DEEPENING OR LESSENING THE DIVIDE BETWEEN DIPHTHONGS?

AN ANALYSIS OF THE QUEEN'S ANNUAL CHRISTMAS BROADCASTS

Jonathan Harrington
Institute of Phonetics and Digital Speech Processing (IPDS), University of Kiel, Germany.

Sallyanne Palethorpe
Macquarie Centre for Cognitive Science (MACCS), Macquarie University, Sydney, Australia.

Catherine Watson
Macquarie Centre for Cognitive Science (MACCS), Macquarie University, Sydney, Australia.

1. Introduction

This study is about the changes that have taken place in the last fifty years to one variety of English, Received Pronunciation (RP), the so-called standard accent of Britain. It is also about defining some of the pronunciation characteristics of Her Majesty Queen Elizabeth II. Close to the time of writing, Professor John Laver CBE has accepted The Queen's Anniversary Prize, on behalf of Queen Margaret University College Edinburgh for its pioneering contributions to speech science. Just over two years before this, in a thoughtful and well-reasoned article on a paper we had just published entitled 'Does the Queen still speak the Queen's English', Philip Hensher writing in *The Independent*[1] noted that our study could hardly seem anything but disrespectful to the point of *lèse-majesté*, while *The Times*[2], labelling us as 'three Australian phoneticists' (*sic*) in its Editorial of December 21st 2000 entitled 'Our Common Queen: Rougher Talk from her Majesty' thundered: 'We are not amused — oh no we ain't.' We feel, then, that we have a duty to explain our motives for going down this potentially treacherous research path.

Part of the answer is to be found in the multiple strands of speaker characteristics and variability that have been so carefully documented and modelled by John Laver in his many writings on these subjects over the last thirty years. As discussed in both *Principles of Phonetics* (Laver, 1994), and the *Phonetic Description of Voice Quality* (Laver, 1980), those aspects of vocal performance that define the characteristics of a speaker are due to considerably more than the anatomical and physiological properties of the vocal organs (although these of course play a considerable role as well). Within the same speaker there can be variations along a number of dimensions. The formality of speaking style can vary resulting in a greater or lesser tendency to reduce and assimilate speech. There can be variation in the paralinguistic tone of voice in which the loudness, pitch and segment duration can change according to a speaker's mood and emotional state. Articulatory settings can vary between speakers: there may be short-term differences, as in the tendency for some speakers of British English Received Pronunciation (RP) to produce /s/ with lip-rounding (Laver, 1994); or there may be long-term effects if, for example, a speaker habitually produces speech with hyponasality. Beyond these issues, there are the well-known differences in speakers due to dialect and social status, a factor that is especially relevant to the standard accent of England, Received Pronunciation. But the general point, as summarised by Laver (1994, p.66) is that:

> It will be apparent on reflection that the phonetic realizations of every phoneme of every speaker have the potential of being slightly different from those of many other speakers even of the same general accent. A lifetime of settling to a habitual mode of speaking has given the personal accent of every speaker an individualizing realizational flavour, within the overall systemic, structural and selectional conventions of his or her own accent-group.

The variability in the speech signal due to speaker differences poses a considerable problem if we wish to provide an experimental basis for how sounds have changed in time. Typically, in what has come to be known in sociolinguistics as an *apparent time* study (Chambers & Trudgill, 1980; Labov, 1994), pronunciation change is inferred by comparing the speaking characteristics of two different age-groups of the same community. But any such study is confronted with the difficulty that the many dimensions on which speakers can differ may mask the changes in pronunciation that we seek to validate experimentally, especially if - as is so often the case in sociophonetic analyses – we wish to base our comparison on continuous, or even completely spontaneously produced speech. A further potential difficulty may lie in the assumption that underlies apparent time studies that the spoken characteristics of, say, present-day 60 year old speakers is equivalent to those of 20 year old speakers from the early 1960s (Chambers & Trudgill,1980). As Labov (1994) comments: 'apparent time studies may understate the actual rate of sound change, since older speakers show a limited tendency towards communal change, participating to a small extent in the changes taking place around them'. A similar view is held by Laver & Trudgill (1979) who, in discussing the four major ways in which accents between two speakers can differ state that: 'These comments about accent differences between speakers make the assumption that a speaker's accent is fixed and unchanging. It seldom is, of course, and a further area where linguistic concepts can help us to refine our analysis of social markers in speech concerns certain aspects of the notion of *linguistic variability*. (their emphasis)'. Laver & Trudgill's (1979) view about the changing accent within the same speaker as well as Labov's (1994) caveat are consistent with a few so-called 'real-time' analyses which have found pronunciation changes in the same adult speakers recorded at intervals of between 15 and 20 years in Montreal French (Yaegor-Dror, 1994) and RP (Bauer, 1985).

On Christmas day in 1952, Queen Elizabeth II first adopted a tradition that was started by her grandfather King George V in 1932 and then continued by her father King George VI who had 'broadcast a message to his people in all parts of the world' (Christmas Broadcast, 1952). Since that time, the Queen has made broadcasts annually on Christmas Day and since 1957 these have been televised. From the point of view of studying sound change, the Christmas Broadcasts are unique: as far as we know, there are no other spoken materials recorded annually for over a fifty year period from the same person reading a text with a broadly similar communicative intent (a message to people in all parts of the world) in any variety of English. The advantages of analysing the Christmas Broadcasts for studying sound change are not only that we eliminate many of the confounding influences due to variation in spoken language between speakers, but also those due to variation within a speaker, since neither the speaking style nor the communicative purpose of the Christmas broadcasts vary a great deal from year to year. For these reasons, experimental evidence that points to sound change in the Christmas Broadcasts can, in all likelihood, be attributed to a phonetic change, rather than to an artifact of variation due to speaker-characteristics or changes to speaking style. To make use of a metaphor from Laver (1980), in which the relationship between voice quality and the linguistic-phonetic content of vocal performance is likened to that between a figure and the ground against which it is set, a very detailed study of (in our case almost 3000 vowels and diphthongs) of one speaker over a fifty year period clarifies the figure (the phonetic content) precisely because ground (the speaker-specific aspects) becomes so sharply delineated.

There is another reason why the Christmas Broadcasts are of great interest to phonetics and its relationship to sociolinguistics: they can tell us something about the extent to which adults can resist pronunciation changes that are taking place in the community. We can be in no doubt that RP has changed in the last fifty years (Crystal, 1988; Laver, 1994) and it is also likely that at least some of these changes are linked to the considerable change that took place in the social structures and hierarchies in the second part of the 20th Century. In the 1950s, the demarcation between the social classes was sharp but throughout the 1960s there was a progressive blurring between the class boundaries (Cannadine, 1998). Many professions, which in the 1950s would have been exclusively the preserve of the Establishment and those educated at public schools have been increasingly pursued by those from a variety of different class backgrounds. The weakening of the boundaries that define the class system has a corollary in the change in accent in England and the attitude towards Received Pronunciation, a label first used by Alexander J Ellis (1869, p. 23) to refer to the accent of the Establishment: that is, an educated pronunciation of the metropolis, of the court, the bar, the pulpit and one that did not differ across regions in England (see also Parsons, 1998 for a comprehensive review of RP). In the 1950s, Received Pronunciation was still very much a prestige accent associated with the middle and upper classes, who had typically been educated in Britain's public schools. It was perhaps unthinkable to most living in the 1950s that RP and the stigmatised Cockney English, associated with the working classes, had very much in common at all (beyond the fact that both were accents of the same language). But in the 1980s and 1990s, the situation is very different. The form of RP spoken by the upper classes is declining according to Gimson & Cruttenden (1994) while a greater proportion of the educated middle classes now apparently speak 'Estuary English', a term first used in 1984 by David Rosewarne. Estuary English is an accent of the South East of England (spoken around the Thames Estuary, and hence the name) whose foundation is Received Pronunciation but which has considerable influences from London Cockney (but see Maidment, 1994 for a critique of Estuary English). Wells (1994) provides evidence that RP has been influenced by Cockney in recent times and Crystal (1995) notes that Estuary English is increasingly heard in public domains and has begun to penetrate the Establishment. The rise of Estuary English has also occurred over a period which has seen a change in the attitude towards RP: whereas in the 1950s, RP enjoyed prestige status and Cockney was stigmatised, since at least the 1970s, many younger speakers have rejected RP (Burridge, 1998, Gimson, 1966) precisely because of its unfashionable association with the Establishment. Indeed, a form of RP associated with the upper classes is in more recent times often regarded as affected and the its speakers have for many 'become a figure of fun' (Gimson & Cruttenden, 1994). It is against this backdrop of the collapsing hierarchical social structures accompanied by the blurring between accents that mark social class as well as change in attitude to RP that the question of changes to the pronunciation in the Christmas broadcasts over a period of time are of great interest. As the representative of the very pinnacle of the Establishment as well as both the pre-eminent speaker and personification of the Queen's English, a form of English upheld and defended by many against 'debasement, ambiguity and other forms of misuse'[3], we can reasonably speculate that the Queen might be resistant to accent change from below. From another perspective, if we find further evidence for a shift in the Queen's accent in the direction of changes that have taken place in the wider community, then such a finding lends support to the view that adults throughout their lives cannot be immune from accent change that is taken place around them and it would also give credence to Labov's (1994) comment that the rate at which pronunciation changes may well be underestimated in apparent time studies.

