Journal of Phonetics 38 (2010) 526–539

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Vowel systems and accent similarity in the British Isles: Exploiting multidimensional acoustic distances in phonetics

Emmanuel Ferragne *, François Pellegrino

Laboratoire Dynamique Du Langage—UMR 5596, CNRS—Université Lyon 2, 14 avenue Berthelot, 69007 Lyon, France

ARTICLE INFO

Article history:
Received 23 September 2008
Received in revised form
5 January 2010
Accepted 16 July 2010

ABSTRACT

We illustrate how a high-dimension feature space typically used in speech technology can be adapted to the phonetic description of vowels in 13 accents of the British Isles. In a previous work (Ferragne & Pellegrino, 2010), we carried out a formant investigation of the vowel systems of the British Isles; due to erroneous formant estimation, two-thirds of the speakers had to be left out. The present article is therefore an attempt to overcome the methodological difficulties brought about by the use of formants. This novel methodology makes use of distances between vowels in the Mel-Frequency Cepstral Coefficient (MFCC) space. First, hierarchical clustering and multidimensional scaling (MDS) are applied, and tree diagrams and MDS plots are displayed in order to make the data phonetically interpretable. By making distances explicit, this approach to acoustic vowel description facilitates the spotting of phonemic mergers and splits. This part of the study is complemented with an exploratory analysis of the duration of some vowel pairs whose members are acoustically very close to each other. Second, correlations between individual vowel distance matrices are computed, yielding an estimate of the acoustic distance between accents. The explanatory power of these distances is then assessed with hierarchical clustering and MDS. Our ultimate goal is to draw a parallel between the findings obtained with our unconventional method and previous phonetic descriptions, and to benchmark this new methodology against the results in Ferragne and Pellegrino (2010).

1. Introduction

1.1. Acoustic description of vowels

For the past sixty-odd years, at least ever since Martin Joos published his *Acoustic Phonetics* in 1948, it has been common practice in the phonetic literature to plot vowels in the F1/F2 plane with reversed axes so that a close match appears between formant values and the more impressionistic dimensions of openness and backness. Most acoustic studies of vowels, including those concerned with accent variation, still rely on formant measurements, very probably because they combine two advantages. Firstly, they allow a parsimonious representation of vowel quality, i.e. the first two or three vocalic formants are generally deemed sufficient for phonetic descriptive purposes. Besides, trivial though this may seem, vowel coordinates (like any other coordinates) can be satisfactorily represented on graphs in parameter spaces not exceeding 3 dimensions; so including more parameters would lead to either giving up the idea of graphing the data, or resorting to dimension-reduction techniques (with the risk of distorting the original data).

Secondly, the formant-based representation of vowels correlates very well with traditional triangular/trapezoidal plots. In other words, when the formant vowel chart appeared, it came down to nothing more than applying technological advances to a pre-existing descriptive framework. Hence, nowadays, the two approaches (formant charts and auditory plots based on cardinal vowels) can co-exist, and even reinforce each other.

Despite such advantages, the formant approach has a serious drawback: formant detection and formant tracking are not totally reliable techniques (de Wet et al., 2004), which, at least partly, explains why formants are not normally used in automatic speech recognition. As a consequence, the intervention of a human expert is required, which, in turn, means that analyzing large speech corpora can take an overly long time, strict reproducibility is not fully guaranteed, and formants are not suitable for real-time applications. In addition to that, high-dimension feature spaces typically used in speech technology (Gold & Morgan, 2000, pp. 295–308) produce a more extensive representation of vowel quality, and cepstral analysis is known to be less sensitive to individual variation than LPC-based formant estimation (Gold & Morgan, 2000, p. 290). This is supported by de Wet et al. (2004): they performed automatic vowel classification with linear discriminant analysis using 3 types of parameters: hand-labelled formants, automatically extracted formant-like features, and Mel-Frequency Cepstral Coefficients (MFCCs). Their results show

---

* Corresponding author. Tel.: +33 4 72 72 65 35; fax: +33 4 72 72 65 90.
E-mail address: Emmanuel.Ferragne@univ-lyon2.fr (E. Ferragne).

0095-4470/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2010.07.002

that, not only do MFCCs outperform automatic formant-like features, but they are also found to be superior to hand-labelled formants. A final distinction between formants and MFCCs is that formant estimation for vowel description is based on prior expectations as to what values should show up in a given vowel according to models of the vocal tract. As real vocal tracts hardly ever behave exactly like a model, unexpected peaks in the spectrum are sometimes mistakenly identified as formants. Whenever mismatches occur between expected and observed formant values, the estimate in question can clearly be said to be wrong. On the contrary, cepstral analysis does not specify a priori a limited set of events that should be located in vowel-dependent frequency bands. So there is no prior reference against which MFCCs can be compared; in other words, MFCCs may be sensitive to background noise, individual non-phonetic variation, but cannot really be wrong.

1.2. *Accent studies*

As opposed to many recent accent studies in the British Isles (e.g. Foulkes & Docherty, 1999) whose primary focus is on sociophonetic variation in urban centres, our goal – which is mainly determined by the nature of the corpus (see Section 2.1) – is to explore geographical phonetic variation. The concept of an accent region is very deceptive; some caveats must therefore be borne in mind. Following Kortmann and Upton (2004), we must insist that there are no such things as clear accent boundaries; the reality of the isoglosses inherited from sound atlases (e.g. Orton, Sanderson, & Widdowson, 1978) should not be overstated. In addition, not all speakers in one region use the same features; and some do not to the same extent as others. Additionally, depending on linguistic history, education and social class, some speakers may use features from other regions or vary from a typically local pronunciation to supra-local, or more standard variants.

The classic descriptive framework of our study relies on Wells (1982) whose lexical sets have become the gold standard in accent studies in the British Isles. Lexical sets are groups of words that tend to have the same phonological vowel whatever the accent; they are symbolized by key words in small capitals. The acoustic and computational side of this study draws on previous work of our own (Ferragne, 2008, pp. 309–338; Ferragne & Pellegrino, 2007b) and also on multidimensional studies of vowels, especially Heeringa (2004), and Huckvale (2004, 2007).

Huckvale (2004) carried out accent classification with a subset of the corpus used in the current study. He reported classification experiments involving formants and a spectral envelope metric based on a 19-channel auditory filterbank. Huckvale built on the approach initiated in Barry, Hoequist, and Nolan (1989) whereby accent-specific phonetic knowledge is more explicitly included in the classifier by computing distances between vowels rather than using individual absolute vowel coordinates. He introduced the use of matrix correlation as a reliable and gender-insensitive means of estimating the distance between the vowel distance matrices of two speakers. With this method, Huckvale (2004) achieved up to 89.4% correct classification.

In addition to automatic classification, Huckvale (2004, 2007) performed cluster analysis and proved that the method involving vowel distance matrices could be useful to highlight how accent groups relate to one another. A similar approach was explored in Ferragne and Pellegrino (2007b) and Ferragne (2008): several classifiers (Huckvale's method, artificial neural networks, linear discriminant analysis, and Gaussian mixtures) were used to sort 264 speakers into 13 accent groups. Correct classification rates reached as high as 94%, and preliminary results on both accent and vowel clustering were reported.

1.3. *Goals*

The goal of this article is to illustrate how feature spaces used in speech technology can be adapted for the description of vowels in 13 English accents of the British Isles. This novel methodology is not only interesting per se—as it includes valuable features such as speaker normalization, but it is also an alternative to the error-prone formant approach. We show, using a distance metric between vowels that had originally been designed for automatic accent classification (Ferragne, 2008, pp. 309–338; Ferragne & Pellegrino, 2007b; the approach builds on Barry et al., 1989; Huckvale, 2004), that a high-dimension feature space can easily be reduced to a more manageable, phonetically interpretable space.

The vowel systems of the 13 accents are represented with tree diagrams (dendrograms) and MDS plots: although the information of openness and backness is not explicitly mapped onto the representation, the diagrams allow the quick spotting of short distances and, in some cases, they display an interesting picture of the overall structure of some systems.

Then, the distance between individual vowel spaces is used to compute distances between accents. Hierarchical clustering and MDS are applied again in order to obtain a graphic representation of acoustic distances between accents. The interpretation of the resulting graphs leads to a taxonomy which is consistent with classical studies and textbooks on the accents of the British Isles (Hughes, Trudgill, & Watt, 2005; Wells, 1982).

2. Materials and methods

2.1. *Corpus*

The data come from the Accents of the British Isles (ABI) corpus, which is a commercially available collection of recorded read materials covering 14 areas in the British Isles (D'Arcy, Russell, Browning, & Tomlinson, 2004). Recordings were carried out with a close-talking microphone and the signal was digitized with 22,050 Hz sampling rate and a 16 bit quantization. On account of lack of homogeneity, one of the original accents (the Inner London subset) was not included in the analysis after auditory assessment by a British phonetician. The remaining sample therefore consists of 13 accents, each represented by 10 male and 10 female speakers on average, for a total of 261 speakers. Table 1 shows the abbreviations that will be used to designate the accents and Fig. 1 shows a map of the British Isles. A list of 11 /hVd/ words was read five times by the participants: *heed, hid, head, had, hard, hod, hoard, hood, who'd, Hudd, heard*. We will not delve into the advantages and drawbacks of such word

Table 1
Accents of the ABI corpus and corresponding abbreviations.

