Proceedings of the Institute of Acoustics

THE DESIGN AND PERFORMANCE OF TWO ACCENT DIAGNOSTIC 'SHIBBOLETH' SENTENCES

K Edwards(1), J Iaver(1), M Jack(1) & A Simons(2)

(1) Centre for Speech Technology Research, University of Edinburgh
(2) BT Laboratories, Martlesham Heath

1. INTRODUCTION

Speech recognition systems which claim to be speaker independent must either address, or somehow bypass, the problems of inter-speaker variability. The issue may be avoided altogether with the simplest of recognition tasks, but they must be accounted for as complexity increases. Variability results from physical differences in our speech organs, and in the way we use them. We may be categorised by our typical speech style or range, or by a favourite phrase, but the most commonly used identifier of this sort is our regional accent. Although different accents do not cause major comprehension difficulties for speakers of the same language, simple HMM recognition strategies may not be able to cope with all accents and a system for accurately automatically classifying regional accent may be desirable.

This paper presents criteria for the diagnosis of nine accents of British English, based on the analysis of 'Shibboleth' speech data. Some initial results of this approach are presented, including suggestions towards an automatic system for accent decision-making. The speech used to test this strategy came from the Subscriber database, collected by BT Laboratories and annotated at CSTR as part of a collaborative research project.

2. CONSTRUCTION OF A SHIBBOLETH SENTENCE

The idea of a Shibboleth is a biblical one - an utterance which would give a fool-proof indication of the origin of talkers, based on their pronunciations of certain sounds. Provided the purpose of the Shibboleth is not revealed to the talkers, and the setting does not cause them to significantly alter their speech style, they are likely to betray their regional origin by using their normal accent. To distinguish between several accents in this way, an utterance must be constructed which exploits as many differences between accents as can be easily and consistently identified.

In the phonemic analysis of accents, [e.g. 1, 2] there are four major areas of difference: first, a systemic difference, where accents have different sets of phonemes in their consonant and/or vowel systems, i.e. not all the same sounds are available to be combined into words. Differences of this sort usually involve vowels - for example, speakers from most parts of England would use one vowel, /aɪ/, in their pronunciations of: 'side', 'like', 'cry', 'tied', but in Scottish vowel systems

there are two: /ʌi/ in 'side' and 'like' and /ae/ in 'cry' and 'tied'. Scottish accents also have a consonantal systemic difference to English ones, the velar fricative /x/ of 'loch', 'Buchan' etc.

The second difference is **distributional** or **structural**. In these cases, the same phoneme exists in two accents, but it is not used in all the same environments, and indeed could be used in completely different environments. An important distributional difference between accents of English concerns rhoticity. In rhotic accents, an 'r' in the spelling of a word is always pronounced with some kind of /r/ sound. In non-rhotic accents, preconsonantal and word-final 'r' is not pronounced, but word-initial and pre-vocalic 'r' is pronounced. Thus for a rhotic accent like that of the West Country, the words 'rich', 'poor', 'hard' and 'narrow' would be pronounced /rɪtʃ/, /pʊər/, /hɑrd/, /narou/. In a non-rhotic accent like that of Liverpool the words would be pronounced /rɪtʃ/, /pɔ/, /ha:d/, /narou/, so rhoticity is clearly an important divider of accents.

Selectional differences involve the distribution of a phoneme common to two accents, but only with respect to particular lexical items and not otherwise rule-governed. Thus in American English, the word 'herb' is pronounced /ɜrb/ with *no* initial aspirate, whereas in RP English an initial /h/ is always present. The dropping of initial /h/ in American English cannot be generalised - rather it is an exception to the rule, which has been selected for the word 'herb'.

The last major difference is **realisational** or phonetic: the same phoneme can have consistently different phonetic realisations in different accents. For example many Glaswegian speakers use a velarised, or dark [ɫ] in all 'l' environments, but amongst speakers from the Gaelic-influenced North West of Scotland, the /l/ is usually palatalised or clear [l] throughout.

The following principles were used in the design of the Shibboleth sentence:

- Coverage of nine selected accents had to be achieved.
- At least two indicators of each accent were present.
- Short words were used, to avoid multiple pronunciations of words within the same accent.
- The sentence had to be fairly easy to read aloud.

The sentences composed for this task were:

'I hear that some young bears were caught off guard by the rising tide.'

'One of the cubs soaked its brown pelt in a bath of muddy grey sea-water.'

These sentences were no more difficult for the callers to say than the other sentences collected for the Subscriber database - 64% of all the accent determining sentences followed the text exactly and contained no line or background noises.

3. THE ACCENT GROUPS

It should be emphasised that it is impossible to identify the regional origin of every speaker on the basis of phonetic or phonological cues. Aside from the fact that other cues are present in the speech signal (e.g. intonation and duration), the accent of many speakers is a product of their education and social aspiration, as well as the areas they may have lived in. The collection of reliable accent data can be further complicated by the 'telephone voice' formal speech style adopted by some callers.