Our investigation is principally concerned with the relationship between what has been termed upper-class, upper-crust or *U-RP* and *mainstream RP* (Wells, 1982). The distinction between "U" and and "non-U" goes back to Ross (1954) and Mitford (1956) and U-RP is described and partly caricatured by Wells (1982) as the accent of a dowager duchess, a Noel Coward sophisticate, a Terry Thomas cad or an upper-class army officer. This description of U-RP is more or less equivalent to *refined RP* used by Gimson & Cruttenden (1994). C*onservative RP* is defined by Gimson (1966) as a form of RP "used by the older generation and, traditionally, by certain professions or groups" and it is distinguished from *advanced RP*, a form of RP used by the younger generations (Gimson, 1964). Mainstream RP, which is also sometimes known as *Standard Southern British* (Deterding, 1997) is, according to Wells writing in 1982, spoken by the majority of RP speakers. It is the type of accent that is characteristic of many of the professional middle classes in England and is still exemplified by many non-regional BBC news presenters[4]: it is closest to what is sometimes described as *BBC English* which is the term used in the later editions of Daniel Jones' *English Pronouncing Dictionary* instead of Received Pronunciation (Roach & Hartman, 1997). Mainstream RP is distinct from *Estuary English* which, as discussed earlier, is a form of RP that is coloured by a London accent and Cockney (Coggle, 1993; Parsons, 1998; Rosewarne, 1994, 1996; Wells, 1994, 2003).

Auditory impressions suffice to show that the Queen's accent of both the 1950s and 1980s is not the same as mainstream RP. There are certainly some aspects of the Queen's accent that are characteristic of U-RP, such as: a markedly back /u/ (in the 1950s broadcasts at least), a noticeably intervocalic /r/ in words like 'very', an open-mid realisation of /æ/ (typical of 'conservative RP') and some lexically-specific U-RP characteristics such as a high back /ɔ/ vowel in 'lost' (to rhyme with 'forced').

In Harrington, Palethorpe and Watson (2000a, b) we suggested that there had been a shift in the Queen's monophthongal vowel space towards that of 1980s mainstream RP. We reached this conclusion by analysing the Christmas broadcasts from the 1950s and 1980s and demonstrating that the Queen's vowels from the 1980s were closer (in an auditorily transformed formant space) than those from the 1950s broadcasts to the averaged vowel space of five mainstream RP female BBC presenters from the 1980s. The main change was in the direction of a 'vertical' expansion of the vowel space, so that between the 1950s and 1980s, the high vowels had become higher and the open vowels had become more open. We also found a much smaller tendency for a 'horizontal' change, such as a marginally (but significantly) more fronted /u/ in the 1980s than in the 1950s broadcasts. The changes were in all cases in the direction of the vowels of 1980s mainstream RP speakers, *without actually attaining those vowel positions*: for example, the Queen's 1980s /æ/ was shown to be intermediate in phonetic openness (as judged from formant data) between the Queen's 1950s /æ/ and the /æ/ of the 1980s mainstream RP speakers, while the Queen's 1980s /u/ was not as front as the 1980s mainstream RP /u/, but more advanced than in the 1950s broadcasts.

In our present contribution, we aim to extend this analysis to the front and back rising diphthongs /aɪ eɪ aʊ oʊ/ in the same corpora (see Table I for the system of RP transcription used in this paper). As a precursor to the experimental analyses, we begin with a brief summary of some of the reported phonetic characteristics of these RP diphthongs, and of the way in which they are distinguished in some of the RP varieties. We also report whenever possible on the phonetic changes that may have taken place to these RP diphthongs in the last fifty years or so. Since, with the exception of some formant frequencies in /hVd/ words in Gimson & Cruttenden (1994), there are, as far as we know, no systematic experimental studies of RP diphthongs, this brief review is necessarily based on auditory analyses.

According to Gimson & Cruttenden (1994), RP /aɪ/ extends from a position midway between cardinal vowels 4 and 5 towards RP /ɪ/, although the second target is typically undershot and therefore more open and centralised compared with RP /ɪ/. In refined or U-RP, the first target of /aɪ/ has 'a very back starting point' (Gimson & Cruttenden, 1994). A back first component may also occur in Cockney (Wells, 1982) which, as in a broad Australian English accent (Harrington, Cox & Evans, 1997; Bernard, 1970, 1981), may also be rounded. There is some evidence that first component of mainstream RP /aɪ/ has retracted since the early to mid part of the 20^{th} Century (Wells, 1982), which would then correspond to a diachronic shift in the direction of a Cockney accent.

RP /eɪ/ extends from somewhere between cardinal vowels 2 and 3 (close to RP 'head') towards RP /ɪ/. Both Gimson & Cruttenden (1994) and Wells (1982) note that a phonetically close first target of

/eɪ/ is old-fashioned and characteristic of older RP speakers. In U-RP the first target of /eɪ/ is more open (presumably than in mainstream RP) and the entire diphthong may be monophthongised. Wells (1982) also comments that a London variety of RP has a backer first target while in London Cockney (as in Australian English) it is noticeably more open.

The phonetic definition of /oʊ/ is complex and there seems to be a good deal of variation both across RP varieties and age groups. There is some agreement that the first target of RP /oʊ/ has a central position and that the movement is in the direction of RP /ʊ/. The first component is raised and further back for older speakers whose first component starts near RP /ɔ/ (Wells, 1982). Refined RP has a fronted first target relative to mainstream RP (Gimson & Cruttenden, 1994) whereas in Cockney, the first component is more open than in mainstream RP and the second component may be completely unrounded.

There is some agreement that the first component of RP /aʊ/ has a starting position close to RP /ɑ/ and is therefore somewhat further back than the first component of RP /aɪ/. The movement of the diphthong is in the direction of RP /ʊ/ although it is often undershot to a mid position (Gimson & Cruttenden, 1994). There are few comments about changes to RP /aʊ/ in the last fifty years, although Wells (1982) has suggested that the first component is becoming backer while Gimson & Cruttenden (1994) list the increasing convergence of the first targets of /aɪ/ and /aʊ/ as a current RP change.

The experiments of this study are designed to test the hypothesis that the Queen's accent between the 1950s and 1980s has changed in the direction of a 1980s mainstream RP or Standard Southern British (SSB) accent. If this is so, then we should be able to predict more or less the positions of the diphthongs in the 1980s Christmas broadcasts by an extrapolation (in an F2 x F1 auditorily transformed space) between the diphthong positions in the 1950s Christmas broadcasts and those in a corpus of 1980s Standard Southern British speakers: that is, the 1980s Christmas broadcast diphthongs should be located in a space that is intermediary between the 1950s Christmas broadcast and 1980s SSB corpora, if our hypothesis is tenable.

In order to test this hypothesis, the experiments are divided into two parts. Firstly (section 2) we establish and compare the positions in an F2 x F1 space of the diphthongs and monophthongs in the 1980s SSB and 1950s Christmas broadcast corpora. Secondly, we form a set of predictions (section 3) about where in this space the 1980s Christmas broadcast diphthongs should be located, based on the hypothesis that they are located in approximately intermediary positions between those of the 1950s Christmas broadcasts and 1980s SSB positions. The predictions are then tested by comparing the relative proximity of monophthongs and diphthongs in these three spaces using classifications based on Bayesian distances.