Abbreviation	Accent
brm	Birmingham
crn	Cornwall
ean	East Anglia
eyk	East Yorkshire
gla	Glasgow
lan	Lancashire
lvp	Liverpool
ncl	Newcastle
nwa	North Wales
roi	Republic of Ireland
shl	Scottish Highlands
sse	Standard Southern English
uls	Ulster

Fig. 1. Accents of the British Isles.

lists, suffice it to say that they neutralize information-related phonetic variation (e.g. varying degrees of vowel-target achievement due to predictability of occurrence or to the lexical vs grammatical status of the word; Aylett & Turk, 2006; Wright, 2003); so, in a way, this is the closest one can get to a controlled experiment. More vowels (19 in total) are available in the corpus, but we chose to concentrate on a restricted set; the reason for this is explained in Section 3.1.

Consonants too play an indexical role in the study of accents, but we concentrated on vowels since the data did not contain a list of stimuli (equivalent to the /hVd/ words) with minimal pairs involving consonants. Consonants are beyond the scope of this study; some diagnostic consonantal features on the same corpus are briefly presented in Ferragne (2008).

2.2. General method

The vowel nuclei were extracted using automatic F0 detection with the Snack Sound Toolkit (Sjölander, 2004). The resulting voiced interval comprises more than just the vowel—it also includes a portion of the closure phase of /d/ and the /r/, whenever it is realized as an approximant in rhotic accents, but this bias was played down by the fact that the acoustic parameters were measured at the temporal mid-point of the duration of the whole voiced interval. The melfcc Matlab routine (Ellis, 2005) was used to compute 12 Mel-Frequency Cepstral Coefficients (MFCCs) at 50% of the duration of the interval. The window length was set to 20 ms, the analysis step, to 10 ms, and the maximum frequency was 8000 Hz. All other options in the melfcc function were those that Ellis (2005) recommends to duplicate the MFCCs obtained with the Hidden Markov Model Toolkit (Young, 1994).

Duration itself, although an important parameter in most vowel systems, was not included on the grounds that its weight relative to the MFCCs was difficult to define a priori for the computation of distances between vowels. Duration will be dealt with in a separate subsection.

For each speaker, the 5 repetitions of each vowel type were averaged: at this stage, a speaker's vowel system was represented by a raw data matrix of size 11 (vowels) $\times$ 12 (parameters). Then, following Barry et al. (1989) and Huckvale (2004), distances were computed between the $11\times(11-1)/2=55$ vowel pairs. Several types of distance metrics were experimented before we eventually opted for the Manhattan distance on the grounds that it yielded the highest correct classification rates in Ferragne and Pellegrino (2007b).

3. Vowel systems

3.1. Method

The goal of Section 3 is to illustrate the vowel systems of the 13 accents. In this section, the 11 vowels that are conventionally labelled monophthongs in descriptions of Standard British English (e.g., Wells, 2008; Jones, 2003) are included in the analysis (recall, as mentioned earlier, that the corpus actually contains more /hVd/ words). We concede that this choice, whose primary motivation was to enhance the legibility of dendrograms and MDS plots by graphing a smaller set of vowels, is not fully satisfactory. By way of example, opinion is divided as to whether the SQUARE vowel (instantiated in the corpus by the word *hared*) in Standard British English should be called a diphthong (Wells, 2008; Jones, 2003) or a monophthong (Olausson & Sangster, 2006; Upton, 2004); so it goes without saying that agreeing on a single subset of the 19 original vowels that could be phonetically matched across accents would be a hard task. It should therefore be remembered that the 11 remaining vowels were chosen for the sake of convenience, rather than on grounded phonetic principles.

For each of the 13 accents, the mean distance matrix over all speakers' 11×11 distance matrices was computed. The resulting 13 matrices were converted to dendrograms using hierarchical clustering with average linkage in Matlab. Average linkage is known to be more robust than other standard methods such as single, complete linkage or Ward's method (Everitt, Landau, & Leese, 2001, pp. 59–64), and it was successfully used by Huckvale (2007) in a study whose principles are very close to our own, although Huckvale concentrated on clustering accents only, not vowels. Incidentally, among standard agglomeration methods, average linkage is also known to yield the highest cophenetic correlation coefficients (Holgersson, 1978, p. 291). The cophenetic correlation coefficient measures the correlation (the closer to 1, the stronger the positive correlation) between the original distance matrix and the corresponding distances as they appear in the tree. Multidimensional scaling (MDS) was also computed from each of the 13 accent distance matrices. The specific, non-metric, algorithm used in this article was requested to produce a two-dimensional configuration of points whose pairwise distances approximate the original dissimilarities between vowels while minimizing a stress criterion defined as the sum of the Euclidean distances between original dissimilarities and output distances, normalized by the sum of squares of the interpoint distances. Goodness-of-fit is better when the stress index is lower.

3.2. Results

The cophenetic correlation coefficients and the stress indices for the MDS are listed in Table 2. The correlation coefficients range from 0.669 (*eyk*) – which is quite low – to 0.824 (*lvp* and *shl*), reflecting discrepancies in how faithful the trees are to the original distances. Statistical significance tests are not normally performed with such correlations because the two matrices are not independent (Legendre, 1998, p. 376). The third column shows the percentage of explained variance, which is the squared correlation coefficient multiplied by 100. The stress values returned by the MDS are shown in column 4. As a rough guideline, a value close to 0.025 generally means that the fit is excellent, while a stress value between 0.10 and 0.20 is considered fair to poor (Izenman, 2008, pp. 500–501).

Table 2
Cophenetic correlation coefficients and stress.

Accent	Cophenetic correlation	% Variance explained	Stress
brm	0.801	64.2	0.086
crn	0.685	46.9	0.094
ean	0.730	53.3	0.066
eyk	0.669	44.8	0.040
gla	0.744	55.4	0.072
lan	0.818	66.9	0.044
lvp	0.824	67.9	0.037
ncl	0.724	52.4	0.059
nwa	0.730	53.3	0.047
roi	0.772	59.6	0.096
shl	0.824	67.9	0.068
sse	0.740	54.8	0.057
uls	0.769	59.1	0.059

The dendrograms and MDS plots are shown from Figs. 2–14. The description of the trees and MDS plots will be complemented with an auditory analysis conducted by the first author. The absolute location of vowels in the phonetic space cannot be retrieved from the distances we have computed. However, we will use prior knowledge from Ferragne and Pellegrino (2010) in the following description: it so happens that while some vowels vary quite a lot across accents (e.g. *who'd*), others tend to exhibit a relatively high degree of stability (e.g. *heed*, *hoard*, and *had*). In Ferragne and Pellegrino (2010) it appears that, whatever the accent, *heed* shows the highest amount of frontness and closeness, *hoard* is a rather stable reference for maximal backness, and *had* can serve as a reference for maximal openness. In the remainder of this section, we address each accent in turn, starting with a short description of what is to be expected on the systemic and acoustic level based on the subsample of speakers analysed in Ferragne and Pellegrino (2010).

3.2.1. Standard Southern British English (sse)

The 11 /hVd/ words in the corpus contain distinct vowel phonemes in *sse*; in other words, no phonemic splits or mergers are expected. On the phonetic level, the most striking feature is perhaps the front position of the *hood* and *who'd* vowels (Ferragne

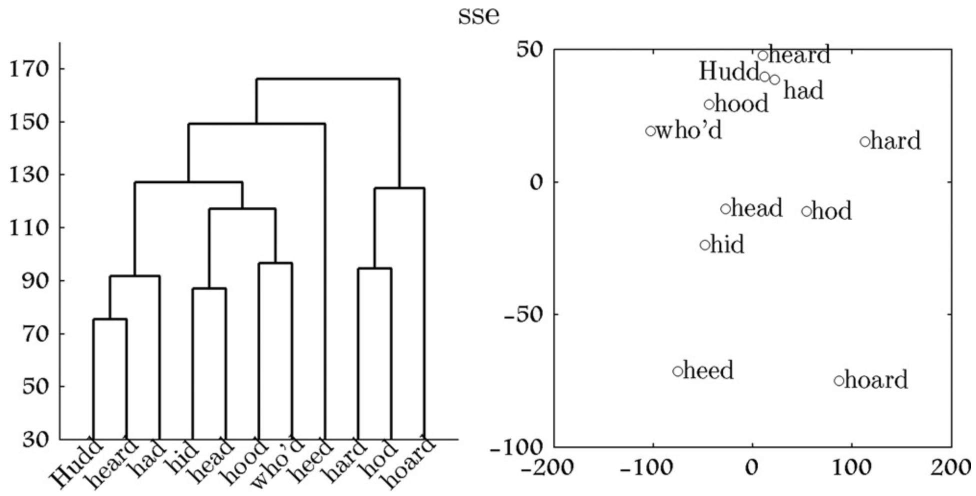

Fig. 2. Dendrogram and MDS plot of the *sse* system.

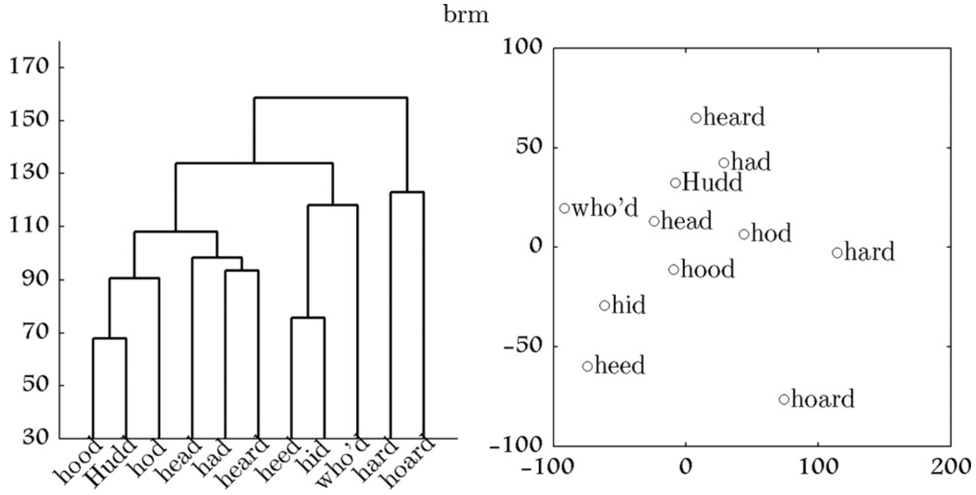

Fig. 3. Dendrogram and MDS plot of the *brm* system.

