# THE SPEAKER DISCRIMINATING POWER OF SOUNDS UNDERGOING HISTORICAL CHANGE: A FORMANT-BASED STUDY

*Gea de Jong, Kirsty McDougall, Toby Hudson, Francis Nolan*

Department of Linguistics, University of Cambridge
gd288|kem37|toh22|fjn1@cam.ac.uk

## ABSTRACT

This study investigates whether patterns of diachronic sound change within a language variety can predict phonetic variability useful for distinguishing speakers. An analysis of Standard Southern British English (SSBE) monophthongs is undertaken to test whether individuals differ more widely in their realisation of sounds undergoing change than in their realisation of more stable sounds. The vowels /æ, ʊ, uː/, demonstrated by previous research to be changing in SSBE, are compared with the relatively stable /iː, ɑː, ɔː/. Read speech of 50 male speakers of SSBE aged 18-25 from the DyViS database is analysed and compared with earlier results for 20 speakers. First, the data confirm the stability of /iː, ɑː, ɔː/, the fact that /ʊ, uː/ have indeed fronted and that the articulation of /æ/ has become more open. Results from discriminant analysis based on F1 and F2 frequencies show speaker classification rates well above chance. The non-stable vowels all achieved higher levels of discrimination than the stable /ɔː/. However, the highly variable pronunciation of some changing vowels in the case of a few individuals and the 'special' status of F1 for /ɑː/ and F2 for /iː/, increasing the rate for those vowels, made the overall picture more complicated.

**Keywords:** Speaker identification, sound change, vowels, formant frequencies, SSBE.

## 1. INTRODUCTION

The system of sound contrasts in a language is constantly in flux. Linguistic variation leads to change as new realisations of existing contrasts become established, as old contrasts are subject to merger, and as new contrasts are formed. At any point in time, certain sounds are changing while others appear more stable. The present study examines such variation as a potential source of speaker-distinguishing information. We hypothesise that, within a given homogeneous speech community, those sounds which are undergoing diachronic change are more likely to exhibit individual variation than sounds which are relatively stable. It is likely that certain speakers within the group will differ in terms of their realisations of variables which are undergoing change. Certain speakers may exhibit more conservative or more novel realisations than others. Although in the longer term a particular change would be expected to characterise all members of a speech community, in the shorter term patterns of usage may be valuable in distinguishing different speakers [9].

A recent and comprehensive acoustic study of diachronic change in SSBE monophthongs is provided by Hawkins and Midgley [6] (henceforth referred to as H&M). These authors analysed the F1 and F2 frequencies of monophthongs produced in /hVd/ contexts by male speakers of RP in four age groups: 20-25 years, 35-40 years, 50-55 years and 65-73 years. There were five speakers in each age group and directional patterns of differences in formant frequencies across successive age groups were interpreted as evidence of a time-related shift in the acoustic target for the relevant vowel.

The monophthongs H&M identified as having undergone the largest changes were /ɛ, æ, uː, ʊ/. For /ɛ/ and even more so for /æ/ the frequency of F1 was progressively higher for younger cohorts of speakers. These two vowels also exhibited a slight lowering in their F2 frequencies for successively younger age groups. Phonetic lowering of /æ/, the vowel found in HAD, is consistent with other studies [5: 83, 15: 291-2, 7: 44]. The frequency of F2 for /uː/, as in WHO'D, was progressively higher for younger speakers in H&M's study, consistent with the percept of /uː/, becoming increasingly centralised, or even fronted and less rounded [cf. 5, 7, 15]. Finally, H&M's data for /ʊ/, as in HOOD, showed a higher F1 and a much higher F2 frequency for the youngest speakers.

This study examines individual variation in productions of three changing vowels and three stable vowels in SSBE, by a group of speakers of the same sex and similar age. Formant frequency measurements of /æ, ʊ, uː/, (changing) and /iː, ɑː, ɔː/ (stable) are compared, to investigate whether patterns of sound change may inform the selection of indices useful for speaker identification.

## 2. METHOD

### 2.1 Database and Subjects

The DyViS database is a large-scale database of speech collected under simulated forensic conditions. It includes recordings of 100 male speakers of SSBE aged 18-25 (years of birth: 1981-1988) to exemplify a population of speakers of the same sex, age and accent. Each speaker is recorded under both studio and telephone conditions, and in a number of speaking styles. Further details about the content of the database and elicitation techniques are given in Nolan *et al.* [12]. In the present study, read sentences are analysed for 50 speakers, who were recorded between February and April 2006. The subjects had no history of speech or hearing problems, and their status as speakers of SSBE was judged by a phonetician who is a native speaker of that variety.