2. Q50 and SSB corpora

2.1 Method

The materials included the diphthongs /aɪ aʊ ou eɪ/ in subsets of the same two Christmas broadcast and SSB corpora that were analysed in Harrington, Palethorpe and Watson (2000a, b). The Christmas broadcast corpora in the present study included six broadcasts, three each from the 1950s (henceforth the 'Q50 corpus') and three from the 1980s (henceforth the 'Q80 corpus'). The analysed vowels and diphthongs in the Q50 corpus are from 1952, 1954 and 1957 and in the Q80 corpus they are from 1983, 1985, and 1988. Harrington, Palethorpe and Watson (2000a, b) also analysed broadcasts from the late 60s/early 70s but we do not include these in the present study, since our analyses had shown a minimal change in the vowels between them and the broadcasts from the 1980s. For the Standard Southern British corpus (henceforth the 'SSB corpus'), we analysed the same five female BBC broadcasters from the 'Machine Readable Spoken English Corpus' (Roach, Knowles, Varadi and Arnfield, 1994) that were also the subject of the analysis in Deterding (1997). These speakers were recorded in the 1980s they are described by Deterding (1997) as having a 'Standard Southern British accent', a style of speech that 'may be familiar...through listening to the BBC World Service'.

We segmented and labelled in these three corpora all rising diphthongs /aɪ/, /eɪ/, /aʊ/, /oʊ/ that occurred in prosodically accented words from acoustic displays including a waveform, spectrogram, superimposed formant tracks and occasionally an amplitude contour using the EMU speech database analysis system (Cassidy & Harrington, 2001). (We did not include /oɪ/ because there were too few tokens in any one corpus).We marked the first target either in the most steady-state part of roughly the first half section of the diphthong trajectory and/or when F1 reached a maximum in /aɪ/ and /aʊ/. The formant trajectories were occasionally manually corrected in the case of obvious formant tracking errors. The segmentation and labelling were carried out by the first two authors of this paper. The formant values were also converted to the Bark scale (Zwicker, 1961) and all statistical analyses were applied to values in Bark. The numbers of tokens for each diphthong category in the three corpora are shown in Table I. The monophthong data in Table I are taken from Harrington, Palethorpe and Watson (2000a, b) and include the SSB monophthongs from Deterding (1997). In the Queen's data, we differentiated between /u/ that followed /j/ (as in 'few' and marked as /ju/ in Table I) from those that did not follow /j/ (marked as /u/ in Table I) since /j/ has a marked influence on the formants of this vowel.

Various plots of the diphthongs in an F2 x F1 Bark space were derived by linearly time-normalising all diphthong trajectories to the same 20 equally spaced intervals on the time-axis between their acoustic onset and offset (Fig. 1b) and then averaging the trajectories separately for each diphthong category (Fig. 1c) and plotting averaged time-normalised F2 against F1 (Fig. 1d). In the SSB corpus, the averages were calculated across the five female SSB speakers without applying any speaker-normalisation techniques.

2.2 Results

2.2.1 SSB diphthongs

Fig. 2 shows the average positions of the monophthongal vowels in the F2 x F1 Bark space of the five 1980s SSB speakers as reported in Harrington, Palethorpe and Watson (2000a, b) and Deterding (1997) with superimposed, linearly time-normalised and then averaged trajectories for the four separate diphthong categories. The figure shows a separation between the front-rising /aɪ eɪ/ and back-rising diphthongs /aʊ oʊ/ and there is also a separation in this F2 x F1 Bark space between /eɪ ou/ and /aɪ aʊ/ based on the first formant frequency.

The positions of the SSB diphthong trajectories in the F2 x F1 space are consistent with only some of the auditory analyses of mainstream RP reviewed earlier. There is an elbow (and hence a first target) for both /aɪ/ and /aʊ/ between /æ/ ('had') and /ɑ/ ('hard') supporting the interpretation that these diphthongs originate between these two vowels. The F2 x F1 trajectories are also consistent with auditory interpretations that the diphthongs end at positions which are phonetically somewhat higher than /ɛ/ ('head') and /ɒ/ ('pod') respectively. Fig. 2 also shows that the trajectories point towards /ɪ/ /ʊ/ and this, together with some evidence that the direction in which the diphthong points (rather than the final target attained) is the main factor that determines the phonetic identity of the second component (Gottfried, Miller and Meyer, 1993; Nearey & Assmann, 1986; Pols, 1977) may underlie the auditory impression that the second components of /aɪ/ and /aʊ/ are phonetically close to /ɪ/ and /ʊ/ respectively.

At first sight, there seems to be little support for the auditory impressions that the first target of /aʊ/ is phonetically more back than that of /aɪ/ since the elbow and therefore the targets of these diphthongs both originate from similar positions between the /æ/ and /ɑ/ spaces. However, although both diphthongs have their starting point in a similar location in the F2 x F1 space, it is also clear that /aɪ/ and /aʊ/ pass predominantly through the /æ/ and /ɑ/ ellipses respectively (Fig. 2). The movement of the trajectories through these separate vowel spaces may give rise to the percept that the first components of these diphthongs differ in backness with /aɪ/ being predominantly associated with /æ/ and /aʊ/ with /ɑ/.

For the first part of the /eɪ/ and /oʊ/ diphthong trajectories, F1 is low and considerably lower than the centroids of /ɛ/ ('head') and /ɜ/ ('herd') with which the first component of these diphthongs is assumed to be most closely associated in auditory impressions of mainstream RP. A separate analysis by speaker (Fig. 3) shows that the starting trajectory of /eɪ/ is nearer to /ɪ/ than to /ɛ/ for 4/5 speakers while for one speaker (S1) it is

about midway between /ɪ/ and /ɛ/. Fig. 3 also confirms that /oʊ/ originates from a position that is above /ɜ/ (i.e. with a lower F1) for 4/5 speakers.

As far as the second component is concerned, /eɪ/ moves through the /ɪ/ ellipse in the direction of the /i/ for all five speakers

(Fig. 3) without attaining the /i/ centroid. Fig. 3 also shows that /oʊ/ ends in the ellipse space, and nearer to the centroid of /u/ rather than /ʊ/.

2.2.2 Q50 diphthongs

The averaged diphthong trajectories superimposed on a monophthongal vowel space from the Queen's 1950s broadcasts in Fig. 4 show some similarities but also some very clear differences in comparison with the SSB vowels and diphthongs in Figs 2 and 3. One of the most striking differences is that there is considerably more compression along the F1 dimension so that in the Q50 data, the end of the /aɪ aʊ/ trajectories overlap with the beginning of the /eɪ oʊ/ trajectories whereas in the SSB data they do not. Another is that the Q50 diphthongs are confined to the front-central part of the F2 x F1 space and do not extend into, or come as close to, the space occupied by the rounded vowels and /ɑ/.

We consider first the acoustic structure of the Q50-/aɪ aʊ/ diphthongs. Fig. 4 shows that the first components of these diphthongs originate between Q50-/æ/ and Q50-/ɑ/. As in the SSB-corpus, there is no evidence that the first 'target' (corresponding to the elbows of the diphthong trajectories in Fig. 4) is phonetically more retracted in /aʊ/ than in /aɪ/; however, as in the SSB-corpus, /aɪ/, but not /aʊ/, passes through the Q50-/æ/ space on its way towards Q50-/ɪ/ (Fig. 4). On the other hand, whereas the SSB-/aʊ/ trajectory actually passes through SSB-/ɑ/, Q50-/aʊ/ has a higher F2 than Q50-/ɑ/ along its entire trajectory: this suggests that Q50-/aʊ/ is phonetically more front relative to Q50-/ɑ/ than SSB-/aʊ/ is to SSB-/ɑ/.