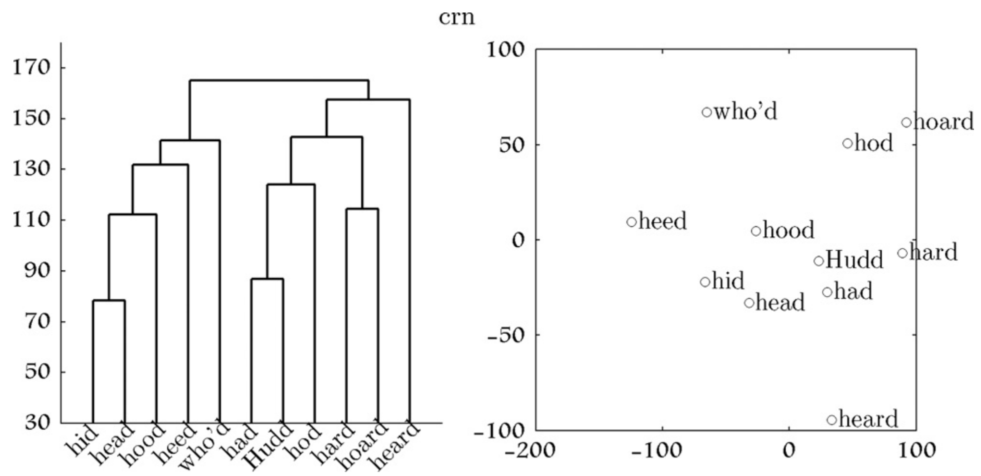

Fig. 4. Dendrogram and MDS plot of the *crn* system.

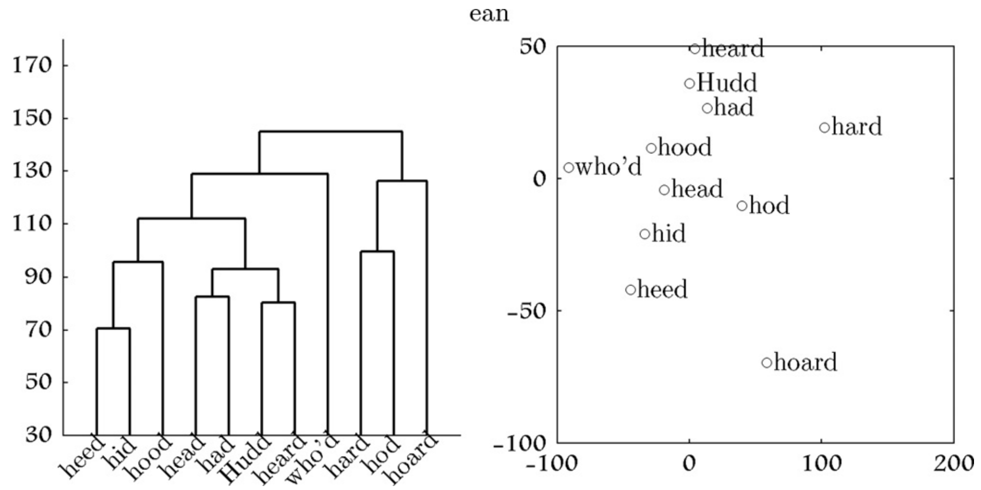

Fig. 5. Dendrogram and MDS plot of the *ean* system.

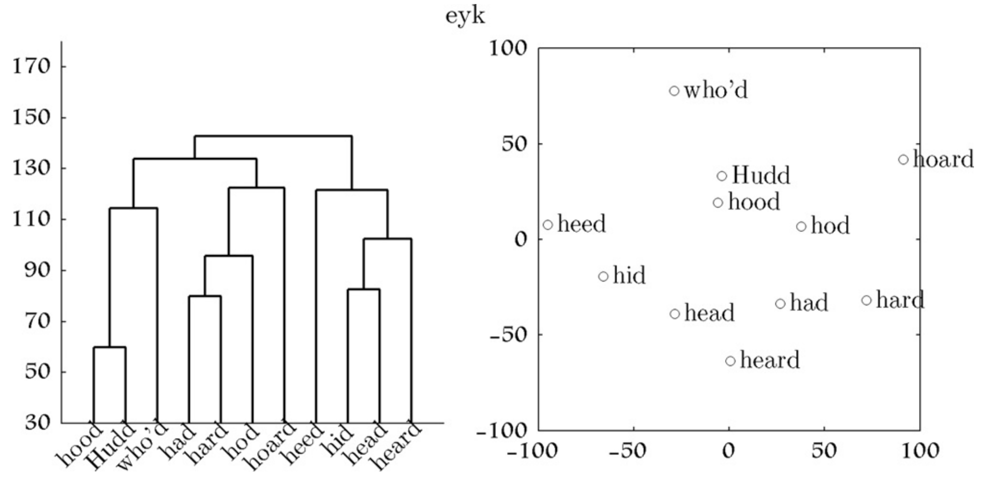

Fig. 6. Dendrogram and MDS plot of the *eyk* system.

& Pellegrino, 2010). Fig. 2 represents the vowel system of *sse*: the overall agreement between the two types of representations seems rather poor. However, short distances (e.g. between *Hudd*/*heard*/*had*, *hid*/*head*, and *hood*/*who'd*) have been preserved in both graphs. A cut-off distance value of about 160 in the dendrogram reveals a first split between back and non-back vowels. Then, around 120, the leftmost cluster is composed of central vowels, and the following one comprises front vowels. If we discard the unexpected behaviour of *heed*, three groups roughly corresponding to (from left to right) central, front, and back vowels seem to

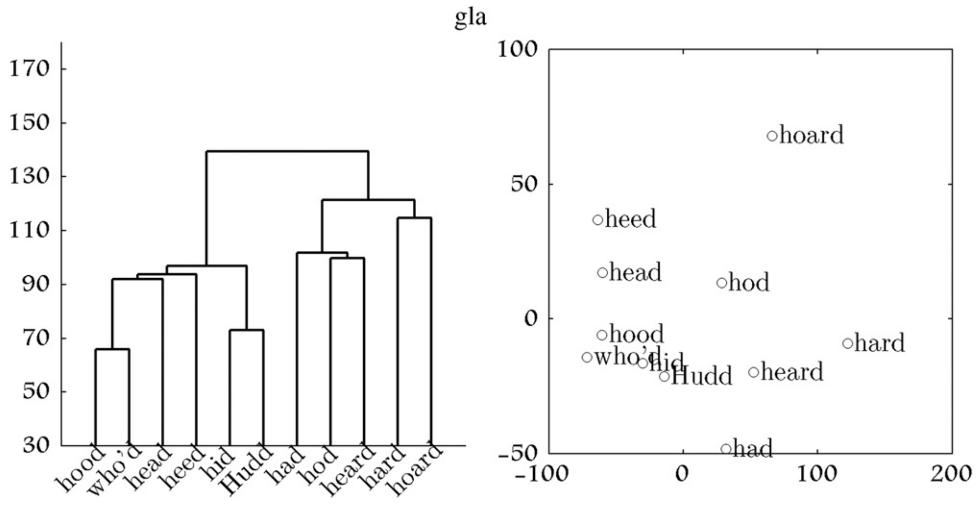

Fig. 7. Dendrogram and MDS plot of the *gla* system.

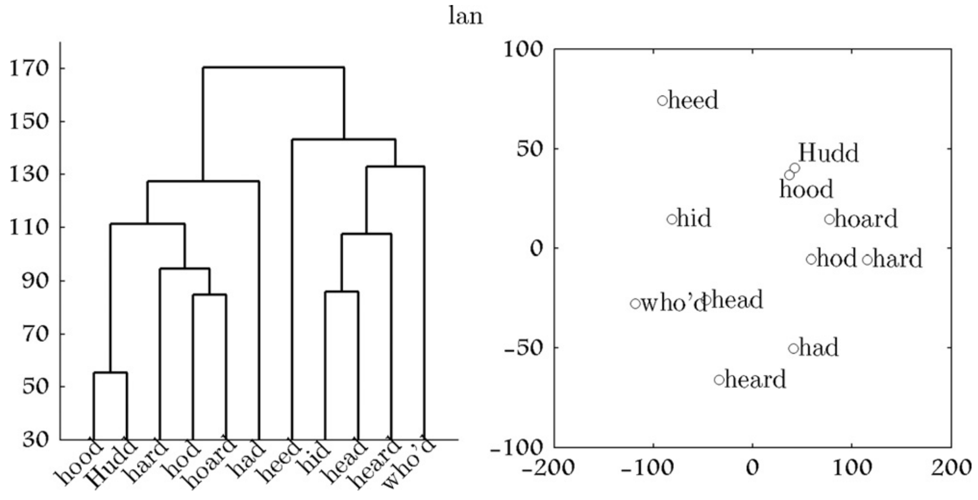

Fig. 8. Dendrogram and MDS plot of the *lan* system.

