PITCH AND DURATION IN RP: A CORPUS-BASED HISTORICAL EXPLORATION

Joanna Przedlacka and Ladan Baghai-Ravary

Phonetics Laboratory, University of Oxford
{joanna.przedlacka, ladan.baghai-ravary}@phon.ox.ac.uk

ABSTRACT

As a reference variety, Received Pronunciation (RP), a non-regional accent of British English, is extensively described both in terms of synchronic variation and its historical development. There is a long tradition of describing changes over time in the traditional phonemic framework. However, modern corpus-based acoustic investigations have not been attempted on material older than the 1950s and most of such work focuses on vowel systems, with other phonetic features less extensively covered. In the present paper, we examine data from two closely matched groups of middle-aged male RP speakers, recorded in 1929 and 2010, demonstrating that corpus-based comparisons can be extended further back in time than previously done. Investigating pitch and durational characteristics of their speech, we also show historical investigations can go beyond the segment. Possible reasons for the discrepancies between the groups are explored.

Keywords: pitch range, segment duration, diachronic change, male speech, Received Pronunciation.

1. INTRODUCTION

Received Pronunciation (RP) is an accent of British English, which has no regional traits. Having served as a reference variety for textbook descriptions of British speech and used as a (previously exclusive) model in language teaching materials in Europe for almost a century, it is undoubtedly one of the best described accents of English. There also exists a long tradition of describing historical changes in this variety in the classic phonemic framework [3, 10, 13].

In more recent decades these classic studies of RP have been augmented by modern corpus-based acoustic investigations. The chief focus of those enquires has been the vowel system [5, 6, 14], with other phonetic features less extensively covered. In the present study we examine data from two closely matched groups of middle-aged male RP speakers, recorded in 1929 and 2010, comparing their pitch and phoneme durational characteristics. We carry out an acoustic analysis of the data, about which we had formed auditory impressions beforehand. On listening, it struck us that the early 20th century speakers sound higher-pitched than their modern counterparts and that their speech is also slower with vowels appearing to have longer durations. We hope that the present contribution can serve to demonstrate that historical investigations of speech can go beyond segmental features and corpus-based comparisons can be extended further back in time than had been previously done.

The research aims were: (1) to compare pitch and phoneme duration of two groups of male RP speakers, whose birth dates span almost a century, (2) to test whether reliable acoustic analysis can be carried out on data older than previously attempted. (3) to corroborate auditory impressions with acoustic data. The general prediction was that auditory judgements could be corroborated by acoustic analyses.

2. DATA

The present investigation makes use of two pre-existing RP speech corpora. The 1929 material is freely available from the British Library Sound Archive *Early spoken word recordings* [2]. The data was obtained by recording the version played via the web directly into *Speech Filing System* [8] via Stereo-Mix input, at a sampling rate of 10 kHz. The sample forms part of the Linguaphone English series *English Conversation*. Each lesson deals with an aspect of daily life such as visiting the doctor, being invited to dinner or going on holiday to the country. Each unit consists of a short passage describing the topic, followed by a conversation between two or sometimes three speakers. The voices on the recordings belong to eminent linguists and other prominent figures of the era. For the purposes of the present study we have selected five male speakers. Three of them are phoneticians connected with UCL Department of Phonetics: Daniel Jones, Arthur Lloyd-James and J.R. Firth, who later became professor of linguistics at SOAS. The remaining ones are Victor Clinton-Baddeley, an actor, radio

poetry reader and editor for the *Encyclopaedia Britannica* and Dennis Drew Arundell, a theatrical historian and producer. At the time of the recording all men were between 30-50 years of age, and university educated.

Our 2010 data is a closely-matched sample. This material had been gathered as part of another research project on voice characteristics in English. One of the informant selection criteria was the lack of regional features in their speech. The five male speakers we have chosen to analyse were in the same age range (30-50) as the 1930s group and British born. All had university degrees and hailed from families with high level of educational achievement. The informants were recorded reading *The Grandfather* passage widely used in speech and language therapy to collect data from patients. To keep the style variable constant, we compare the contemporary speech sample only to the read passages data in the 1929 Linguaphone lessons, excluding the conversations. Altogether our data constitutes 20 minutes of running speech in total from two groups of males, born in the first and the third quarter of the 20th century.

3. PROCEDURES

Source files were processed with commercial noise removal software [11], which performs noise removal by adaptive linear filtering, preventing artefacts due to the non-linear processing employed by most such systems. This avoids bias in results due to different levels of background noise in different recordings. All recordings were processed in this way, even if they had negligible levels of background noise, as a precaution against bias in results due to the processing itself.