It is also evident from a comparison of Figs.2 and 4 that Q50-/aɪ aʊ/ start around one Bark higher than SSB-/aɪ aʊ/. This difference parallels the findings for the open monophthongs in Harrington, Palethorpe and Watson (2000a, b) which showed that F1 of /æ ɑ ʌ/ in the Q50 corpus were all significantly lower than in the SSB-corpus. In general, these data are consistent with the view that open vowels, including both open monophthongs as well as diphthongs with an open first component are phonetically closer in the Q50 compared with those in the SSB corpus. But by far the most marked open vowel difference between these corpora is in /æ/. Its F1 is almost 2 Bark lower in the Q50-corpus and, as discussed in Harrington, Palethorpe and Watson (2000a, b), this lends strong support to the phonetic impressions that /æ/ has become more open since the 1950s. Since the extent of F1 difference between the Q50 and SSB data is greatest for /æ/, it follows that the relationship between /æ/ and the first component of the /aɪ aʊ/ diphthongs is also slightly different in the two corpora: in the Q50 corpus, F1 of the first components of /aɪ/ and /aʊ/ are higher than F1 of /æ/ whereas in the SSB corpus, the first components of /aɪ aʊ/ and the monophthong /æ/ have quite similar F1 values. Taken together, these findings suggest firstly that the Q50-/aɪ aʊ/ diphthongs are phonetically more open than Q50-/æ/; secondly that all the open monophthongs and /aɪ aʊ/ are phonetically more open in the SSB than in the Q50 corpus; and thirdly that the extent of the openness differences between these two corpora is more pronounced in /æ/ than in /aɪ aʊ/.

As far as the second component of these diphthongs is concerned, Q50-/aɪ/ passes through the middle of the Q50-/ɛ/ ellipse and extends into the Q50-/ɪ/ space, terminating just below (i.e., with a higher F1) the Q50-/ɪ/ centroid. Q50-/aʊ/ passes from the Q50-/ʌ/ into the Q50-/ʊ/ space, ending within the Q50-/ʊ/ ellipse at a position roughly between the Q50-/ɪ/ and Q50-/ʊ/ centroids. The extent of the diphthong trajectories through the F2 x F1 space is quite comparable in the two corpora, but since all the phonetically close vowels in the SSB corpora are shifted upwards in the F2 x F1 space, as inferred by the lower F1 values for SSB-/i ɪ ʊ u ɔ/, the SSB-/aɪ aʊ/ diphthongs terminate well short of SSB-/ɪ ʊ/ and near SSB-/ɛ ɒ/, as discussed earlier. Another important difference is in the direction of the /aʊ/ trajectory. In the SSB data, /aʊ/ bends towards the back of the F2 x F1 space as it passes through SSB-/ɑ/, ending slightly above SSB-/ɒ/ and pointing towards SSB-/u ʊ/. In the Q50-corpus by contrast, the trajectory moves almost vertically upwards through the F2 x F1 space and is well forward (i.e. with a higher F2) of any of the Q50 rounded vowels. In general, these data suggest that /aɪ/ and /aʊ/ are a good deal less differentiated in the Q50 than in the SSB corpus.

Turning now to /eɪ/, in the Q50 corpus this diphthong originates between the Q50-/ɛ/ and Q50-/ɪ/ centroids and then extends in an arc around the high-F2 part of the Q50-/ɪ/ ellipse ending about midway between the Q50-/ɪ/ and Q50-/i/ centroids, but well within the Q50-/ɪ/ ellipse. In the Q50 and SSB corpora, the relationship between /eɪ/ and /ɪ/ is therefore quite similar. On the other hand, both the /eɪ/ trajectory and the /ɪ/ ellipse are acoustically much nearer to /i/ and further from /ɛ/ in the SSB than in the Q50 corpus.

The first component of the /oʊ/ diphthong is within the /ɜ/ and /ɛ/ ellipses in the Q50 corpus, but a good deal further from these monophthongs in the SSB corpus. The differences are evident both in the positional relationship between /oʊ/ and these monophthongs and in terms of absolute values (F1 and F2 of the first component of /oʊ/ are lower in the SSB corpus). These data suggest that the first component of /oʊ/ and /ɜ/ are phonetically fronted in the Q50 compared with the SSB corpus (see also Fig. 3 which shows that the starting trajectory of /oʊ/ does not extend into /ɛ/ for any of the SSB speakers). Considering now the second target, /oʊ/ extends through the edge of the /ʊ/ ellipse and points towards /u/ rather than /ʊ/ in both the Q50 and SSB corpora. Whereas /oʊ/ extends into the /u/ space for 3/5 SSB speakers, it reaches the edge of, but does not extend into, the /u/ ellipse in the Q50 corpus.

3. Q80 corpus

3.1 Predictions for Q80

We can make a number of predictions about the positions of the Q80 vowels in the F2 x F1 space based on a comparison between the Q50 and SSB corpora: that is, if the Queen's accent of the 1980s has shifted in the direction of a more mainstream 1980s RP, as Harrington, Palethorpe and Watson (2000a, b) have proposed, we should expect the Q80 diphthongs to be at approximately intermediate locations between those in the Q50 and SSB corpora. For example, because the first component of /aɪ/ and /æ/ are acoustically nearer to each other in the SSB corpus than in the Q50 corpus, we predict (see P1 below) that the first target of /aɪ/ will be acoustically nearer to /æ/ in the Q80 corpus. A summary of predictions about the differences in the positions of the diphthong in the F2 x F1 spaces between the Q50 and Q80 corpora based on an extrapolation from the Q50 to the SSB corpora are as follows. All these predictions are statements about the position of the Q80 diphthongs relative to those in the Q50 corpus:

Q80 /aɪ/

P1: The 1st target will be nearer to /æ/

P2: The 2nd target will be further from /ɪ/

P3: F1 of the 1st target will be higher

P4: F2 of the 1st target will be lower

Q80 /aʊ/

P5: The 1st target will be nearer to /æ/

P6: F1 of the 2nd target will be further from /ʊ/

P7: F1 of the1st target will be higher

P8: F2 of the 2nd target will be lower

Q80 /eɪ/

P9: The 1st target of /eɪ/ will be further from /ɛ/

P10: The 2nd target will be nearer to /i/

P11: F1 of the 1st target will be lower

P12: F1 of the 2nd target will be lower

Q80 /oʊ/

P13: The 1st target will be further from /ɜ/

P14: The 2nd target will be closer to /u/

P15: F1 of the 1st target will be lower

P16: F1 of the 2nd target will be lower

Some additional predictions can be made about how the diphthongs have changed relative to each other. The predictions are once again to do with the positional change of the diphthongs in the Q80 corpus relative to the Q50 corpus:

P17: The 2nd target of /aɪ/ will be further from the 1st target of /eɪ/

P18: The 2nd target of /aɪ/ will be further from the 1st target /oʊ/

P19: The 2nd targets of /oʊ/ and /aʊ/ will be further from each other

P20: The 2nd targets of /aɪ/ and / aʊ/ will be further away from each other

3.2 Method

We used a Gaussian probabilistic technique to quantify the extent to which two vowel types, let us call these /X/ and /Y/, overlapped with each other in the F2 x F1 Bark space. We did this by calculating the probability that a given token of /X/ could be a member of type /Y/; and we also calculated the probability that a given token of /Y/ could be a member of type /X/. We did this for all tokens of /X/ and for all tokens of /Y/ so that each token from /X/ and each token from /Y/ is associated with a probabilistic distance (that it could be a member of type /Y/ and /X/ respectively). We compared two sets of probabilistic distances, one set obtained for /X/ and /Y/ in the Q50 corpus and one set obtained for /X/ and /Y/ in the Q80 corpus. The probabilistic distances are in fact Bayesian distances which are derived from the density function of the normal curve with a known mean, variance and (in the multidimensional case) covariance matrix (Harrington & Cassidy, 1999, p. 252).

These steps are illustrated in Fig. 5 with reference to the first prediction, P1, that /æ/ and (the first target of) /aɪ/ overlap to a greater extent in the Q80 corpus. Firstly, a two-dimensional Gaussian model is fitted to the distribution of Q50-/æ/ which essentially converts every data point in the Q50 F2 x F1 space into a Bayesian distance that quantifies the likelihood that it could be a member of /æ/. Equal Bayesian distances form ellipses that are proportional to standard deviations away from the mean, or centroid, of /æ/ in the F2 x F1 space (Fig. 5b). The Bayesian distances to /æ/ can therefore be obtained for all the individual tokens (data points) of /aɪ/ (Fig. 5c). The process is then repeated, but now the Gaussian model is calculated for /aɪ/ (Fig. 5d) and the Bayesian distances are obtained from the tokens of /æ/ to /aɪ/ (Fig. 5e). The two sets of Bayesian distances (Figs.5c, 5e) thus obtained are combined in a single distribution (Fig. 5f). The more the values in Fig. 5f tend to zero, the greater the overlap between /aɪ/ and /æ/ in the Q50 corpus, while increasingly negative values imply a greater separation between these vowel types. The same methodology is used to derive a Bayesian distance distribution for /æ/ and /aɪ/ in the Q80 corpus. That is, we derive two versions of Fig. 5f: one distribution showing the pooled Bayesian distances between /æ/ and /aɪ/ in the Q50 corpus; and another showing the pooled Bayesian distances between /æ/ and /aɪ/ in the Q80 corpus. These two distributions are then compared statistically.