The data analysed are six repetitions per speaker of the vowels /iː, æ, ɑː, ɔː, ʊ, uː/ in hVd contexts with nuclear stress. Each hVd word was included in capitals in a sentence, preceded by schwa and followed by *today*:

It's a warning we'd better HEED today.
It's only one loaf, but it's all Peter HAD today.
We worked rather HARD today.
We built up quite a HOARD today.
He insisted on wearing a HOOD today.
He hates contracting words, but he said a WHO'D today.

Six instances of these sentences were arranged randomly among a number of other sentences and presented one at a time using *PowerPoint*. Subjects were asked to read each sentence aloud at a normal speed, in a normal, relaxed speaking style, emphasising the word in capitals. They practised reading a few sentences at the start. During the recording they were asked to re-read any sentences containing errors. Subjects were recorded in a sound-treated studio. Each subject was seated with a Sennheiser ME64-K6 cardioid condenser microphone positioned approximately 20 cm from his mouth. The recordings were made with a Marantz PMD670 portable solid state recorder using a sampling rate of 44.1 kHz.

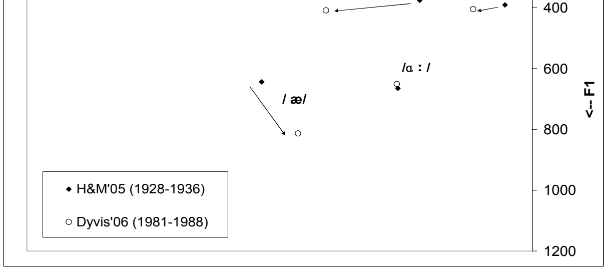
### 2.2 Measurements

Analysis was carried out using *Praat* [1]. Wide-band spectrograms were produced for each utterance. LPC-derived formant tracks were generated by *Praat*, and formant frequency values written to a log file for the time-slice judged by eye to be the centre of the steady state of each hVd vowel. In cases where no steady state for the vowel was apparent, the time-slice chosen was that considered to be the point at which the target for the vowel was achieved, according to movement of the F2 trajectory (i.e. a maximum or minimum or in the F2 frequency). All measurements were compared with visual estimates based on the spectrogram, values from adjacent time-slices, and the peak values of the frequency-amplitude spectrum at the target time-slice. When values generated by *Praat* were judged to be incorrect, they were replaced by correct values from a time-slice immediately preceding or following the slice being measured.

## 3. RESULTS AND DISCUSSION

Figure 1 compares the mean F1 and F2 frequency values of /iː, æ, ɑː, ɔː, ʊ, uː/ from the 65+ cohort of H&M [6] with the means calculated across the 50 speakers from the DyViS project. The figure mainly confirms the patterns of change noted by other authors and found in a similar analysis based on a subset of the speakers [3].

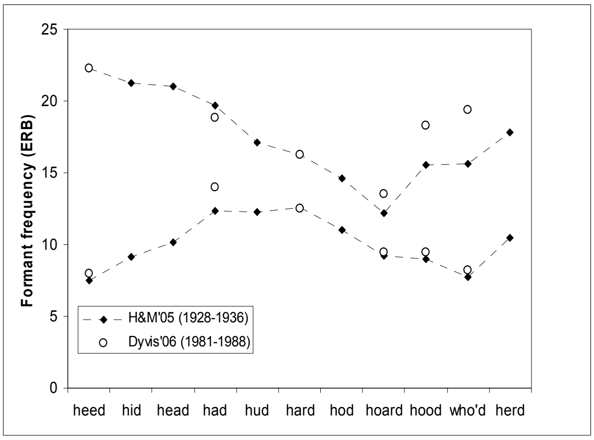
**Figure 1:** Mean F1 and F2 frequency values for the 65+ cohort of Hawkins & Midgley (2005) (diamonds) and for DyViS (circles).