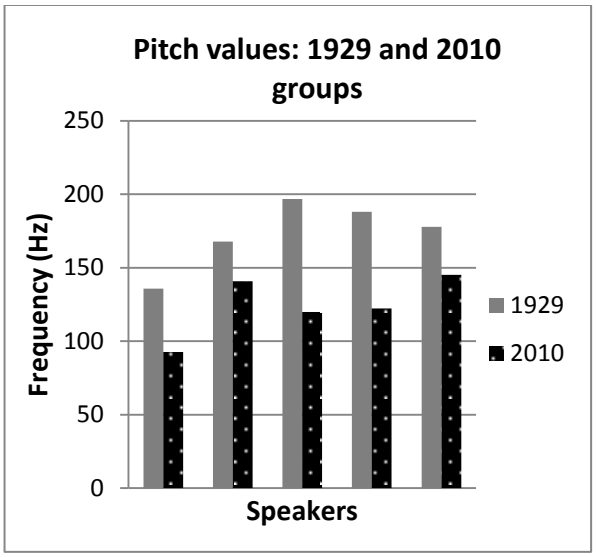
The signals were then normalised in amplitude and resampled to 16 kHz sample rate, mono, 16-bit PCM format using *SoX* [12]. Voice parameters were calculated using *Praat* [1], with the default methods and parameters ('To Pitch (ac)...', 'To PointProcess (cc)' and 'Voice report...'). This provided measures of mean and standard deviation of pitch as well as other voice parameters.

Prior to phonetic alignment, signals were processed with the standard *HTK* [7] Perceptual Linear Prediction Cepstral Coefficient front-end, with 25 ms Hamming windows, and a frame rate of 1 every 10 ms, a first-order pre-emphasis filter with a coefficient of 0.97, 20 frequency channels, and 12th order linear prediction analysis. The analysis included velocity (delta) and acceleration (delta-delta) coefficients, including cepstral coefficient zero. Phonetic alignment was performed using HTK based on a Standard Southern British English pronunciation dictionary and a set of acoustic models, developed in the University of Oxford Phonetics Laboratory for previous projects. The dictionary includes alternative pronunciations for many words, and the acoustic models correspond to a large subset of ARPABET [4] monophones.

4. RESULTS

4.1 Pitch range

Figure 1: Comparison of pitch values (Hz) between the 1929 and 2010 groups.



It is evident from Figure 1 that there is a clear mean pitch difference between 1929 and 2010 groups. In the former, the pitch range is between 136 and 197 Hz, with mean at 173 Hz, while in the latter group the range is between 93 and 145 Hz, with mean at 124 Hz. The difference between means was assessed by a non-parametric test of difference (Mann-Whitney U test) and was found to be significant (U = 54.00; p (exact 2-sided) = 0.000). Standard deviation pitch values were 39.9 and 22.4 for the 1930s and 2010 data respectively.

4.2 Phoneme durations

Figure 2: Comparison of mean phoneme durations (s) 1929 and 2010 groups.

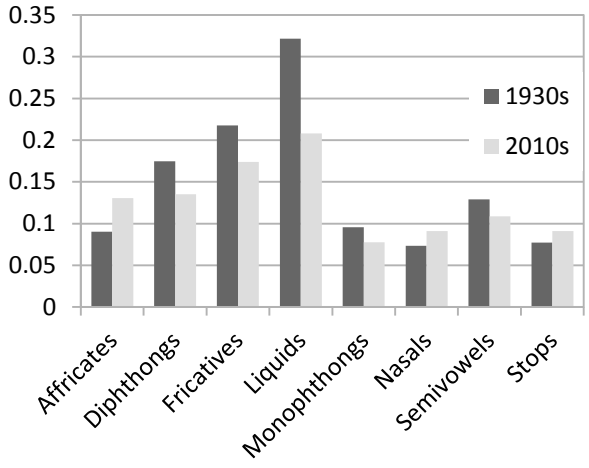


Figure 2 compares mean phoneme durations between 1929 and 2010 groups. Differences in means by broad phonetic class are presented. The mean durations differ both between classes of sounds and within each class there is also a difference between the two groups of recordings.

5. DISCUSSION

The results show a clear difference between the groups with respect to pitch range. The speakers recorded in 1929 have a higher mean pitch than the contemporary group. The mean durations differ between categories of sounds and between the two groups of recordings, with five out of eight classes having longer mean durations in the 1929 recording.

5.1 Pitch range

Speculatively, the pitch differences between the groups can be accounted for in terms of physical characteristics of the speakers. Admittedly, we have no data regarding those particular speakers' stature. However, it has been shown that over the course of the 20th century the average height has markedly increased in the population as a whole, which may be attributable to changes in diet and overall higher living standards. Taller speakers tend to have bigger vocal tracts and longer vocal folds, and those physical traits contribute to a lower pitch range [9]. This explanation is two-faceted. Firstly, we can consider that this particular group of speakers may have been of average shorter height than was typical of the era. Secondly, if the male population as a whole was of smaller stature, the norms for male speakers may have been those of smaller speakers. So even if particular individuals, in this case the speakers in our sample, were fairly tall, they may have accommodated to the mean determined by the general population.