The null hypothesis for all the predictions is that these two distributions are not significantly different, which we would interpret to mean that the overlap in the F2 x F1 space between the two vowel types we are testing is about the same in the Q50 and Q80 corpora. Tests of significance in comparing these distributions of Bayesian distances are carried out using the non-parametric Wilcoxon rank sum test since (as subsequently confirmed by the Shapiro-Wilk W-statistic on these data) distributions of the kind shown in Fig. 5f are not Gaussian.

The above test of significance was used when we wanted to quantify the extent of overlap between two vowel types. When we wanted to compare the change in F1 or F2 Bark values between the Q50 and Q80 corpora, we applied a (parametric) Welch two-sample *t*-tests. Finally, when comparisons were made with the second target diphthong target, we obtained a notional second target by averaging the final 20% of the diphthong trajectory. We did this because finding a reliable second target for a diphthong is very difficult in view of the strong tendency for it to be undershot (Gay, 1968, 1970; Jha, 1985; Nâbélek & Dagenais, 1986; Pols, 1977).

3.3. Results: Q80 vowel positions

In general, the results support strongly the hypothesis that the Queen's diphthongs have shifted in the direction of the corresponding 1980s SSB positions. Both Fig. 6 and in particular Fig. 7 show that the linearly time-normalised Q80 trajectories occupy positions in the F2 x F1 Bark space that are intermediate between those in the Q50 and SSB corpora; the statistical analysis confirmed that 16/20 predictions were significant, most of these well below $p < 0.01$ (Table II).

Considering now the predictions in further detail, the results show that /æ/ is acoustically closer to the first targets of /aɪ/ and /aʊ/ (Table II: P1, P5) and that F1 of the first targets of /aɪ/ and /aʊ/ (P3 and P7) are higher in the Q80 corpus. These findings support the interpretation that all the vowels that can be characterised as phonetically open have become more open in the Q80 corpus. The extent of the shift has been greatest for /æ/; however, since /æ/ starts in the Q50 data from a phonetically higher position than /aɪ aʊ/, the net effect is a convergence so that /æ/ and the first components of /aɪ/ and /aʊ/ are less differentiated in the Q80 (and SSB) than in the Q50 corpus.

Whereas the first targets of /aɪ aʊ/ have become phonetically more open, those of /eɪ oʊ/ are associated with a lower F1 suggesting phonetic raising in line with the predicted changes from the Q50 to the SSB corpora (Table II: P11, P15). Since the shift in the mid vowels /ɛ ɜ/ from the Q50 to the Q80 corpora is small, these monophthongs are acoustically further from the first targets of /eɪ oʊ/ respectively in the Q80 corpus, which is a predicted change (Table II: P9, P13) in the direction of the SSB positions.

Turning now to the diphthongs' second targets, /aɪ/ and /aʊ/ terminate near the mid vowels /ɛ/ and /ɜ/ in the Q80 corpus (Figs. 6 and 7) and they are significantly further in the F2 x F1 Bark space from /ɪ/ and /ʊ/ respectively than in the Q50 corpus, in accordance with the predicted changes in the direction of SSB (Table II: P2, P6). The second target of /aɪ/ is also acoustically further from the first targets of /eɪ/ and /oʊ/ (Table II: P17, P18), partly because the entire /eɪ oʊ/ trajectories have shifted upwards in the F2 x F1 Bark space in the Q80 compared with the Q50 corpus. This shift is not only in absolute terms, confirming various predictions (Table II: P11, P12, P15, P16), but also relative to most of the other vowels and diphthongs: as a result, the second targets of /eɪ/ and /oʊ/ are acoustically nearer to /i/ and /u/ respectively in the Q80 than in the Q50 corpus, in accordance with the predicted shift in the direction of SSB (Table II: P10, P14). Since /oʊ/ has shifted upwards in the F2 x F1 Bark space and since the second target of /aʊ/ falls acoustically well short of the phonetically high vowels in the Q80 corpus, the second targets of /oʊ/ and /aʊ/ are also acoustically further apart from each other in the Q80 corpus, in accordance with the predicted shift in the direction of SSB (Table II: P19).

There is however one major non-significant change which is associated with F2 of /aʊ/. In the Q50 data, F2 of /aʊ/ is noticeably higher than in the SSB corpus; indeed, in the SSB corpus, /aʊ/ extends very clearly into the space associated with back vowels. There is no suggestion that F2 of /aʊ/ has lowered in the Q80 corpus (Table II: P8) and the trajectory extends more or less vertically upwards in the F2 x F1 space as in the Q50 corpus. Therefore, there is no evidence from this data to suggest that the second targets of /aɪ/ and /aʊ/ have become more divergent, in contrast to the prediction based on the interpolation from the Q50 to the SSB spaces (Table II: P20) .

4. Discussion

The acoustic analysis of the four diphthongs that we have investigated shows a very clear F1-dependent separation between /aɪ aʊ/ on the one hand and /eɪ oʊ/ on the other. In the 1980s Standard Southern British speakers, /aɪ aʊ/ both originate between SSB-/æ/ and SSB-/ɑ/. Phoneticians may have associated the first components of these diphthongs with open front and open back vowels respectively because, whereas SSB-/aɪ/ passes predominantly through the SSB-/æ/ space, the trajectory of SSB-/aʊ/ bends in the opposite direction through SSB-/ɑ/, as these diphthongs extend upwards in the F2 x F1 formant plane. Their trajectories terminate at positions corresponding approximately to phonetically mid in the SSB space, and they end well below the first components of SSB-/eɪ oʊ/. For all five SSB-speakers, the first components of SSB-/eɪ oʊ/ generally start above the spaces associated with SSB-/ɛ/

and SSB-/ɜ/. The second component of SSB-/eɪ/ is about midway between the SSB-/ɪ/ and SSB-/i/ spaces and the second component of SSB-/oʊ/ is nearer to SSB-/u/ than to SSB-/ʊ/.

As reported in Harrington, Palethorpe and Watson (2000a, b), there has been a 'vertical expansion' of the Queen's monophthongal vowel space between the 1950s and 1980s as evidenced by an F1-raising of the open monophthongs /ɑ ʌ æ ɒ/ and an F1-lowering of the half-close and close monophthongs /ɪ ʊ ɔ i u/. In the light of these and other changes which were shown to be in the direction of the SSB-monophthongal vowels, we were interested to know whether there would be an analogous shift in the diphthong trajectories.

The results of this present study show a change from the 1950s to the 1980s Christmas broadcast corpora such that the diphthongs with an open first component, /aɪ aʊ/, have moved downwards in the F2 x F1 space with /ɑ ʌ æ ɒ/ , while the diphthongs with a mid-high first component, /eɪ oʊ/, have shifted upwards with /ɪ ʊ ɔ i u/. Since the extent of the trajectory between the first and second targets within the same diphthong category is about the same across the two sets of years, /aɪ aʊ/ overlap a good deal with (the first components of) /eɪ oʊ/ in the 1950s broadcasts, but they do not in the broadcasts of the 1980s. For the same reason, /aɪ aʊ/ extend well into the /ɪ ʊ/ spaces in the 1950s broadcasts, but they do not in those of the 1980s; and the first components of /eɪ oʊ/ start within the /ɛ ɜ/ spaces in the 1950s broadcasts, but they are on the upper edge of them in 1980 broadcasts.

These are precisely some of the changes that are predicted to occur by extrapolating between the acoustic vowel spaces of the 1950s Christmas broadcast and 1980s SSB corpora. Moreover, we can see that, as for the monophthongs, the diphthongs have shifted towards 1980s SSB positions, but without actually attaining them. These data would seem to corroborate the interpretation in Harrington, Palethorpe and Watson (2000a, b) that between the 1950s and the 1980s the Queen's accent has shifted towards, but not attained, a 1980s Standard Southern British accent.