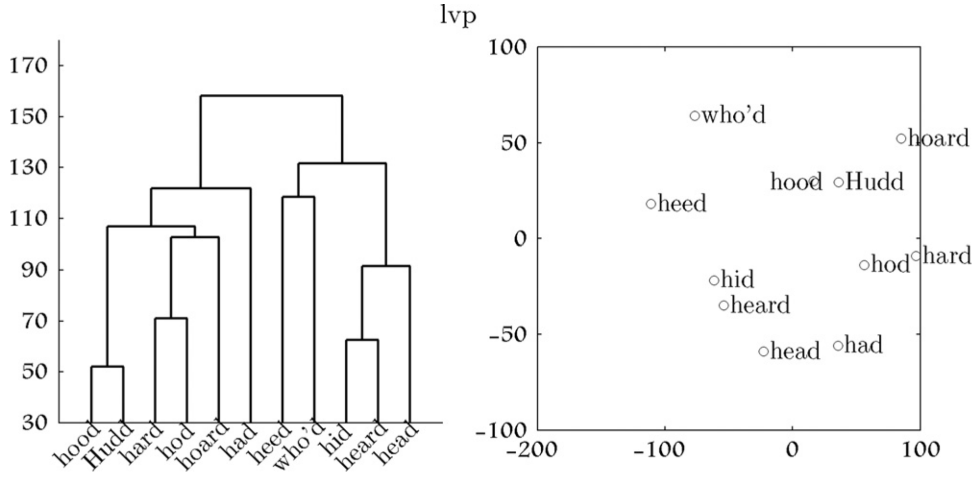

Fig. 9. Dendrogram and MDS plot of the *lvp* system.

emerge. Note however that this structural information does not match what the MDS plot conveys. Then, focusing on front vowels (*hid, head, hood, who'd*) in the tree, they are further split (around 110) between rounded and unrounded vowels. The fact that *hood* and *who'd* cluster with front vowels somewhat contradicts recent pronunciation dictionaries (Jones, 2003; Wells, 2008), but it accords well with formant studies (Ferragne & Pellegrino, 2010; Hawkins & Midgley, 2005; McDougall & Nolan, 2007). Both the tree and the MDS plot suggest that *had* patterns with central vowels. This is not surprising in the light of what Wells noted

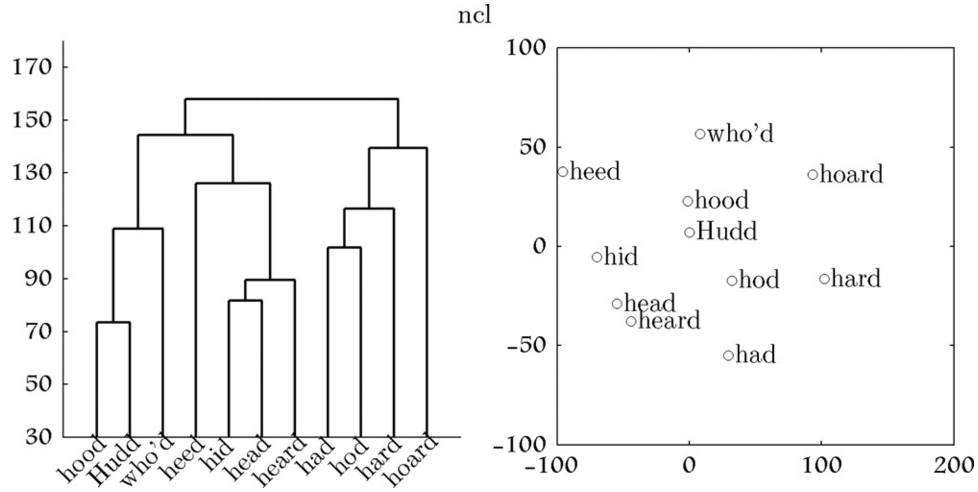

Fig. 10. Dendrogram and MDS plot of the *ncl* system.

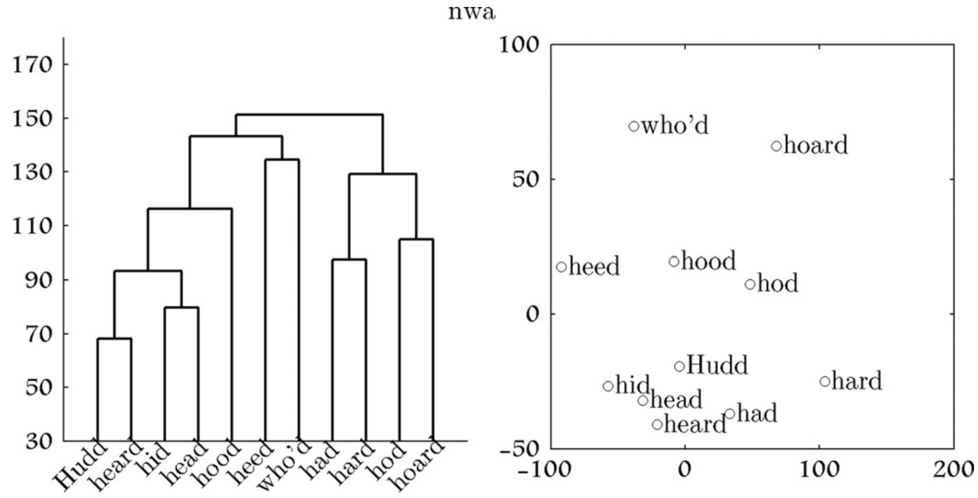

Fig. 11. Dendrogram and MDS plot of the *nwa* system.

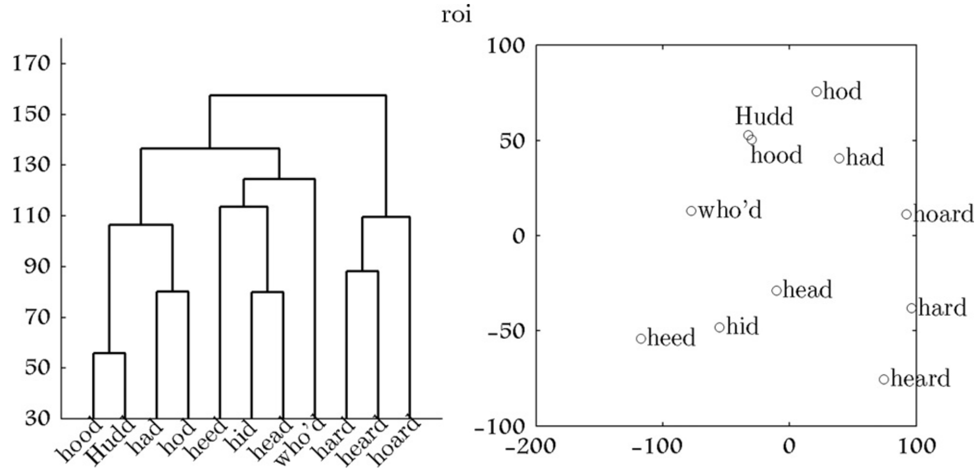

Fig. 12. Dendrogram and MDS plot of the *roi* system.

(Wells, 1982, pp. 291–292): the downward movement of the TRAP vowel leads to a potential merger with STRUT in some speakers. Although we have not found auditory evidence for such a merger, it remains true that the TRAP vowel in our *sse* sample is probably best described as a rather central vowel; the clustering of *had* with *heard* and *Hudd* is even more visible in the MDS plot.

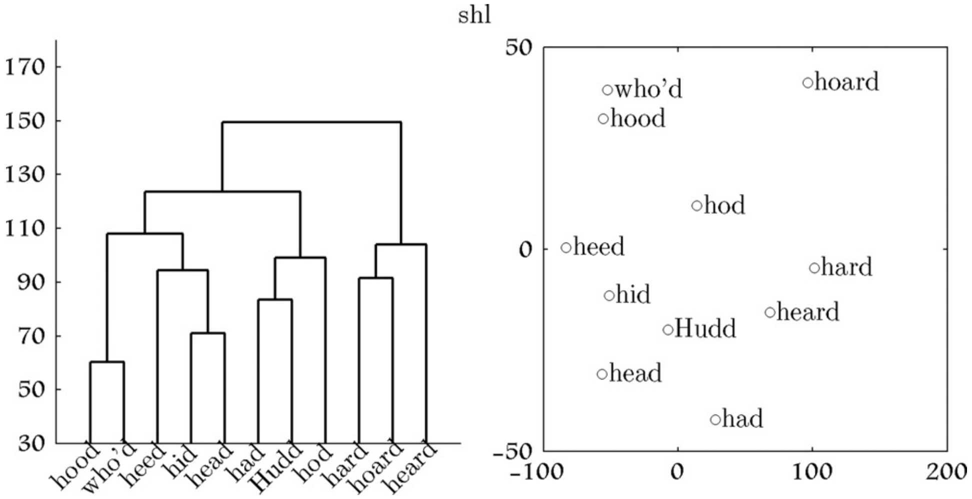

Fig. 13. Dendrogram and MDS plot of the *shl* system.

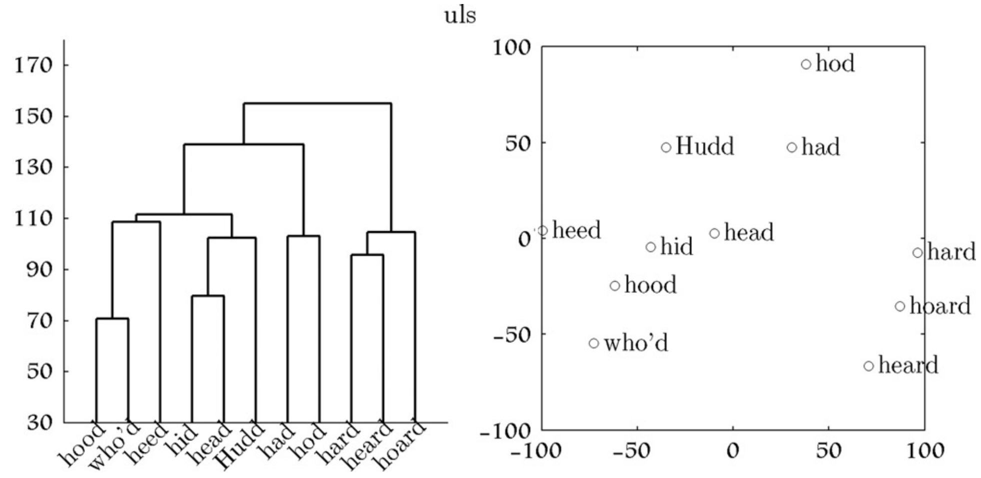

Fig. 14. Dendrogram and MDS plot of the *uls* system.