Another plausible explanation for the higher mean pitch of the 1929 group is related to the recording ambience. In modern times in comfortable studio conditions with high quality microphones, speaking at a natural volume suffices to obtain good quality material. However, prior to electrical recording with microphones (beginning in 1926) recordings were acoustic where voice energy was transferred down a horn to cut the record. This required a different technique. To achieve a result comparable to its modern equivalent in terms of clarity one needed to speak at a volume higher than habitually used. Linguaphone recordings, made in 1929, were not acoustic, but very early electrical recordings. However, it is plausible to assume that our speakers might have had experience of making acoustic recordings and had not yet adapted to the microphone. Consequently, the effort involved caused their pitch to increase.

Finally, we recognise that an increase in pitch may also be a result of nervousness, which in this case would have been caused by a relative unfamiliarity with recording equipment. However, in connection with our data this is the least plausible explanation. While indeed information about the speakers' emotional state cannot be obtained, it is not very likely that a group of educated individuals with a professional interest in speech would have felt uneasy contributing to one of a number of recordings they had made in the course of their careers.

5.2 Phoneme durations

It has been noted that phoneme durations may vary between speakers and depend on the speaking rate [15]. Impressionistically, the 1929 group appear to speak slower, which might be attributable to the fact that the recording was made for teaching purposes. Hence, as the target audience was non-native speakers of English, one aimed to achieve maximum clarity and intelligibility. Speculatively, the reason that 1929 recordings could appear slower in overall speech rate is that they might have longer pauses between intonational phrases, which gives the impression of a slower delivery. This aspect was however not tested in the present study.

5. CONCLUSIONS

Our results demonstrate that it is indeed possible to extend corpus-based comparisons further back in time than previously attempted, even with commercially available recordings. F0 is easily recoverable, which allows for pitch comparison. Moreover, there exists a potential for investigating durational characteristics of speech. Durations of most sounds can be estimated automatically even in a relatively noisy signal. Moreover, this measurements are equally possible both in recordings made are over 80 years ago and in modern ones. Our recordings come from 1929, but potentially the same technique could be applied even to earlier material, thus opening opportunities for new investigations. It is hoped that the present contribution will go some way towards bridging the gap between acoustic corpus-based research and studies of historical development of language features.

We would like to thank Elinor Payne for generously allowing us to use her data from a research project funded by the John Fell Fund, Oxford University.

6. REFERENCES

[1] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, pp. 341-345

[2] British Library Sound Archive: http://sounds.bl.uk/Oral-history/Early-spoken-word-recordings

[3] Cruttenden, A. 2014. *Gimson's Pronunciation of English*, 8th ed., London: Routledge.

[4] Fisher, W. M.; Doddington, G. R. and. Goudie-Marshall, K. M., 1986. The DARPA speech recognition research database: specifications and status. In *Proc. DARPA Workshop on Speech Recognition*, 93–99, Feb. 1986.

[5] Harrington, J., Palethorpe, S. & Watson, C. 2000. Monophthongal Vowel Changes in Received Pronunciation: An Acoustic Analysis of the Queen's Christmas Broadcasts, *JIPA*, 30:1-2, 63–78.

[6] Hawkins, S. & Midgley, J. 2005. Formant Frequencies of RP Monophthongs in Four Age Groups of Speakers, *JIPA* 35:2, 183-199.

[7] HTK Speech Recognition Toolkit. Available for download from http://htk.eng.cam.ac.uk/

[8] Huckvale, M. 2010. Speech Filing System [computer software] available from *http://www.phon.ucl.ac.uk/resource/sfs/*.

[9] Johnson, K. 2006. *Acoustic and Auditory Phonetics*. 2nd ed. Oxford: Blackwell.

[10] Jones, D. 1932. *An Outline of English Phonetics*. 3rd ed., Cambridge: Heffer and Sons Ltd.

[11] Mono-a-Mono - The Wonder Audio-Scrubber, commercial software available from http://mono-a-mono.x10.mx/

[12] SoX - Sound eXchange, the Swiss Army knife of audio manipulation, open-source software available from http://sox.sourceforge.net/sox.html

[13] Wells, J. C. 1982. *Accents of English*, Cambridge: Cambridge University Press.

[14] Wikström, J. 2013. An acoustic study of the RP English LOT and THOUGHT vowels. *JIPA* 43, 37-47.

[15] Ziółko, B. and Ziółko, M. 2011. Time Durations of Phonemes in the Polish Language for Speech and Speaker Recognition. In Vetulani, Z. (ed.) *Human Language Technology. Challenges for Computer Science and Linguistics. Lecture Notes in Computer Science* 6562, 105-114.