But there is also evidence that the Queen's accent has not changed in some respects. Consider for example /aʊ/. The trajectory rises vertically up through the F2 x F1 space in both the 1950s and 1980s Christmas broadcast data, whereas in the 1980s SSB data, /aʊ/ very clearly bends towards the back of it. As a result, the /aɪ/ and /aʊ/ trajectories have not become more divergent in the 1980s Christmas broadcast data, contrary to the predicted change based on their extrapolation towards the SSB positions. It has been suggested[5] that the lack of divergence in the F2 x F1 space between /aɪ/ and /aʊ/ may provide experimental evidence for the popular impression that these diphthongs are less distinct in U-RP, giving rise to the caricature that the pronunciation by some members of the Royal Family of 'house' and 'trousers' is close to 'hice' and 'trisers' (Wales, 1994). Perhaps the fronting of the second component of /aʊ/ has come to be identified as an especially salient marker of the U-RP/SSB distinction, in the same way that glottal stop is a perceptually dramatic marker of London Cockney. We can speculate that if U-RP speakers (and the Queen) choose to set themselves apart from those with a mainstream RP accent, then they may do so precisely by resisting change in those attributes that provide especially prominent cues to the U-RP/SSB distinction.

Although the changes to both the diphthongs reported in this paper as well as to the monophthongs in Harrington, Palethorpe and Watson (2000a, b) can be interpreted as evidence for *phonetic* changes in the direction of Standard Southern British, we must also consider the following alternative explanations: that they are attributable to *age-related* or *speaking-style* changes.

In Harrington, Palethorpe and Watson (2000b) we rejected an age-related explanation because the formant differences in the positions of the 1950s and 1980s Christmas broadcast monophthongs were not compatible with some recent experimental findings by Rastatter, McGuire, Kalinowski, Stuart (1997) who examined formant differences between a young and old age group producing similar materials. It is however clear from the data in Table III that there are global changes to the formant positions between the 1950s and the 1980s broadcasts. These data were derived by calculating means and standard deviations across all the frames of data from vowel onset to vowel offset in the same monophthongs that were analysed in Harrington, Palethorpe and Watson (2000b). Turning firstly to the means, Table III shows that all the F1-F4 means have decreased in the 1980s Christmas broadcasts relative to those from the 1950s by up to 0.4 Bark (in F4). Interestingly, this decrease in predicted by Laver & Trudgill (1979) who suggest, based on remarks by Luchsinger & Arnold (1965), that the entire respiratory system and digestive tract are in a lower position with increasing age. Such a lowering would cause the vocal tract to lengthen resulting in the well-known downwards shift in the formant frequency ranges (Laver & Trudgill, 1979, p.10). We also see in Table III an age-related change in fundamental frequency which is reported to be lower in older women (Helfrich, 1979): as Table III shows, the mean F0 in the 1980s broadcasts is some 0.5 Bark lower in the 1980s than in the

1950s broadcasts. However, although we recognise that some of the formant displacement may be due to maturation of the vocal tract, there is no evidence of a general lowering of F1 and F2 in the Queen's 1980s monophthongs and diphthongs relative to those of the 1950s (Fig. 7). In particular, the dramatic F1-rise in open monophthongs and diphthongs with an open first component as well as the F2-increase in /u/ are not compatible with such an age-related explanation. Nevertheless, we cannot entirely rule out an age-related component in the F1-lowering of the first component of both the /eɪ oʊ/ diphthongs and of some of the close monophthongs, but there are three points that need to be considered with regard to such an interpretation. Firstly, the relationship between formant lowering and increasing age needs to be investigated further: the study by Rastatter et al. (1997) for example suggests that the formants may *rise* from younger to older speakers in some vowels. Secondly, the F1 decrease in /eɪ oʊ/ in the 1980s Christmas broadcasts relative to the 1950s broadcasts is somewhat greater than the average F1 change between these two corpora: in particular, there is approximately a 0.75 Bark decrease in the first targets of /eɪ oʊ/ whereas the data in Table III show an average 0.2 Bark decrease between these two time periods. Thirdly, we would have to reconcile a purely age-related explanation with our evidence that the changes to the vowel and diphthong positions between the two sets of Christmas broadcast materials are very clearly in the direction of the 1980s Standard Southern British corpus. An age-related explanation of the F1-changes in /eɪ oʊ/ would then lead us to the conclusion that the 1980s SSB speakers must have been a good deal older than the Queen (aged 57-62 between 1983 and 1988) given that the SSB F1-values for these diphthongs and close vowels are even lower than those in the 1980s Christmas broadcasts. Deterding (1997) does not list the ages of the SSB speakers, but we have no reason to believe from our auditory impressions of them that they are all considerably older than 57-62. For all these reasons, we continue to reject an age-related explanation for the formant changes we have presented in this study and in Harrington, Palethorpe and Watson (2000a, b).

The influence of speaking-style changes is another matter, however. One interpretation that was considered in Harrington, Palethorpe and Watson (2000a, b) is the following. Newsreaders and broadcasters such as those analysed in this 1980s SSB corpus are likely to have a particularly clear speaking-style. The Queen may have learned to increase the clarity of the delivery of the broadcasts over the 30 year period we have examined; the incremental shifts in the Queen's Christmas broadcast data towards the 1980s SSB values would (under this interpretation) simply be the consequence of an increase in the clarity of speaking-style, not of changes to the phonetic quality of vowels. An expansion on the F1-dimension, of the kind that we have found, could be the acoustic consequence of an increased phonetic height differentiation: that is, an increase in clarity might be effected by a 'vertical hyperarticulation' in which phonetically high vowels are produced with a more raised jaw position and greater vocal tract narrowing and phonetically low vowels are produced with a lower jaw position and greater vocal tract opening (resulting in an expansion along the F1 dimension, given the well-known articulatory-acoustic relationship demonstrated by e.g. Lindblom & Sundberg, (1971) between a greater vocal tract opening/jaw lowering and F1-raising). Certainly, there is very clear evidence for a change in the range (standard deviation) of the formants from the 1950s to the 1980s Christmas broadcasts: in the 1980s broadcasts, the standard deviation of F1 is 0.5 Bark higher than in the 1950s broadcasts (Table III). On the other hand, we also reasoned in Harrington, Palethorpe and Watson (2000a, b) that a hyperarticulation of the vowel space with the goal of increasing communicative clarity has been shown to affect F1 *and* F2: that is, hyperarticulated and more clearly produced vowels become more *peripheral*, i.e. are produced further from the centre of the F2 x F1 vowel space (de Jong, 1995; Harrington, Fletcher and Beckman, 2000; Moon & Lindblom, 1994). But there is no evidence for an F2-expansion in the vowels of the Queen's Christmas broadcasts, nor of a global increase in the standard deviation of formants above F1 (Table III). In fact, F2 is lowered in many of the front vowels from the 1950s to the 1980s Christmas broadcasts while F2 of /u/ is raised (i.e. there is a modest *compression* along much of the F2 dimension) and both of these changes are in the direction of the F2 positions in the 1980s Standard Southern British vowel space. A hyperarticulation of the vowel space also does not seem to be compatible with global changes to the vowel duration between the 1950s and 1980s broadcasts: a subsequent analysis has shown that the average vowel duration in the 1980s Christmas broadcasts is significantly less than the 1950s for almost all of the monophthongal and diphthongal vowel types, which suggests that the rate at which the Christmas broadcasts is delivered has increased in the later years. An explanation couched in terms of a general hyperarticulation of the vowel space between the 1950s and 1980s broadcasts would also need to reconcile an expansion along the F1-dimension in monophthongs with a non-significant F1-change that we have also established in the *extent* of the diphthong trajectories between their first and second components. That is, if there were global changes of speaking style from the 1950s to the 1980s Christmas broadcasts resulting primarily in a 'vertical' hyperarticulation and F1-expansion, we might expect the extent of F1-change from the first to the second components of /aɪ/ and /aʊ/ to increase from the 1950s to the 1980s broadcasts but the data are not consistent with this interpretation.