For both /uː/, the vowel in WHO’D, and /ʊ/, the vowel in HOOD, the frequency of F2 has increased considerably, indicative of a more fronted pronunciation of those vowels, while the F1 has remained relatively unaffected. The change is most marked for /uː/, going from H&M’s mean of 994 Hz to a DyViS mean of 1612 Hz. This increase is similar to formant values reported by Wells [14] and slightly smaller than that noted by Deterding [4]. F2 of /ʊ/ increased from 990 Hz (H&M) to 1398 Hz. An increase, from 644 Hz to 813 Hz, in the frequency of F1 is observed for the vowel /æ/, giving it a more open articulation. This increase is slightly smaller compared to the formant values reported by Wells [14] and Deterding [4].

Despite the fact that in previous research /ɔː/ has been reported as a stable vowel, the F2 increased slightly from H&M’s 619 Hz to 758 Hz for the DyViS data, a pattern that would indicate slight fronting. Compared to Wells [14] however, this vowel has remained stable.

**Figure 2:** Mean F1 and F2 frequency values for the 65+ cohort of Hawkins & Midgley (2005) (diamonds) and for DyViS (circles) in ERB.



In Figure 2, F1 and F2 are shown separately per vowel with the formant frequencies in Hz converted to an auditory scale, the Equivalent Rectangular Bandwidth (ERB-rate) using the formula from Moore [8]:

ERB-rate=21.4*log(10)(0.00437*f+1)

where f is frequency in Hz.

Consistent with an earlier DyViS analysis of 20 speakers [3], the current data based on 50 speakers confirm that the pronunciations of /iː, ɑː, ɔː/ have indeed remained quite stable, whereas /ʊ, uː/ have fronted and the articulation of /æ/ has become more open.

The mean values of the frequencies of F1 and F2 of /iː, æ, ɑː, ɔː, ʊ, uː/ for each individual speaker are shown in Figure 3. Each data point represents the average realisation of the relevant vowel for a given speaker across 6 tokens.

**Figure 3:** Mean F1 and F2 frequency values for 50 SSBE speakers for the vowels in HEED, HAD, HARD, HOARD, HOOD and WHO’D. Each datapoint is a mean across 6 tokens of the vowel.

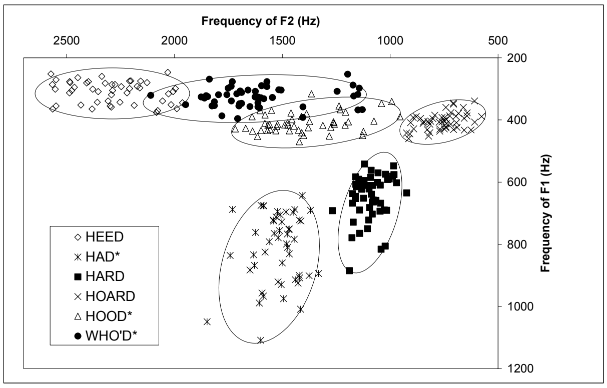


Figure 3 shows that these vowels differ considerably from one another in the degree of between-speaker variation they exhibit. For example, /ɔː/, the vowel in HOARD, is tightly clustered in the vowel space, with an F1 range of 121 Hz and an F2 range of 340 Hz. The /ɑː/ vowel in HARD, on the other hand, has a similar spread for F2, 345 Hz, but a much larger one, 343 Hz, for the F1 dimension. The largest F1 spread was found for /æ/ with a range of 466 Hz. The largest F2 spread is found for /uː/ in WHO’D, ranging from 1129 Hz for the lowest mean to 2110 Hz for the highest.

A result not predicted by sound change data for SSBE is that of considerable differences among speakers in their average F2 frequency of /iː/. Research has shown, however, that the frequency of F2 may be less crucial perceptually than a weighted average of F2, F3 and F4, which is because these formants may merge auditorily into one spectral prominence and as a result compensate for each other in the achievement of a specific phonetic quality [2, 11: 337-341]. The vowel /ɑː/ is also more variable in the F1 dimension than might be expected. This may be explained by the fact that the F1 of /ɑː/ (and /æ/) is highly sensitive to pharynx length [10: 171-172, 13: 270-271].

So, are formants of vowels undergoing change more useful indicators of speaker identity than stable vowels? The degree of speaker-specificity exhibited by each vowel was tested using discriminant analysis. This analysis is a multivariate technique which can be used to determine whether a set of predictors (here the F1 and F2 frequencies) can be combined to predict group membership (here membership of speakers S1-S50). Discriminant functions are determined to maximise differences between speakers relative to differences within speakers. Each token in the data set is then allocated to one of the speakers and the percentage of correct allocations (the 'classification rate') calculated. Here, this is done using the 'leave-one-out' method where each token is classified by discriminant functions derived from all tokens except for the token itself. The classification rates resulting for each vowel are given in Table 1.