3.2.2. Birmingham (brm)

As a linguistically northern English accent, *brm* is expected to exhibit the same vowel in *Hudd* and *hood*. Phonetically, a rather close realization of the KIT vowel is expected in typical *brm* (O'Connor, 1973, p. 155), and it was confirmed in Ferragne and Pellegrino (2010). Fig. 3 shows the vowel system of *brm*. The vowel pair with the shortest distance is composed of the words *hood* and *Hudd* in the tree, while the MDS suggests otherwise. As a matter of fact, no strong similarity between two vowels seems to emerge from the latter representation (contrary to Fig. 2 where *had* and *Hudd* showed strong similarity). Regarding the KIT vowel, *heed* and *hid* are very close to each other in the tree from Fig. 3, though not in the MDS plot.

In terms of classic accent typology (Beal, 2004; Wells, 1982), the lack of FOOT–STRUT split is probably the most cited criterion that distinguishes the accents of the north from those of the south of England. So, according to the tree in Fig. 3, we would conclude that *brm* patterns with northern accents. Actually, listening to the test words reveals that only half of the 20 *brm* subjects preserve the homophony between *hood* and *Hudd*, so using averages obviously blurs individual variation. However, compared to the formant study we have carried out elsewhere (Ferragne & Pellegrino, 2010), the tree representation is apparently just as informative as the F1/F2 plots for the *hood*/*Hudd* pair. A final point regarding the tree in Fig. 3 is *who'd* patterning with front vowels, which, as in *sse*, suggests that the GOOSE vowel is better described as a fairly front, close vowel, rather than a back vowel. This point is highly supported by the F1/F2 plot of the subsample used in Ferragne and Pellegrino (2010), quite visible in Fig. 3 in the dendrogram, but not really inferable from the MDS plot.

3.2.3. Cornwall (crn)

From a systemic point of view, we can anticipate that the 11 test words will have distinct phonemes in *crn*. On the realizational level, the insufficient data (two speakers) in Ferragne and Pellegrino (2010) do not allow us to predict any particular feature with certainty. In Fig. 4, the vowels of *crn* are represented. Both graphs agree that the shortest distances are found between *hid* and *head*, and *had* and *Hudd*. Although the cepstral distance between *hid* and *head* is relatively small, two perceptually distinct vowels can be heard. This confirms that our method is really a preliminary tool intended to draw the researcher's attention to potential interesting phenomena, not a totally self-contained method. As was the case for *sse* and *brm*, *who'd* patterns with front vowels, and so does *hood* according to the dendrogram. In terms of system, given the set of vowels we are looking at, *crn* is similar to *sse*.

3.2.4. East Anglia (ean)

In terms of system, *ean* is expected to be very similar to *sse* (Ferragne & Pellegrino, 2007a). The vowels of *hood* and *who'd* should pattern with front vowels. Fig. 5 shows the vowel system of *ean*. As was the case for *heed* in *sse*, the *who'd* vowel here exhibits a somewhat unpredicted pattern in the dendrogram. The vowels of *hood* and *who'd* tend to cluster with front vowels. The vowel in *had* patterns with central vowels. It is surprising to see that *heed* and *hid* are separated by such a short distance. A possible explanation would be that, given that the FLEECE (i.e. *heed*) vowel is known to be rather diphthongized in *ean*, the acoustic measurement at temporal mid-point captures a vowel quality that is slightly more central than the targeted [i].

3.2.5. East Yorkshire (eyk)

It is expected that the vowel system of *eyk* will lack the FOOT–STRUT split. From Ferragne and Pellegrino (2010), the spectral proximity between *had* and *hard* constitutes a remarkable phonetic phenomenon. The vowel system of *eyk* (Fig. 6) exhibits a remarkable northern structure. A cut-off distance of about 130 splits the tree into three clusters roughly corresponding to (from left to right) back vowels, open vowels, and front close vowels. The remarkably short distance between *hood* and *Hudd* – which is also observable in the MDS plot – obviously illustrates the absence of FOOT–STRUT split. The fact that *heard* patterns with front vowels can be accounted for by the fact that many speakers seem to distinguish between *heard* and *head* by relying on duration only (median duration: *head*: 166 ms; *heard*: 281 ms). The same remark applies to the *had*/*hard* pair (median duration: 175 and 281 ms, respectively).

3.2.6. Glasgow (gla)

The *gla* vowel system is expected to exhibit the FOOT–GOOSE merger (Stuart-Smith, 2004; Wells, 1982). On the phonetic level, the realization of the KIT vowel is markedly different from the other accents. Fig. 7 illustrates the *gla* system. The vowels of *hood* and *who'd* cluster with front vowels and the proximity between the two seems to confirm the existence of the FOOT–GOOSE merger in Glasgow English Actually there is a perceptibly evident difference in duration between the two vowels (approx. 80 ms in median duration); which indicates that, had duration been included (with a perceptually motivated weight) in the computation of the distance, the two vowels would have been further apart. Another important feature in Fig. 7 is *hid* patterning with *Hudd*. This is perfectly consistent with our auditory impression. This striking realizational characteristic is well-documented (Eremeeva & Stuart-Smith, 2003; Stuart-Smith, 1999, 2004), and it has been shown to correlate with social class: in a nutshell, lower and more retracted variants are used by lower-class speakers.

3.2.7. Lancashire (lan)

Systemically, the *lan* vowel system should have the same vowel in *Hudd* and *hood*. Our sample is almost exlusively non-rhotic. The system of *lan* is illustrated in Fig. 8. Starting from the root, the first split separates front vowels from the rest. The lack of FOOT–STRUT split appears clearly. Contrary to *eyk* – which probably has the vowel system that is closest to that of *lan* in the corpus – the *lan* realization of the *who'd* vowel is rather front.

3.2.8. Liverpool (lvp)

Potential systemic peculiarities in *lvp* are the NURSE–SQUARE merger and the lack of FOOT–STRUT split. The subset of /hVd/ words unfortunately does not allow us to test the former. Fig. 9 shows the *lvp* vowel system. As expected from a northern English accent, both the tree and the MDS diagram suggest that FOOT and STRUT constitute one single phoneme. The next shortest distance in both graphs is between *hid* and *heard*. The auditory analysis confirms that *heard* is realized as a long monophthong whose quality is close to that of *hid*. The auditory analysis confirms that *heard* is realized as a front vowel.

3.2.9. Newcastle (ncl)

As a typically northern English accent, *ncl* should not normally have distinct phonemes for *Hudd* and *hood*. The typical vowel in *heard* should show a high degree of frontness (Watt & Allen, 2003). The vowel system of *ncl* is illustrated in Fig. 10. The shortest distance in the tree diagram (i.e. between *hood* and *Hudd*) seems to support the lack of FOOT–STRUT split. It is slightly greater than the distance found between these two vowels in other accents lacking the distinction; for instance, the tree distance between *hood* and *Hudd* for *eyk*, *lan*, and *lvp* is closer to 50 than 70. One possible reason is that some speakers phonologically interpreted the vowel in *hood* as being the same as that in *who'd* (Ferragne & Pellegrino, 2010). The MDS plot implies that the smallest dissimilarity occurs between *heard* and *head*, which makes sense because, *head* being a fairly stable reference across accents for frontness, the *heard* vowel is clearly front in most speakers. However, the fact that *heard* and *head* are so close to each other on the MDS plot not only conceals the fact that duration is a totally reliable cue to distinguish them, but also falls short of signalling that there is a perceptible difference in rounding—the vowel in *heard* being rounded.

3.2.10. North Wales (nwa)

In the *nwa* vowel system, the 11 test words are assumed to contain distinct phonemes. On the phonetic level, we expect rather front qualities for *heard* and *Hudd* and a back quality in *who'd*. Fig. 11 shows the *nwa* vowel system. The similarity between *Hudd*, *heard*, and front vowels is attested in both graphs. The vowels in *Hudd* and *heard* are very close to each other in the dendrogram (less so in the MDS plot) and their distance—slightly less than 70, seems typical of an absence of spectral difference. In spite of this, the vowels can be easily distinguished by ear (median duration: *Hudd*: 150 ms; *heard*: 278 ms) The distance between *had* and *hard* actually conceals the fact that, in some speakers, only duration seems to be used to tell one from the other; in other words, had all speakers behaved like the latter, the distance between *had* and *hard* on the graph would have been smaller.

3.2.11. Republic of Ireland (roi)

On the systemic level, Dublin English may have the NURSE–SQUARE merger which, for want of adequate test words, could not be analysed here. From Ferragne and Pellegrino (2010), we know that our *roi* sample quite unexpectedly has the same phoneme in *hood* and *Hudd*. Phonetically, rhoticity is expected to influence the measurements, especially in *heard* since the latter has a strong tendency to be r-coloured throughout (Ferragne, 2008). The vowel system of *roi* in Fig. 12 shows that *hood* and *Hudd* are very close to each other. Auditorily, the lack of FOOT–STRUT split is confirmed: this is, according to Hickey (2004), not typical of Dublin English in general. Although the distance between them is quite substantial, it is worth noticing that the three words containing a graphic ⟨r⟩ tend to cluster together, which, perhaps reflects their similarity in terms of rhoticity.