But on the other hand, if we abandon completely an interpretation of the results of this paper in terms of speaking-style changes, then we are forced into an explanation based on phonetic changes (assuming that we can generally discount the age-related explanation). Under a purely phonetic interpretation, we would have to conclude that speakers of U-RP from the 1950s compress their vowels a good deal on the phonetic height dimension and expand them slightly on the phonetic backness dimension in comparison with 1980s SSB speakers. Certainly, there are many reported phonetic changes from an older and more conservative variety RP to its more modern usage that are entirely compatible with such a general interpretation, including: the lowering of /æ/ (Gimson, 1966; Ramsaran, 1990; Wells, 1982); the fronting of /u/ (Gimson, 1966); the fronting of the second component of /oʊ/ (Wells, 1982), the fronting of /ʌ/ (Gimson, 1964 -- although, like Bauer, 1985, we have found no evidence for this change in our data) the raising of /ɪ/ (Bauer, 1985, but see Wells, 1982 for a different interpretation), the raising of /ɔ/ (Gimson, 1966; Wells, 1982) and the possible convergence of the first targets of /aɪ/ and /aʊ/ (Wells, 1982). Yet some of the changes we have found, such as the upward shift in the F2 x F1 space of the entire /eɪ/ and /oʊ/ trajectories find scant support in impressionistic analyses of RP change in the literature.

It seems then, we must come back to John Laver's most apt metaphor, that even after our analyses of almost 3000 vowels and diphthongs from the same speaker producing speech with a similar communicative intent, we have not yet quite managed to delineate sufficiently the figure from the ground against which it is set. The matter of what is phonetic and what is speaker-specific and more generally the characterisation of speakers should "rest on a scientific base, and not rely on impressionistic description which is idiosyncratic to the individual analyst" (Laver, 1980, p. 7). Extending the range of available large-scale speech corpora for the 'real time' longitudinal analysis of many other types of speakers of Received Pronunciation and other varieties of English will help us to achieve this important aim.

Table I. The number of accented vowel and diphthong tokens analysed in the Q50, Q80 and SSB corpora. In the SSB corpus, there is no differentiation between /u/ and /ju/ which are collapsed in the /u/ (n=42) category.

Corpus		Q50	Q80	SSB
lax monophthongs				
ɪ	(hid)	103	123	60
ɛ	(head)	138	156	62
æ	(had)	96	77	68
ʌ	(run)	80	83	62
ɒ	(pod)	62	64	51
ʊ	(hood)	8	21	26
tense monophthongs				
i	(heed)	82	82	57
ɜ	(herd)	37	33	42
ɑ	(hard)	35	42	68
ɔ	(hoard)	78	82	55
u	(who'd)	24	17	42
ju	(few)	31	37	
diphthongs				
aɪ	(hide)	90	78	198
eɪ	(hay)	89	118	276
aʊ	(how)	24	19	128
oʊ	(hoe)	75	64	214

Table II: Results of the statistical analysis of twenty predicted changes in the Q80 corpus relative to the Q50 corpus. W: Wilcoxon rank sum test with continuity correction. t: t-statistic of Welch Two Sample t-test.

Code	Prediction	Result
	/aɪ/	
P1	1st target of /aɪ/ is closer to /æ/	W = 5978, $p < 0.00001$
P2	2nd target of /aɪ/ is further from /ɪ/	W= 25268, $p < 0.00001$
P3	F1 of the 1st target of /aɪ/ is higher	t = -2.00, df = 164.6, $p < 0.05$
P4	F2 of the 1st target of /aɪ/ is lower	NS
	/aʊ/	
P5	1st target of /aʊ/ is closer to /æ/	W = 3329, $p < 0.00001$
P6	2nd target of /aʊ/ is further from /ʊ/	W = 1263, $p < 0.00001$
P7	F1 of the 1st target is higher	t = -2.81, df = 40.6, $p < 0.001$
P8	F2 of the 2nd target is lower	NS
	/eɪ/	
P9	1st target of /eɪ/ is further from /ɛ/	W = 44232, $p < 0.00001$
P10	2nd target of /eɪ/ is closer to /i/	W = 13275, $p < 0.0001$
P11	F1 of the 1st target of /eɪ/ is lower	t = 10.36, df = 198.2, $p < 0.00001$
P12	F1 of the 2nd target of /eɪ/ is lower	t = 9.83, df = 164.2, $p < 0.00001$
	/oʊ/	
P13	1st target of /oʊ/ is further from /ɜ/	W = 8056, $p < 0.00001$
P14	2nd target is of /oʊ/ is closer to /u/	W = 4599, $p < 0.00001$
P15	F1 of the 1st target of /oʊ/ is lower	t = 10.4, df = 198.2, $p < 0.00001$
P16	F1 of the 2nd target of /oʊ/ is lower	t = 8.00, df = 134.2, $p < 0.00001$
	Interdiphthong comparisons	
P17	2nd target of /aɪ/ is further from the 1st target of /eɪ/	W = 25558, $p < 0.00001$
P18	2nd target of /aɪ/ is further from the 1st target /oʊ/	W = 15404, $p < 0.00001$
P19	2nd target of /oʊ/ is further from the 2nd target of /aʊ/	W = 6440, $p < 0.00001$
P20	2nd target of /aɪ/ is further from the 2nd target of /aʊ/	NS

Table III. Means (m) and standard deviations (s) of the first four formants (F1-F4) and of the fundamental frequency (f0) in the Q50 and Q80 corpora in Bark (the third row shows the difference in the mean and standard deviation values between the two corpora). The means and standard deviations were calculated across all the frames (Q50: *n*= 20429 frames; Q80: *n* = 17972 frames) from acoustic onset to offset of the monophthongs in Harrington, Palethorpe and Watson (2000a, b). The values are corrected to one decimal place.

	F1		F2		F3		F4		f0	
	m	s	m	s	m	s	m	s	m	s
Q50	5.5	1.1	12.0	2.3	15.5	0.5	17.5	0.6	2.8	0.6
Q80	5.3	1.6	11.8	2.2	15.4	0.5	17.1	0.6	2.3	0.5
Change	-0.2	0.5	-0.2	-0.1	-0.1	0	-0.4	0	-0.5	-0.3

Notes.

1. "Don't be fooled: the Queen is not speaking our language." Article by Philip Hensher in *The Independent*, 22nd December, 2000.

2. Editorial column, *The Times*, 21st December, 2000.

3. From the Queen's English Society: http://www.author.co.uk/qes/

4. For example, Julian Marshall, presenter of Newshour on the BBC World Service.

5. My thanks to Francis Nolan for this observation.

Acknowledgements

We thank Buckingham Palace for granting permission to analyse the Christmas broadcasts and the BBC for providing us with the recordings from their archives. The paper has benefited from the comments from a number of Colleagues at the Institut für Phonetik und Sprachliche Kommunikation, Munich, the Department of Linguistics, Cambridge University, the Istituto di Fonetica e

Dialettologia, del CNR, Padua and the Institute of Phonetics, Saarland University where earlier versions of this research were presented in 2002.

Figures

Fig.1. (a) F1 and F2 trajectories of three /aɪ/ diphthongs. In (b) these trajectories are linearly time-normalised and in (c) the time-normalised trajectories in (b) are averaged at 20 equally spaced time intervals (shown by dotted lines in (c)). (d) is a plot of the averaged F1 against the averaged F2 derived from (c).

Fig. 2: Averaged SSB diphthong trajectories for /aɪ/ /eɪ/ (solid: /aɪ/ is below /eɪ/) and /aʊ/ /oʊ/ (dashed: /aʊ/ is below /oʊ/) superimposed on the SSB vowel monophthongs in an F1 x F2 Bark space. The vowel symbol is plotted at the centroid (mean) and the ellipses include 2 standard deviations around the mean.

Fig.3. As Fig.2, but plotted separately for the five (S1-S5) speakers and showing only /eɪ/ (solid) and /oʊ/ (dotted). (There are no /ʊ/ tokens for S2 and S4 because of the number of tokens was too small to form an ellipse).

Fig. 4: Averaged Q50 diphthong trajectories for /aɪ/ /eɪ/ (solid: /aɪ/ is below /eɪ/) and /aʊ/ /oʊ/ (dashed: /aʊ/ is below /oʊ/) superimposed on the Q50 vowel monophthongs in an F1 x F2 Bark space with (a) and

without (b) 2 standard deviation ellipses around the monophthongal means. (c) is an enlarged version of (a) but showing only the lax vowels and /ɑ/ on a .

Fig.5. An illustration of the technique for comparing the extent to which two vowel categories (here /æ/ and /aɪ/ in panel (a)) overlap in the F2 x F1 Bark space in the Q50 corpus. Panel (b): /aɪ/ tokens superimposed on ellipses at σ standard deviations away from the centroid of /æ/. Panel (c): the Bayesian distances of the /aɪ/ tokens to /æ/ calculated using the model in (b). Panels (d) and (e): as panels (b) and (c) respectively, but in which Bayesian distances are calculated from the tokens of /æ/ to /aɪ/. Panel (f): a combination of the distributions in panels (c) and (e).