There are also difficulties in selecting a set of accent categories because no matter how many groups are chosen, variations and sub-groups within each accent will exist. However, it seems likely that even a fairly crude accent decision made from reliable phonetic cues will be of use to speech recognition. For example, if a system contained lexica for several major accent groups, then a Shibboleth-based accent decision could be used to activate the appropriate lexicon for each user. A simple accent description may sound inadequate, but we do not yet know the optimal level of accent description for improving speech recognition. Some research has been carried out into the automatic diagnosis of accent using analysis of Shibboleth utterances [3] and the potential gain from regional accent modelling in speech recognition has also been demonstrated [4]. Bearing the above reservations in mind, the accent groups chosen were a first attempt at coverage of the UK:

R-IRISH	Rhotic accent for Northern Ireland
R-SCOTS	Rhotic accent for Scotland
R-LANCS	Rhotic accent for Lancashire
NB	Northern British accent
NB-LIV	Liverpool area accent
WAL	Welsh accent
R-WEST	Rhotic accent of the West Country
LON	Accent of the London area
SBS	Southern British Standard, RP accent

4. LABELLING STRATEGY

For the accent diagnosis, and annotation of the speech in the Shibboleth sentences, two levels of description were considered: firstly, a strategy using one large set of phonetic labels intended to cover all the accent groups, i.e. a 'superset' approach. Alternatively, the speech could be annotated phonemically with several smaller sets of labels, unique to each accent group, i.e. an accent-specific approach. There are advantages and disadvantages to each approach. Segmenting and labelling speech to a phonetic level with a very large set of labels is a more complex and time-consuming task than providing a simple phonemic description, but the number of phonetic labels and the level of detail in the description can be adapted to suit the purpose of the database. A phonetic description which is too detailed can be simplified if some phonemic information is known, but the reverse cannot take place. Phonemic labelling would provide a simpler, accent specific description of the speech using fewer labels, but some way would have to be found of describing callers with hybrid accents. If the database is to be used for HMM-based speech recognition experiments then phonemic labels are not suitable because the varying phonetic realisations described as a single

phoneme (e.g. [r], [ɹ]) would pollute the HMMs. The phonetic superset approach avoids this problem and has the extra advantage that vowels and consonants from different accents which are given the same label can be used to train the same HMM. It is also possible to recover accent specific descriptions and train accent specific HMMs from the superset of phonetic labels.

For the above reasons, a superset of some 75 quasi-phonetic labels was compiled, to describe the speech of every caller. The set of labels was based on the most thorough works available on accents of English [e.g. 1, 2] and was open to expansion if it proved inadequate.

The accent decision for each caller was based on particular vowel and consonant qualities found in the Shibboleth data, and is described in detail elsewhere [5]. Although each caller had to be placed into one of the categories, the phoneticians were allowed to note the influences of other accents, where they could be justified in terms of phonetic qualities. By basing the decision solely on the choice of phonetic labels, the human task is made to mirror that of an automatic system which could read the label file and build up scores for each accent from the vowels and consonants present, in a manner similar to that implemented by Barry et al. [3]. Although such a system has not yet been built, the accent decisions and labels produced for the Subscriber database are an ideal source of information for its construction.

5. THE SHIBBOLETH DATA

In a database of this nature, collecting an even spread of accents is a very difficult task. Even though their distribution in the database may reflect that in the UK's population, important minority accents still need to be accounted for. The set of accent groups chosen to categorise the callers was a first pass at covering the UK and was compiled without any knowledge of call locations or details of caller recruitment. For this reason, some of the nine accents are very strongly represented while others are not represented well enough to be compared with the rest. The coverage of accents achieved in the database is given in table 1:

Accent Decision	Callers
NB	376
SBS	177
R-SCOTS	154
R-WEST	124
R-IRISH	83
LON	58
WAL	25
NB-LIV	11
R-LANCS	6

Table 1. Numbers of callers diagnosed with each accent category

Given this uneven coverage, the three or four most poorly represented accents would need to be augmented with further data before they could be safely compared with the stronger accents, and

their predictive power cannot be relied upon. Therefore, at this stage, the analysis of Shibboleth label data has been restricted to the NB, SBS, R-SCOTS, R-WEST, R-IRISH and LON accents.

Unless a word spotting strategy is employed, an automatic analysis of Shibboleth label data would have to assume that the text of the sentences has been followed strictly. The numbers of correctly produced Shibboleth sentences from each accent are given in table 2:

Accent Decision	Shibboleth sentences
NB	452
SBS	231
R-SCOTS	198
R-WEST	166
R-IRISH	121
LON	71

Table 2. Numbers of correctly pronounced, uncontaminated Shibboleth sentences for each accent group.