3.2.12. Scottish Highlands (shl)

The *shl* system is expected to show the FOOT–GOOSE merger and optionally, in terms of realization, some retraction for the KIT vowel. We anticipate that rhoticity may affect the measurements;

Ferragne and Pellegrino (2010) showed that – as opposed to *gla* – rhoticity is maintained here by all speakers. In Fig. 13 the vowel system of *shl* is illustrated. The shortest distance (both in the tree and the MDS plot) can be found between *hood* and *who'd*, which supports the FOOT–GOOSE merger (see however Section 3.3). Notice that, contrary to *gla*, the vowel of *hid* does not cluster with central vowels, which is borne out by our perceptual impression. Rhoticity being the norm, it follows that the cluster containing *hard, hoard, and heard*, may have arisen as a consequence of this phenomenon.

3.2.13. *Ulster (uls)*

The accent of Ulster (*uls*) is expected to show a vowel system very close to that of Scottish English in that *hood* and *who'd* should have the same phoneme. On the phonetic level, these two vowels are quite front, and rhoticity may affect the quality of the vowel in *hard*, *hoard*, and *heard*. The vowel system of *uls* is shown in Fig. 14. There is rather good agreement between the tree and the MDS plot. According to the dendrogram, the shortest distance can be found between *hood* and *who'd*. This is not exactly the case in the MDS plot although the distance between *hood* and *who'd* is among the shortest. Both graphs in Fig. 14 show that *hood* and *who'd* pattern with front vowels as anticipated. The separate group composed of *hard*, *hoard*, and *heard* may arise from their proximity in terms of rhoticity. It must be borne in mind that rhoticity sometimes surfaces as an r-colouring spanning the whole vowel. Therefore, in such cases, our attempt to play down the impact of rhoticity by making acoustic measurements at temporal mid-point is ineffective. Now, back to the FOOT–GOOSE merger, Ferragne and Pellegrino (2010) tend to suggest that the spectral quality of these two vowels is perceptually equivalent (see however Section 3.3).

3.3. *Duration*

As mentioned above, duration was not included in the computation of distances because estimating its a priori weight, relative to cepstral coefficients, was problematic. In spite of this, it was thought that an exploratory analysis of the duration in pairs of vowels separated by a small acoustic distance could constitute an interesting complement to our study. For each accent, only the vowel pairs containing the two vowels separated by the shortest distance in the matrix are treated below.

Kernel smoothing density estimates (Everitt et al., 2001, pp. 16–20) were computed from the raw duration of the restricted set of vowels under study. The difference between the two duration distribution estimates in each accent is expressed in Fig. 15 as the Jensen–Shannon divergence (multiplied by10^4 for the sake of legibility), which is a symmetric version of the Kullback–Leibler (Kullback, 1968) distance used to gauge the distance between two probability distributions. A rapid look at Fig. 15 shows that four accents (*ean*, *gla*, *nwa*, *sse*) exhibit a strong divergence between the two vowels, which very likely reflects a reliable distinction in duration. Conversely, the other 9 accents probably have no robust durational differences between the analysed vowels.

We are faced with three situations: (1) the case of FOOT–STRUT, where the expected lack of phonemic split is strongly supported, (2) the case of FOOT–GOOSE, where duration raises doubts about an expected phonemic merger, and (3) the case of acoustically very close vowels which, given their great distance in terms of duration, very likely constitute two separate phonemes. There is also a fourth, quite unexpected, case – namely the vowels of *hid* and *head* in *crn* – in which there was no reason to expected that the distance (acoustic and durational) between them would match the distance generally observed between two occurrences of the same phoneme.

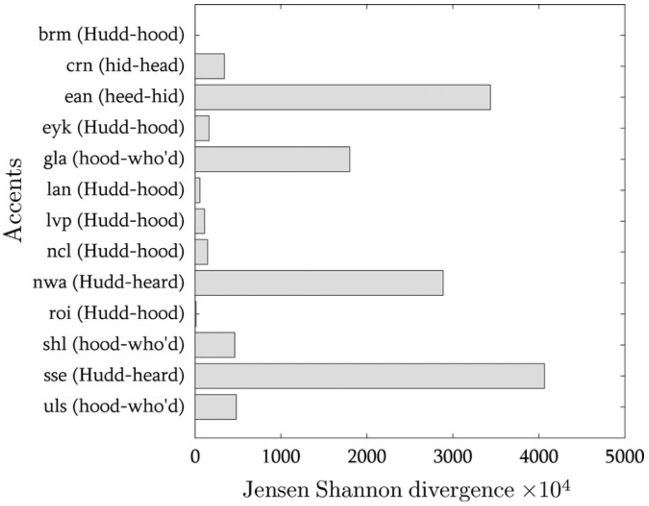

Fig. 15. Jensen–Shannon divergence between the probability densities of the duration of the closest pair of vowels in each accent.

As far as *hood* and *Hudd* are concerned (*brm*, *eyk*, *lan*, *lvp*, *ncl*, and *roi*), the low divergence in Fig. 15 and a careful inspection of the density estimates (there is an almost perfect overlap in each accent) suggest that duration could in any case not constitute a reliable cue to distinguish the two vowels. So the general picture tends to confirm that in the accents we have just mentioned, *Hudd* and *hood* (i.e. FOOT and STRUT) are one single phoneme, whose expected realization is, incidentally, a short rather back vowel. A finer-grained analysis (see Ferragne, 2008, pp. 256–258) would actually reveal that, at least in *ncl*, one speaker has a long vowel in *hood* because, as is sometimes the case in the north of England, some words generally thought to belong to the FOOT lexical set pattern with GOOSE words.

Regarding *hood* and *who'd* in the accents *gla*, *shl*, and *uls*, the divergences are slightly (*shl* and *uls*) or much (*gla*) greater than those observed between the duration of *Hudd* and *hood* in the previous paragraph, which was then interpreted as evidence that there was no difference in duration. The direction of the divergence is consistent across accents: the *who'd* vowel tends to be longer than the *hood* vowel. While it would be far-fetched to reach definite conclusions about *shl* and *uls*, the substantial divergence in *gla* warrants closer inspection. It so happens that, as a result of the Scottish Vowel Length Rule, some vowels followed by the suffix [d] are appreciably longer than their counterparts followed by a tautomorphemic [d]. According to Scobbie, Hewlett, and Turk (1999), /u/ is a good candidate for this phenomenon, and our results support this claim strongly for our *gla* sample, and also, perhaps – though to a much lesser extent – for *shl* and *uls*.

Finally, the biggest divergences in Fig. 15 (*ean*, *nwa*, and *sse*) exemplify pairs of vowels whose acoustic proximity is mainly disambiguated thanks to duration. In Fig. 15, we did not expect any difference in duration between *hid* and *head* in *crn* because there is absolutely no indication in the literature (Wells, 1982) that one of these two vowels should be longer than the other. So, the spectral similarity between *hid* and *head* in *crn*, as it is suggested by the dendrogram, is not counterbalanced by a difference in duration.

3.4. *Discussion*

This section shows how the type of high-dimension parameter spaces used in speech technology can be adapted to the needs of phoneticians. The main drawback of our method is that the traditional dimensions of backness and openness are no longer accessible. However, as already noted, the procedure is completely automatic; it therefore saves time and yields reproducible results.

Contrary to the formant method, cepstral analysis has, as far as we know, never been interpreted within a phonetic framework, so, it lacks predefined expected values for a specific vowel. Therefore, although it may be subject to extraneous factors (individual variation, background noise, etc.), one cannot conclude that it is wrong, because there is no phonetic theory against which it can be compared. Granted that distances in the MFCC space achieve good phonetic interpretability, we feel that the argument that MFCCs cannot be wrong (while formants can) provides strong support for the use of MFCCs in phonetic studies, if only for practical reasons.

Our procedure is still subject to improvement, especially regarding the mapping from the distance matrix to the final visual output. This aspect requires further investigation because different techniques yield very different results, as evidenced by the comparison between dendrograms and MDS plots. If we compare the shortest pairwise distance per accent from the matrix (the vowel pairs in question are shown Fig. 15) with those computed with hierarchical clustering or MDS, it appears that the tree diagrams in Section 3 are more faithful to shortest dissimilarities. Depending on the MDS technique, more or less weight will be given to the accuracy of small or large distances (Izenman, 2008); thus, a genuine benchmark of the various MDS algorithms available should be carried out to compensate the shortcomings in the current article.

Another possible improvement yet concerns the computation of distances, although, based on the high classification scores obtained in Ferragne and Pellegrino (2007b), we can be quite confident that the MFCC distances really capture accent-specific vowel features. But the question of weighting the original features deserves further investigation, especially with respect to the possible inclusion of duration. The Manhattan distance is sensitive to differences in units in the raw data, which means that the variables measured on scales involving greater numbers are (implicitly) given more weight; hence the exclusion of duration in this study. It would however be desirable to compute distances with an appropriate weighting for all features, including acoustic changes within a vowel.

Finally, it must be pointed out – but this is not specific to hierarchical clustering or MDS – that average values often hide relevant individual variation. Within-accent differences have been highlighted in Ferragne and Pellegrino (2010): for example, some *brm* speakers have two phonemes for FOOT and STRUT, not all speakers of *gla* are rhotic, the realization of GOOSE can vary from a diphthong to a monophthong (back or front) in many accents, etc. So, ideally, dendrograms representing individual systems should be obtained before averaging the data.

To be fair, the technique presented here is a complement to auditory analysis rather than a self-contained tool. However, it illustrates how methods from biostatistics can be used by phoneticians.