Fig. 6: Averaged Q80 diphthong trajectories for /aɪ/ /eɪ/ (solid: /aɪ/ is below /eɪ/) and /aʊ/ /oʊ/ (dashed: /aʊ/ is below /oʊ/) superimposed on the Q80 vowel monophthongs in an F2 x F1 Bark space. The vowel symbol is plotted at the centroid (mean) and the ellipses include 2 standard deviations around the mean.

Fig. 7. Q50 (left), Q80 (middle) and SSB (right) averaged diphthong trajectories superimposed on the vowel centroid positions from the corresponding corpora. /aɪ eɪ/ are the solid trajectories (/aɪ/ is below /eɪ/) and /aʊ oʊ/ are the dotted trajectories (/aʊ/ is below /oʊ/).

References

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

Burridge, K. (1998). *English in Australia and New Zealand: an Introduction*. Melbourne: Oxford University Press.

Bernard, J. R. L. (1970). Towards the acoustic specification of Australian English, *Zeitschrift fur Phonetik*, 2/3, 113–128.

Bernard, J. R. L. (1981). Australian pronunciation. In *The Macquarie Dictionary*, edited by A. Delbridge. Macquarie Library, Sydney, pp. 18– 27.

Cannadine, D. (1998)*Class in Britain* (Yale Univ. Press, New Haven,.

Chambers, J.K. & Trudgill, P. (1980). *Dialectology*. Cambridge: Cambridge University Press.

Cassidy, S. and Harrington, J. (2001). Multi-level annotation in the Emu speech database management system. *Speech Communication*, 33, 61-77.

Coggle P (1993). *Do You Speak Estuary?* London: Bloomsbury.

Crystal, D. (1988) *The English Language*. Mouton: The Hague.

Crystal, D. (1995). Estuary English. In Crystal, D. *Cambridge Encyclopaedia of the English Language*. p. 327. Cambridge: Cambridge University Press. (Also available from the website in Wells, 2003 listed below).

de Jong K. (1995). The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, **97**, 491-504.

Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association*, **27**, 47-55.

Ellis A.J. (1869) *On Early English Pronunciation*. English Early Text Society: London.

.

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

Gay, T. (1970). A perceptual study of American English diphthongs. *Language & Speech*, 13, 65-88.

Gimson, A.C. (1964). Phonetic changes and the RP vowel system. In D. Abercrombie, D. Fry, P. MacCarthy, N. Scott, J. Trim (eds.), *In Honour of Daniel Jones*. Longmans: London (pp. 131-136).

Gimson, A.C. (1966, 1970, 1980). *An Introduction to the Pronunciation of English* (1st, 2nd, and 3rd editions). London: Edward Arnold

Gimson, A.C. and Cruttenden, A. (1994). *Gimson's Pronunciation of English* (5th edition, revised by Alan Cruttenden). London: Edward Arnold.

Gottfried, M., Miller, J.D., and Meyer, D.J. (1993). Three approaches to the classification of American English diphthongs. *Journal of Phonetics*, 21, 205-229.

Harrington J. & Cassidy S. (1999). *Techniques in Speech Acoustics*. Kluwer Academic Publishers: Foris, Dordrecht. ISBN: 0-7923-5731-0.

Harrington J., Cox F, and Evans Z. (1997) An acoustic phonetic study of broad, general, and cultivated Australian English vowels. *Australian Journal of Linguistics*, 17, 155-184.

Harrington J., Fletcher J. and Beckman, M.E. (2000) Manner and place conflicts in the articulation of accent in Australian English. In Broe M. (editor), *Papers in Laboratory Phonology*, 5. (p. 40-55). Cambridge University Press: Cambridge.

Harrington, J., Palethorpe, S., and Watson, C. (2000a). Does the Queen speak the Queen's English? *Nature*, 408, 927-928.

Harrington, J., Palethorpe, S., and Watson, C. (2000b). Monophthongal vowel changes in Received Pronunciation: an Acoustic Analysis of the Queen’s Christmas Broadcasts. *Journal of the International Phonetic Association*, 63-78.

Helfrich, H. (1979). Age markers in speech. In K. Scherer & H. Giles (eds.) *Social Markers in Speech*. Cambridge University Press: Cambridge (p. 63-107).

Jha, S.K. (1985). Acoustic analysis of Maithilli diphthongs. *Journal of Phonetics*, 13, 107-115.

Jones, Daniel. (1924). *English Pronouncing Dictionary* (2nd Edition). London: J.M. Dent & Sons.

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

Laver, J. (1980). *The Phonetic Description of Voice* Quality. Cambridge University Press: Cambridge.

Laver, J. (1994). *Principles of Phonetics*. Cambridge University Press: Cambridge.

Laver J. & Trudgill, P. (1979). Phonetic and linguistic markers in speech. In K. Scherer & H. Giles (eds.) *Social Markers in Speech*. Cambridge University Press: Cambridge. (pp. 1-26).

Lindblom, B. & Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *Journal of the Acoustical Society of America*, 42, 830-843.

Luchsinger, R. & Arnold, G.E. (1965). *Voice-Speech-Language*: Belmont: Wadsworth Publishing Company.

Maidment, J.A. (1994). Estuary English: Hybrid or hype? Paper presented at the 4th New Zealand Conference on Language & Society, Lincoln University, Christchurch, New Zealand, August, 1994. (The full text is available from the website given in Wells, 2003, below).

Mitford, N. (1956) *Nobless Oblige: Enquiry Into Identifiable Characteristics of English Aristocracy*, London: Penguin.

Moon, S.-J., and Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed words, *Journal of the Acoustical Society of America*, 96, 40-55.

Nâbélek, A.K. & Dagenais, P.A. (1986). Vowel errors in noise and in reverberation by hearing-impaired listeners. *Journal of the Acoustical Society of America*, 80, 741-748.

Nearey, T.M. & Assmann, P.F. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America*, 80, 1297-1308.

Parsons, G. (1998) From "RP" to "Estuary English": the concept 'received' and the debate about British pronunciation standards. MA Manuscript, Hamburg. Available from the website listed under Wells, 2003).

Pols, L.C.W. (1977). Spectral analysis and identification of Dutch vowels. Unpublished doctoral dissertation, University of Amsterdam, the Netherlands.

Ramsaran, S. (1990). RP: fact and fiction. In S. Ramsaran (Ed.) *Studies in the Pronunciation of English*. London and New York: Routledge. (p.178-190).

Rastatter, M., McGuire, R., Kalinowski, J. and Stuart, A. (1997). Formant frequency characteristics of elderly speakers in contextual speech. *Folia Phoniatrica*, 49, 1-8.

Roach, P. & Hartman. J. (1997) *English Pronouncing Dictionary* (15th Edition). Cambridge: Cambridge University Press.

Roach, P., Knowles, G., Varadi, T. and Arnfield, S. (1994). MARSEC: A Machine-Readable Spoken English Corpus. *Journal of the International Phonetic Association*, 23, 47-54.

Rosewarne, D. (1984). Estuary English. *Times Educational Supplement*, 19. (The full text is available from the website in Wells, 2003 listed below).

Rosewarne, D. (1994). Pronouncing Estuary English. *English Today* 40, Vol. 10/4 3-6.

Rosewarne, D. (1996). Estuary as a World Language*. Modern English Teacher* 5 13-17.

Ross A S C 1954 Linguistic class indicators in present-day English. *Neuphilologische Mitteilungen* 16: 171-85

Wales, K. (1994) Royalese: the Rise and Fall of "The Queen's English"'. *English Today* 39, Vol. 10/3, 3-10.

Watson, C.I. and Harrington, J. (1999) Acoustic evidence for dynamic formant trajectories in Australian English Vowels. *Journal of the Acoustical Society of America*, 106, 458-468.

Wells, J. (1994). English Accents in England. In P. Trudgill (Ed)., *Language in the British Isles*. Cambridge University Press: Cambridge.

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

Wells, J.C. (2003). Web documents relating to Estuary English. http://www.phon.ucl.ac.uk/home/estuary/home.htm

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*. , 267-292. Cambridge: Cambridge University Press.

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *Journal of the Acoustical Society of America*, **33**, 248.