The discriminant analyses allocated the tokens to the correct speaker 16-24% of the time, rates much higher than chance (1/50 = 2%) and quite promising considering the fact that only F1 and F2 were used. Although some vowel qualities perform better than others, the differences are minimal ranging over 8% only. All changing vowels (HAD, HOOD, WHO'D) do better than the stable vowel in HOARD, but only slightly. This may be due to a few speakers who exhibit high variation for these changing vowels as if caught between two competing targets. The high classification rates for the other stable vowels may be caused by special factors like formant merging (e.g. HEED) and vocal tract size (e.g. HARD).

**Table 1:** Speaker classification rates resulting from the discriminant analysis for each vowel. * indicates vowels changing in SSBE, according to previous research.

| | Classification rate |
|---|---|
| HEED | 23.7% |
| HAD* | 24.0% |
| HARD | 22.3% |
| HOARD | 16.0% |
| HOOD* | 19.7% |
| WHO'D* | 19.0% |

## 4. CONCLUSIONS

This study has provided a formant analysis of read data from 50 SSBE speakers from the DyViS database to assess whether sounds which are undergoing change are those most likely to differ between speakers. Discriminant analyses based on F1 and F2 showed speaker classification rates well above chance. All changing vowels clearly exhibited better speaker discrimination than the stable /ɔː/ vowel. However, the highly variable pronunciation of changing vowels in the case of a few individuals and the 'special' status of F1 for /ɑː/ and F2 for /iː/ made the overall picture more complicated.

## 5. ACKNOWLEDGEMENTS

This research is supported by the UK Economic and Social Research Council as part of the project 'Dynamic Variability in Speech [DyViS]: A Forensic Phonetic Study of British English' [RES-000-23-1248].

## 6. REFERENCES

[1] Boersma, P., Weenink, D. 2005. *Praat: Doing Phonetics By Computer* [computer program]. Available at: <http://www.praat.org/>.

[2] Carlson, R., Fant, G., Granström, B. 1975. Two-formant models, pitch and vowel perception. In: Fant, G., Tatham, M.A.A. (eds.), *Auditory Analysis and Perception of Speech*. London: Academic Press, 55-82.

[3] deJong, G., McDougall, K., Nolan, F. 2007 Sound change and speaker identity: an acoustic study. In: Christian Müller and Susanne Schötz (eds.), *Speaker Classification*. Springer.

[4] Deterding, D. 1990. *Speaker Normalisation for Automatic Speech Recognition*. Ph.D. Dissertation, University of Cambridge.

[5] Gimson, A.C., revised Cruttenden, A. 2001. *Gimson's Pronunciation of English (6th edition)*. London: Arnold.

[6] Hawkins, S., Midgley, J. 2005. Formant frequencies of RP monophthongs in four age-groups of speakers. *Journal of the International Phonetic Association* 35(2): 183-199.

[7] Hughes, A., Trudgill, P. 1996. *English Accents and Dialects*. New York: Oxford University Press.

[8] Moore, B.C.J. 1997. Aspects of auditory processing related to speech production. In: Hardcastle, W.J., Laver, J. (eds.), *The Handbook of Phonetic Sciences*. Oxford: Blackwell, 539-565.

[9] Moosmüller, S. 1997. Phonological variation in speaker identification. *Forensic Linguistics* 4(1): 29-47.

[10] Nolan, F. 1983. *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.

[11] Nolan, F. 1994. Auditory and acoustic analysis in speaker recognition. In: Gibbons, J. (ed.) *Language and the Law*. London: Longman, 326-345.

[12] Nolan, F., McDougall, K., de Jong, G., Hudson,T. 2006. A Forensic Phonetic Study of 'Dynamic' Sources of Variability in Speech: The DyViS Project. In: Warren, P. and Watson, C.I. (eds.), *Proceedings of the 11th Australasian International Conference on Speech Science and Technology*, 6-8 December 2006, Auckland: Australasian Speech Science and Technology Association, 13-18.

[13] Stevens, K.N. 1998. *Acoustic Phonetics*. Cambridge, Massachusetts: MIT Press.

[14] Wells, J.C. 1962. *A Study of the Formants of the Pure Vowels of British English*. MA Dissertation, University College London.

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