Sections of the Shibboleth sentences were selected as vowel/consonant frames as follows:

'I h**EAR** that sOme y**OU**ng b**EAR**s w**ERE** c**AU**ght off-g**UAR**d by the r**I**sing t**I**de'

'One of the c**U**bs s**OA**ked its br**OW**n p**E**lt in a b**A**th of m**U**ddy gr**EY** sea-wat**ER**'

Looking at the phonetic labels used for the above sentences, several important observations can be made: the range of vowels and the most frequent vowels in a single frame can be established for each accent, giving an idea of the most robust and the most variable vowel cues for each accent, and also which words from the Shibboleths are most effective for each accent. Table 3 for example shows the three most common vowels in the word 'yOUng' for six accents. From it, we see that [ʌ] is by far the most frequently occurring vowel in 'young' in all the accents except NB, where [ʊ] is much more common. This difference can be exploited in distinguishing speakers of the NB accent from the others, but cannot separate all the accents. The importance of the [ʊ] vowel in this frame for NB accent is shown by its high intra-accent percentage and inter-accent probability. Looking at these 2 figures for all vowels in all frames allows a set to be compiled of the most important vowels and vowel frames for each accent. This information would be central to an automatic accent detector based on scanning phonetic label files. For each accent, the vowel present in the four or five most important frames could form the basis of a score for that accent. Other evidence from the rest of the Shibboleth data could also contribute to the score, although some kind of weighting would be appropriate. Tables 4 and 5 show the most robust vowels for the identification of the NB and R-SCOTS accents within the Subscriber database, i.e. the vowels which appear most often in the word frames shown for the NB and R-SCOTS accents and least often in those for the other accent groups. The figures underneath the accent name are the probability of the accent, given that vowel in the frame. Competing accents and their probabilities (> 0.1) are also shown.

ACCENT DIAGNOSTIC SHIBBOLETH SENTENCES

Accent	Vowel	% of total	P{Accent|Vowel}
LON	ʌ	94.59	0.2001
	ɒ	2.70	0.1218
	ʊ	2.70	0.0284
NB	ʊ	73.39	0.7728
	ʌ	18.88	0.0399
	ə	3.86	0.4012
R-IRISH	ʌ	75.81	0.1603
	ɒ	12.90	0.5819
	ʊ	11.29	0.1189
R-SCOTS	ʌ	97.12	0.2054
	ʊ	0.96	0.0101
	ɒ	1.92	0.0867
R-WEST	ʌ	89.02	0.1883
	ʊ	4.88	0.0514
	ɒ	1.22	0.0550
SBS	ʌ	97.37	0.2059
	ʊ	1.75	0.0184
	ə	0.88	0.0915

Table 3. Vowels in the frame 'yOUng'. The % figure is of the total number of vowels in that frame for that accent. The right-hand column gives the probability of the accent given that vowel in the frame.

ACCENT DIAGNOSTIC SHIBBOLETH SENTENCES

Frame	Vowel	Competing Accents & Probabilities		
'sOme'	ʊ	NB 0.9		
'yOUng'	ʊ	NB 0.8		
'gUARd'	a	NB 0.7	R-WEST 0.2	
'bAth'	a	NB 0.3	R-SCOTS 0.3	R-IRISH 0.3
'mUddy'	ʊ	NB 0.8		

Table 4. The most robust vowel cues to NB accent from Shibboleth sentences in the Subscriber database.

Frame	Vowel/Cons.	Competing Accents & Probabilities	
'hEAR'	iɹ	R-SCOTS 0.5	R-IRISH 0.4
'bEARs	eɹ	R-SCOTS 0.9	
'tIde'	ʌi	R-SCOTS 0.9	
'brOWn'	ʌu	R-SCOTS 0.8	
'watER'	əɹ	R-SCOTS 0.7	R-WEST 0.3

Table 5. The most robust vowel/consonant cues to R-SCOTS accent from Shibboleth sentences in the Subscriber database.

6. CONCLUSIONS

Initial analysis of the accent-diagnostic sentences from the Subscriber database supports the potential of these Shibboleth utterances to discriminate between accents of English, although their ability to separate all nine accent groups selected to cover the UK remains unproven due to the uneven representation of all the accent groups within the database. Despite the difficulty in collecting representative amounts of data, Subscriber remains a very useful resource on at least six accents of English. Detailed analysis of the phonetically labelled Shibboleth material has increased our knowledge of the intra and inter accent variation that exists, which should allow future Shibboleth designs to be optimised and eventually incorporated into an automatic system for detecting regional accent. The phonetically labelled sentence material in the database also provides an excellent opportunity to build accent-specific lexica for speech recognisers, based on observed rather than predicted forms.

7. ACKNOWLEDGEMENTS

The authors would like to thank BT Laboratories for their support in funding this research.

8. REFERENCES

[1] J C Wells, *Accents of English* (vols 1 and 2). Cambridge University Press (1982).

[2] A Hughes, & P Trudgill, *English Accents and Dialects*. London: Edward Arnold (1979).

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

[4] K Edwards, F R McInnes & M A Jack, 'Accent specific modifications for continuous speech recognition based on a sub-word lattice approach', *Proc. Eurospeech 91* 3, 1195-1198 (1991).

[5] C M Scott, R Lickley, K Edwards & A Simons, 'Standards for the phonetic labelling and segmentation of telephone quality speech', *these proceedings* (1992).

[6] J Laver, 'Accent-diagnostic indicators in British English: Implications for automatic speaker-independent speech recognition', *Subscriber Project Report*, CSTR (1990)