4. Acoustic distance between accents

4.1. Method

In Ferragne and Pellegrino (2007b) and Ferragne (2008), we carried out experiments in automatic accent classification in which up to 94% of the speakers in the corpus were correctly classified. The method for obtaining individual vowel spaces was very close to the one presented here (the original idea comes from Barry et al., 1989), and the classification procedure involved the comparison of vowel spaces with a correlation coefficient (following Huckvale, 2004, 2007). So, given the high classification rates achieved with this procedure, it can be inferred that the correlation between two individual distance matrices (see Section 2.2) adequately reflects their proximity. Therefore, for each accent, mean vocalic distance matrices were computed; and distances between pairs of accents were expressed as 1 minus a correlation coefficient which was used to build a hierarchical clustering tree, again with average linkage (Fig. 16). Simultaneously, the distances between the original 261 vowel matrices were computed (as correlations) yielding a proximity matrix between the 261 speakers. The matrix was further converted to a dissimilarity matrix and submitted to non-metric MDS. In Fig. 17, for each accent, mean MDS coordinates were used to express accent centroids, and the 95% confidence interval of the mean represent the radius.

4.2. Results

In Fig. 16, going from the root to the leaves, the first split separates Scottish and Irish varieties (*gla*, *shl*, *roi*, and *uls*) from the rest. This group of 4 accents further shows a division between the 2 Irish and the 2 Scottish accents. As for the remaining accents, *lan* and *lvp* constitute a separate group. Then, the following split draws a distinction between linguistically southern (*sse*, *brm*, and *ean*) and northern (*eyk*, *ncl*, and *nwa*) accents: this rough bipartition may sound inaccurate for most dialectologist; that is why we will return to it shortly (Section 4.3). The MDS plot (Fig. 17) agrees with the dendrogram in that it highlights the separate group composed of *lan* and *lvp*. It also accurately reproduces the short distances between *eyk*, *ncl*, and *nwa*. The Scottish and Irish varieties no longer pattern together: while a cluster comprising *gla*, *shl*, and *uls* seems to emerge, *roi*, contrary to Fig. 16, does not pattern with them.

4.3. Discussion

The general picture obtained with hierarchical clustering and MDS mostly agrees with previous auditory descriptions of the accents of the British Isles (Hughes et al., 2005; Wells, 1982). The north–south partition in England is duly highlighted; it is a well-known fact, even to the layman, that economic and cultural differences between the north and the south of England are paralleled by salient differences in pronunciation (Wales, 1999).

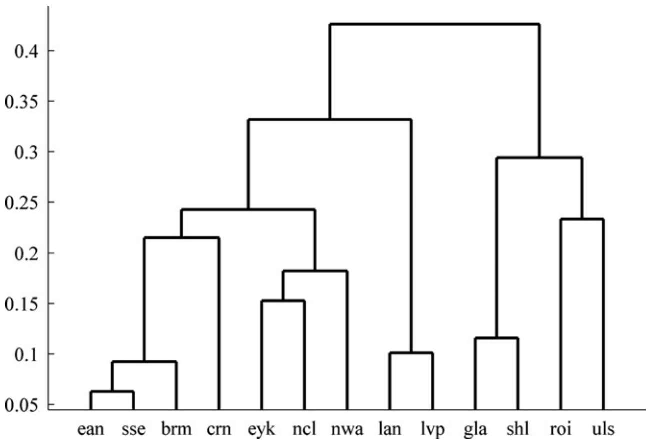

Fig. 16. Dendrogram of the distances between accents.

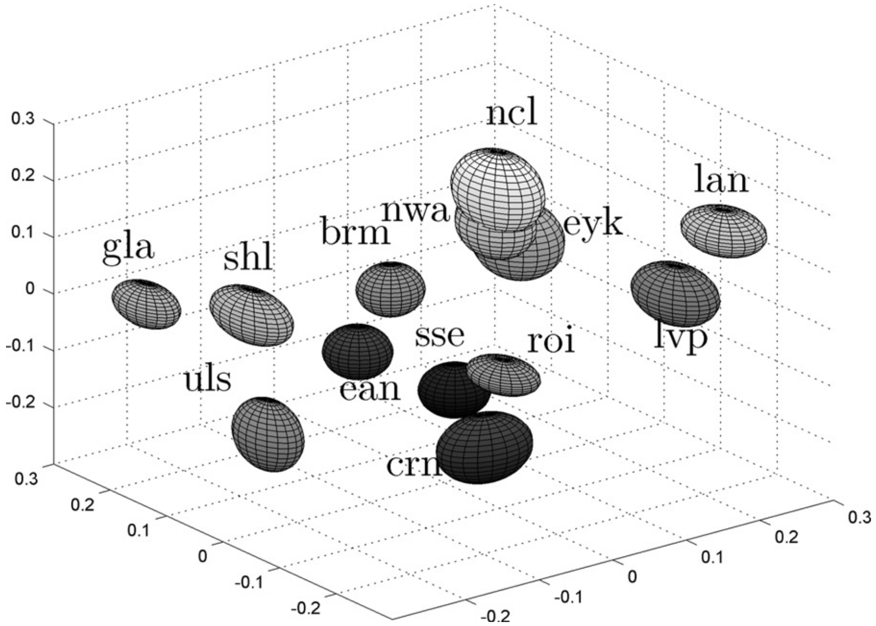

Fig. 17. MDS plot of the distances between accents.

Trudgill (1990, p. 51) even goes so far as to claim that the lack of FOOT–STRUT split, as a typical feature of Northern accents, is known to "everybody who has spent any time in England". The fact that *brm* patterns with southern accents in the dendrogram (while it is generally thought to be rather northern) can be readily explained by our auditory analysis: only half of the speakers in the sample have one single phoneme for FOOT and STRUT words (which is a typically northern trait) while the remaining half have two separate phonemes (typically southern). Note however that the position of *brm* in Fig. 17 is more ambiguous than in Fig. 16. It is not surprising to see *nwa* pattern with the accents of the north of England; as Penhallurick (2004, p. 103) explains, North Wales is geographically close to (linguistically) northern English accents. However, although Penhallurick mentions the possibility of FOOT rhyming with STRUT in North Wales, we have found no such systemic phenomenon that could account for the similarity between *nwa* and northern accents. This proximity is probably best accounted for by similarities on the phonetic, realizational level. For instance, the proximity between *had* and *hard* found in *nwa* is also typical of the *eyk* sample, the rather front quality of *heard* is attested in *brm*, *eyk*, *lvp*, and *ncl*. While *roi* and *uls* cluster together in Fig. 16, our previous work (Ferragne, 2008), which involved more vowel types, showed that *uls* was closer to *gla* and *shl* (than it was to *roi*). The grouping of *uls*, *gla*, and *shl* in Ferragne (2008) was more consistent with the literature (Hughes et al., 2005; Wells, 1982) because – among other possible reasons – there was an influx of Scots settlers in Ulster in the 17th century (Hickey, 2004). Our current result, with Irish accents on one side and Scottish accents on the other, probably arises because fewer vowel types were analysed. Here again the MDS plot in Fig. 17 seems to contradict Fig. 16 since *roi* appears near the accents of the south of England.

The study by Huckvale (2007) is directly comparable to our own. With a similar method, he computed distances between vowels from 20 sentences of the ABI corpus, and applied hierarchical clustering in order to combine speakers into groups. His results agree quite well with those of the present study: the most prominent partition in Huckvale (2007) separates the Scots zone (*gla*, *shl*, *uls*) from the rest. Then the accents of England are roughly split into northern (including *nwa*) and southern (including *brm*). In a further subdivision of the southern group, *brm* is isolated from the others, and in a further subdivision of the northern subset, a subgroup mainly composed of *nwa* speakers appears.

A potential follow-up study would consist in comparing the acoustic distances with similarity judgements by listeners. For example, in a perceptual experiment involving 15 dialects of Norwegian, Heeringa, Johnson, and Gooskens (2008) asked 15 groups of listeners (from each dialect) to judge the perceived distance between their own variety and the remaining 14. Cluster analysis and multidimensional scaling were then performed. The authors found good agreement between the traditional classification of Norwegian dialects, the classification based on perceptual distances, and a classification operated from formant measurements. As far as the English-speaking world is concerned, the perception of American English accents has been thoroughly investigated by Clopper and Pisoni (2004). Their listeners had to carry out a six-alternative forced-choice categorization task; and the resulting confusion matrix was used to compute clustering. Although the clustering solution seems consistent with prior expectations, it must be noted that the categorization was performed with only 30% accuracy. But transforming the confusion matrix of a categorization task into distances is a good alternative to directly asking subjects to estimate a perceptual similarity between two accents. Both methods would however raise thorny issues. For example, it is not a simple matter to put one's finger on what phonetic/acoustic cues listeners attend to when they identify accents. Listening experiments involving whole sentences would be inadequate in our case because we have only measured acoustic cues related to vowels; and consonants and suprasegmental features are also important accent markers. Clopper and Pisoni (2004) tried to find out which acoustic properties could predict the listeners' categorization by using stepwise multiple regression, i.e. the acoustic variables were included one after the other as independent variables in order to assess to what extent they accounted for the resulting categorization. It is true that this procedure can be run on a small set of potential predictors, but a more complete survey would necessitate that all potential phonetic predictors are identified and that their perceptually relevant acoustic correlates are adequately captured. The answer to whether this method is applicable to our corpus highly depends on listeners' aptitude to identify as many as 13 separate accents. Our work on automatic accent identification (Ferragne, 2008; Ferragne & Pellegrino, 2007b) suggests that acoustic parameters make it possible to recover the 13 original classes, but can listeners do the same? Besides, as Daniels (1990, p. 27) observes: "Clearly, not all native speakers of English are interested in regional accents, so that the attempt to identify consciously a speaker's regional accent is not the everyday objective of a listener."

