in [ɜ] and [ju/u], their vowel spaces were more clearly separated in F2 in the 50s than in the 80s.

Lax Vowels	F1			F2		
	mean(50s)	mean(60s)	mean(80s)	mean(50s)	mean(60s)	mean(80s)
ɪ	518	434	453	2201	2171	2160
ɛ	653	642	602	2146	2107	2049
æ	747	928	877	2144	2013	1913
ʌ	800	923	895	1660	1590	1586
ɒ	668	703	689	1205	1223	1212
ʊ	582	514	483	1346	1334	1323

Tense Vowels	F1			F2		
	mean(50s)	mean(60s)	mean(80s) .	mean(50s)	mean(60s)	mean(80s)
i	397	397	369	2831	2759	2750
ɑ	734	854	817	1256	1285	1262
ɔ	532	482	475	853	851	839
ju	445	391	378	1274	1542	1511
u	434	404	394	1148	1238	1253
ɜ	652	642	612	1994	1964	1856

Table 2. Mean F1 and F2 values in Hertz of the monophthongal vowels over the three year groups: 1950s, 1960s and 1980s.

DISCUSSION

The acoustic analysis of these Christmas broadcasts shows that there has been an increase in the 'vertical' distance of the vowel space from the 1950s to the 1980s, together with a slight 'horizontal' compression. The vertical expansion is due to an F1 raising of the open vowels [ɑ ʌ æ ɒ], which suggests that they are phonetically more open in the 1980s data compared with the 1950s data, together with a lowering of all other vowels, particularly [i ʊ]. The horizontal compression of the vowel space is brought about by an F1 raising of [ju u] and F2 lowering of [i ɛ ɑ ɜ ʌ]. Since there are no significant changes in F2 of the back vowels [ɑ ɒ ɔ ʊ], they are less separated from the front vowels in the 1980s than in the 1950s. Gimson (1966) mentioned that "a compression of the front vowels is a characteristic of RP" although this comment is not present in Cruttenden's (1994) edition of Gimson's book; whether Cruttenden felt that the horizontal compression was not present in the RP of the 1990s is not stated.

Our conclusions are that many of the Queen's vowels have changed in phonetic quality over a 35 year period, as shown by changes in the F1 and F2 formant frequency values. Many of these changes have taken place mainly in the period from the mid-fifties to the early seventies with little change between the seventies and late-eighties; and further these changes reflect many of the community changes reported during this time period. We also agree with Bauer (1985) that this adjustment to an adult's vowel space has implications for studies that seek to document diachronic change in terms of apparent rather than real time, as apparent time studies may well underestimate the influence of community changes on those of an adult's personal vowel space.

Acknowledgements

We thank Buckingham Palace for granting permission to analyse the Queen's Christmas messages and the BBC for providing us with the recorded material from their archives.

REFERENCES

Bauer, L. (1985). Tracing phonetic change in the received pronunciation of British English. *Journal of Phonetics*, 13, 61-81..

Gimson, A.C. (1966). *An Introduction to the Pronunciation of English* (First Edition). Edward Arnold: London.

Gimson, A.C. and Cruttenden, A. (1994). *Gimson's Pronunciation of English* (Fifth Edition), revised by Alan Cruttenden). Edward Arnold: London.

Eckert, P. (1988) Adolescent social structure and the spread of linguistic change. *Language in Society*, 17, 183-207.

Harrington, J. and Cassidy, S. (1999). *Techniques in Speech Acoustics*. Kluwer: Dordrecht.

Labov, W. (1994). *Principles of Linguistic Change*. Blackwell: Cambridge Ma.

Lehiste, I. (1964) *Acoustical Characteristics of Selected English Consonants*. Indiana University Press: Bloomington.

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

Yaeger-Dror, M. (1994). Phonetic evidence for sound change in Quebec French. In P.A. Keating (ed.) *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*. (pp. 267-292). Cambridge University Press: Cambridge.

ACOUSTIC COMPARISON OF CHILD AND ADULT FRICATIVES

Akiko Onaka and Catherine I. Watson .
Speech Hearing and Language Research Centre
Macquarie University

ABSTRACT: The study presents acoustic comparisons between child and adult productions of the 9 English fricatives. Fricative tokens are obtained from citation-form words by 4 boys and 4 girls (aged 7 to 11) and 5 men and 5 women. The results show the overall spectral shapes of fricatives produced by the children are similar to those by the adults, however, some significant differences are found in the resonance values and resonant bandwidth. Classification experiments results show that in general the children's fricative data performed much more poorly than the adult data. The implications of the results for automatic speech recognition are discussed.

INTRODUCTION

This study presents acoustic comparisons of fricatives from children and adults and looks at the implications of these results for automatic speech recognition. Existing speech recognition systems have been developed mainly to recognise adult speech, particularly from adult male speakers and accordingly children's speech possesses some problems for current speech recognition systems (Russell *et al.*, 1996). However, few studies have investigated the performance of speech recognition technology with children. A study by Wilpon & Jacobsen (1996) investigated the performance of speech recognition systems on different age groups and found the recognition rates were lowest for children (aged 8-12). Since children's speech production development has been observed to continue until as late as 10 to 12 years of age (e.g. Tingley & Allen, 1975), there will be more variability in children's speech than in adults. Further, fricatives are one of the most difficult class of sounds for English-speaking children to acquire (e.g. Ingram *et al.*, 1980). Therefore, automatic recognition of children's fricatives may be problematic.