Another possible improvement would be to include consonants. In Section 2.1 we mentioned the fact that the corpus used in this study did not contain adequate test words eliciting consonantal accent markers, which is the main reason why consonants were not included. However, other reasons can be put forward. In Hughes et al. (2005), the accents of the British Isles were split into 16 regions based on 10 pronunciation features. While the vocalic criteria, most of which reflect phoneme splits or mergers, can easily be measured with distances in a continuous acoustic space, including the consonantal markers from Hughes et al. (2005), which are all based on the presence vs. absence of a sound, would imply a different methodology (e.g. computing edit distances between strings of hand-labelled phones). Thus a second reason is that, as far as we know, diagnostic consonantal features in the literature on the accents of the British Isles are often discrete. A third reason stems from the difficulty automatically to segment consonants (as opposed to vowels) in a speech corpus. Recall that one of our aims here is to obtain a quick picture of potentially interesting phenomena for accent diagnosis, and our intention is to keep human intervention minimal. It follows from this that manually segmenting consonants from large corpora somewhat clashes with our goals. Whatever the reason for not including consonants (or suprasegmental parameters) in our distance computation, the obvious consequence is that potential accent-specific information has been left out.

5. Conclusions

Our aim was to assess how much phonetic information on 13 English accents of the British Isles could be obtained from high-dimension feature spaces typically used in speech technology. The analysis included two distinct stages: in the first place, distances between vowels were submitted to hierarchical clustering and multidimensional scaling in order to produce tree diagrams and two-dimension plots illustrating vowel systems; then, clustering and MDS were applied to acoustic distances between accents, yielding an informative picture of the relationships between accents. Both stages are based on fully automatic procedures, which means that the results – contrary to those obtained with formants – are nearly instantaneous and entirely reproducible. The dendrograms and MDS plot representing vowel systems, though further improvements are required, constitute very appropriate tools in accent studies: potential phonemic mergers and splits, which are particularly useful in accent diagnostic, can be easily spotted. Turning to distances between accents, the dendrogram and the MDS plot exhibit a structure which is, overall, consistent with standard references on the accents of the British Isles.

The method could easily be employed without modification for the description of accent variation in other languages. Possible improvements include differential weighting of the original parameters, potentially more sophisticated clustering techniques and visual outputs thereof, and comparison with human perception.

References

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, *119*, 3048–3058.

Barry, W. J., Hoequist, C. E., & Nolan, F. J. (1989). An approach to the problem of regional accent in automatic speech recognition. *Computer Speech and Language*, *3*, 355–366.

Beal, J. (2004). English dialects in the North of England: Phonology. In E. W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A handbook of varieties of English. Phonology* (pp. 113–133). Berlin: Mouton de Gruyter.

Clopper, C. G., & Pisoni, D. B. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, *32*, 111–140.

Daniels, H. R. (1990). Accent recognition, listening positions and the case for the archiphone as a dynamic identity in speech perception. In *Proceedings of Colloque d'avril sur l'anglais oral* (pp. 27–49), Villetaneuse, France.

D'Arcy, S., Russell, M. J., Browning, S. R., & Tomlinson, M. J. (2004). The Accents of the British Isles (ABI) corpus. In *Proceedings of Modélisations pour l'Identification des Langues* (pp. 115–119), Paris.

De Wet, F., Weber, K., Boves, L., Cranen, B., Bengio, S., & Bourlard, H. (2004). Evaluation of formant-like features on an automatic vowel classification task. *Journal of the Acoustical Society of America*, *116*, 1781–1792.

Ellis, D. P. W. (2005). *PLP and RASTA (and MFCC and inversion) in Matlab* 〈http://www.ee.columbia.edu/~dpwe/resources/matlab/rastamat/〉 (19 September 2008).

Eremeeva, V., & Stuart-Smith, J. (2003). A sociophonetic investigation of the vowels OUT and BIT in Glaswegian. In *Proceedings of the international congress of phonetic sciences* (pp. 1205–1208), Barcelona.

Everitt, B. S., Landau, S., & Leese, M. (2001). *Cluster analysis*. London: Arnold.

Ferragne, E. (2008). *Etude phonétique des dialects modernes de l'anglais des Iles Britanniques: vers l'identification automatique du dialecte*. Ph.D. dissertation. Lyon: University of Lyon.

Ferragne, E., & Pellegrino, F. (2007a). An acoustic description of the monophthongs of East Anglia. In *Proceedings of the international congress of phonetic sciences* (pp. 1513–1516), Saarbrücken.

Ferragne, E., & Pellegrino, F. (2007b). Automatic dialect identification: A study of British English. In C. Müller (Ed.), *Speaker classification II* (pp. 243–257). New York: Springer.

Ferragne, E., & Pellegrino, F. (2010). Formant frequencies of vowels in 13 accents of the British Isles. *Journal of the International Phonetic Association*, *40*(1), 1–34.

Foulkes, P., & Docherty, G. D. (1999). *Urban voices: Accent studies in the British Isles*. London: Arnold.

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

Gold, B., & Morgan, N. (2000). *Speech and audio signal processing*. New York: Wiley.

Heeringa, W. J. (2004). *Measuring dialect pronunciation differences using Levenshtein distance*. Ph.D. dissertation, Groningen, Rijksuniversiteit Groningen.

Heeringa, W., Johnson, K., & Gooskens, C. (2008). Measuring Norwegian dialect distances using acoustic features. *Speech Communication*. doi:10.1016/j.specom.2008.07.006.

Hickey, R. (2004). Irish English: Phonology. In E. W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A handbook of varieties of English. Phonology* (pp. 68–97). Berlin: Mouton de Gruyter.

Holgersson, M. (1978). The limited value of cophenetic correlation as a clustering criterion. *Pattern Recognition*, *10*, 287–295.

Huckvale, M. (2004). ACCDIST: A metric for comparing speakers' accents. In *Proceedings of the international conference on spoken language processing* (pp. 29–32), Jeju, Korea.

Huckvale, M. (2007). Hierarchical clustering of speakers into accents with the ACCDIST metric. In *Proceedings of the international congress of phonetic sciences* (pp. 1821–1824), Saarbrücken, Germany.

Hughes, A., Trudgill, P., & Watt, D. (2005). *English accents and dialects. An introduction to social and regional varieties of English in the British Isles*. London: Hodder Arnold.

Jones, D. (2003). In P. Roach, J. Hartman, & J. Setter (Eds.), *English pronouncing dictionary*. Cambridge: Cambridge University Press.

Izenman, A. J. (2008). *Modern multivariate statistical techniques*. New York: Springer.

Kortmann, B., & Upton, C. (2004). Introduction: Varieties of English in the British Isles. In E. W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A handbook of varieties of English. Phonology* (pp. 25–33). Berlin: Mouton de Gruyter.

Kullback, S. (1968). *Information theory and statistics*. Mineola: Dover.

Legendre, P., & Legendre, L. (1998). *Numerical ecology*. Amsterdam: Elsevier.

McDougall, K., & Nolan, F. J. (2007). Discrimination of speakers using the formant dynamics of /uː/ in British English. In *Proceedings of the international congress of phonetic sciences* (pp. 1825–1828), Saarbrücken, Germany.

O'Connor, J. D. (1973). *Phonetics*. London: Penguin.

Olausson, L., & Sangster, C. (2006). *Oxford BBC guide to pronunciation*. Oxford: Oxford University Press.

Orton, H., Sanderson, S., & Widdowson, J. (1978). *The Linguistic Atlas of England*. London: Croom Helm.

Penhallurick, R. (2004). Welsh English: Phonology. In E. W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A handbook of varieties of English. Phonology* (pp. 98–112). Berlin: Mouton de Gruyter.

Scobbie, J. M., Hewlett, N., & Turk, A. (1999). Standard english in Edinburgh and Glasgow: The Scottish vowel length rule revealed. In P. Foulkes, & G. D. Docherty (Eds.), *Urban voices: Accent studies in the British Isles* (pp. 230–245). London: Arnold.

Sjölander, K. (2004). *Snack Sound Toolkit* (version 2.2) 〈http://www.speech.kth.se/snack/〉 (7 August 2008).

Stuart-Smith, J. (1999). Glasgow: Accent and voice quality. In P. Foulkes, & G. D. Docherty (Eds.), *Urban voices: Accent studies in the British Isles* (pp. 203–222). London: Arnold.

Stuart-Smith, J. (2004). Scottish English: Phonology. In E. W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A handbook of varieties of English. Phonology* (pp. 47–67). Berlin: Mouton de Gruyter.

Trudgill, P. (1990). *The dialects of England*. Cambridge: Basil Blackwell.

Upton, C. (2004). Received pronunciation. In E. W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A handbook of varieties of English: Phonology* (pp. 217–230). Berlin: Mouton de Gruyter.

Wales, K. (1999). North and south: A linguistic divide? *Reporter*, *439*, 〈http://www.leeds.ac.uk/reporter/439/kwales.htm〉 (19 September 2008).

Watt, D., & Allen, W. (2003). Illustrations of the IPA: Tyneside English. *Journal of the International Phonetic Association*, *33*, 267–271.

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

Wells, J. C. (2008). *Longman pronunciation dictionary*. Harlow: Pearson.

Wright, R. (2003). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI: Phonetic interpretation* (pp. 75–87). Cambridge: Cambridge University Press.

Young, S. (1994). The HTK Hidden Markov model ToolKit: Design and philosophy. *Entropic Cambridge Research Laboratory*, *2*, 2–44.