Previous studies have described various acoustic characteristics on fricative consonants in adult speech and some earlier descriptions of adult fricatives (e.g. Hughes & Halle, 1956; Tabain, 1998) have shown that distinctive characteristics of fricatives can be found in their spectral shapes. Children's fricatives have not been studied as extensively as adult fricatives in terms of their acoustic features. Pentz *et al.* (1979) found that the children's spectra of /f, v, s, z, S, Z/ had higher major resonances than the adults' spectra. Nittrouer (Nittrouer *et al.*, 1989; Nittrouer, 1995) found that the spectra of two fricatives /S/ and /s/ produced by children were significantly different from those of adult fricatives in both the position and shape of the resonant peak.

The first part of this study aims to examine the spectra of all the 9 English fricatives in children's speech and compare them to those from adults. The comparisons are quantified using spectral moments. In the second part we perform a series of small scale classification experiments to investigate whether there are differences in the fricative separation rates between the adult and child data. We discuss the implication of the results for speech recognition systems for children.

METHOD

Subjects

Children's fricatives analysed in this study were taken from a pre-existing database (Cassidy & Watson, 1998). The speech samples were collected from 4 boys and 4 girls aged between 7 and 11 years. All subjects were native speakers of Australian English (AE). The adult fricatives analysed were taken from the Otago speech database (Sinclair & Watson, 1995). The database included 11 men and 10 women aged between 16 and 33 years as subjects. Among these speakers, 5 male and 5 female speakers were selected for this study. All the subjects were native speakers of New Zealand English (NZE). The Otago database was used because it uses the same speech list materials as the Australian children's database. Ensuring maximum control of the (phonetic) context in which fricatives were produced prevents us from attributing differences between child and adult data to the effects of neighbouring segments on sounds in question.

Materials

The speech materials used in the two databases consisted of a set of citation-form productions of 129 (real) words from each subject. Each database contained usually three target words for each phoneme, which exemplified the phoneme in word-initial, word-medial, and word-final positions. For this study, 26 words containing an instance of the fricative sounds /f, v, T, D, s, z, S, Z, h/ were selected. Although their phonetic contexts were not controlled, fricatives which were either preceded or followed by [i] vowel were avoided due to a potential coarticulation effect on fricative sounds. Although AE and NZE vowel spaces are very similar, there are some noticeable differences, in particular the /i/ vowel is quite different being a high front vowel in AE and a retracted mid vowel for NZE (Watson *et al.*, 1998).

As described in Cassidy & Watson (1998), the child data was recorded in a sound-treated studio at the Speech Hearing and Language Research Centre, Macquarie University. The speech data was sampled at 20 kHz and quantised to a 16-bit number. The adult speech database was recorded in a quiet room. The speech was sampled at 22.05 kHz and quantised to a 16 bit number (Sinclair & Watson, 1995). For both of the data, the material was presented to the subjects in a random manner, one word at a time to avoid list intonation. One token for each target word was obtained for the child data and three tokens for the adult data.

The children's speech data was segmented and labelled phonetically by a trained phonetician (Cassidy & Watson, 1998) and the adult speech data was labelled phonemically by the first author. The labelling was carried out in EMU, a speech data management system (Cassidy & Harrington, 1996). The labelling criteria followed those described in Croot & Taylor (1995).

Spectra and spectral moments

Spectra for the adult and children's fricative tokens were obtained by performing a series of overlapping 256 points FFTs across the entire fricative token and averaging the result. Each successive FFT slice was overlapped 50 % over the previous FFT slice. The bandwidth of the spectra from the child data was 10 kHz, whereas for the adult data it was 11.025 kHz, due to the slight different sampling frequencies of the two data sets. In Figure 1 the adult spectra were truncated to 10.0734 kHz (i.e., which is the FFT bin closest to 10 kHz) for comparison purposes. The spectral moments were calculated based on the formula given in Forrest *et al.* (1988) and calculated for the spectra of fricatives in the frequency range of 1 to 9 kHz. The region beneath 1 kHz is only expected to yield information about voicing, since we analyse the voiced and voiceless fricatives separately, we felt there was no need to include this information. Secondly, the children's fricative spectra above 9 kHz all showed a rapid drop in amplitude (see Figure 1), suggesting instrumental distortion due to perhaps the anti-aliasing filter. For this reason the spectra above 9 kHz were not examined either.

Classification experiments

The Gaussian classification technique was used to classify a given token, similar to that used by Tabain (1998). A Bayesian distance metric was used to measure the probability of a given token belonging to a particular class of phoneme, and a "round-robin" procedure was to test and train the classifier. For this study, three classification experiments were carried out: firstly on the adult data, secondly on the child data, and finally on both of the adult and child data where the classifier was trained with the adult data and tested on the child data.

RESULTS

Fricative spectra and spectral moments

Figure 1 shows averaged fricative spectra for the 8 child and 10 adult speakers. The overall spectral shapes of the adult and children's fricatives are similar. The spectra for the adult fricatives are similar to other studies (e.g. Tabain, 1998). The spectra of /f, v, T, D/ are relatively flat with a diffuse spread of energy across the spectra, with /v, D/ having a noticeable voiced peak in the frequency region below 1 kHz. For /s, z/, most of the spectral energy is in the 4-9 kHz region for the children and 4-6 kHz region for the adults. The spectral energy for /S, Z/ is in the 2-6 kHz region for the children and 2-3.5 kHz