

The Accents of the British Isles (ABI), corpus

Shona M D'Arcy¹, Martin J Russell¹, Sue R Browning^{1,2}, Mike J Tomlinson²

¹Electronic, Electrical and Computer Engineering, The University of Birmingham,
Edgbaston, Birmingham B15 2TT, United Kingdom

²20/20 Speech Ltd, Malvern Hills Science Park, Geraldine Road,
Worcs WR14 3SZ, United Kingdom

Darcys@bham.ac.uk, M.J.Russell@bham.ac.uk, Sue_browning_2004@yahoo.co.uk
m.tomlinson@2020speech.com

ABSTRACT

This paper describes the ABI (Accents of the British Isles) speech corpus. The corpus comprises approximately 95 hours of recordings from approximately 300 subjects (equally divided between male and female), whose speech is representative of 15 different regional accents of British English. The paper describes the planning and recording of the corpus, its strengths and limitations, and some of the lessons which were learnt during the recording.

1. INTRODUCTION

In the British Isles, the diverse range of regional accents is often cited as a source of difficulty for both automatic speech recognition and synthesis. However, to our knowledge there has been no systematic study of the effects of accent on speech recognition accuracy.

Some research effort has been applied to understanding the effects of accents associated with different English speaking nations, in particular American, British and Australian English [8] Irish English have also been studied [1].

The inhabitants of the British Isles speak with a broad range of accents, each with specific phonetic properties, and these have which may influence the diversity of accents is the fact that the British Isles includes five different countries, England, Wales, Scotland, Northern Ireland and the Republic of Ireland. These, plus other regions, have their own distinct cultural identities, which include their regional accents. Historically, some of these regions even had their own languages, whose phonetic properties now influence the local English pronunciation. However there is also significant variation within each of these regions. England especially has many different accents. Indeed, the differences between, say, London Cockney and Newcastle 'Geordie' is obvious to most listeners.

One of the main obstacles to a systematic study of the implications of the various accents of the British Isles for speech technology is the absence of suitable speech corpora. The goals of the ABI (Accents of the British Isles) project are to create such a corpus, to analyse the data in order to characterise the acoustic correlates of accent, and to conduct a systematic experimental study of the effects of accent on speech recogniser performance. This paper describes the procedures which were used to plan and record the ABI corpus. It describes the successes and shortcomings of the project and some of the lessons learnt.

2. EXISTING ACCENT CORPORA

An existing corpus of British accents is the IViE corpus developed by The Phonetics Laboratory, University of Oxford, which consists of read and spontaneous speech from 9 areas around the British Isles. The corpus contains 36 hours of speech data from 16 year olds. This corpus is a valuable resource but it is not as comprehensive as the ABI corpus. The work completed to-date on this corpus has mainly focused on studying intonational variations of interest to us here.

3. ACCENTS OF THE BRITISH ISLES

To a first approximation, accents of the British Isles are identified with broad geographical regions, namely southern and northern England, Wales, Scotland and Northern Ireland and the Republic of Ireland. The accents covered by the ABI corpus correspond to more specific areas, but these broad regions are sufficient to highlight some of the difficulties in distinguishing accents at even this high level. One well known distinction between southern English speech and that of most of the rest of the British isles is the use of /æ/ compared with the /a:/. This can be heard in words like 'glass', 'bath' and 'dance', where typically southerners use the /a:/ and northerners (including Irish and Scottish) use /æ/. Unfortunately, rules like this do not generalise. For example, the above rule does not apply to words such as bad or camp where /æ/ is used by all speakers.

Not all accent groups use the same number of phones. Two distinct sounds in one accent may be replaced by a single, common sound in another. For example, in Scotland and Northern Ireland speakers use the same vowel sound in the words 'cot' and 'caught', whereas two distinct sounds would be used in other regions.

Another phenomenon in certain areas of the British Isles is to omit the /h/ consonant from the beginnings of words. This is most usually heard in urban accents in England and Wales, and is a particular characteristic of London (cockney) accents. However this habit has become fashionable and has extended to other urban areas.

These examples illustrate just some of the more obvious phonetic differences between different accents of the British Isles, and these can be heard in the ABI corpus. Of course, any individual will adjust the way in which they speak according to circumstances. Accent is likely to manifest itself much more strongly in conversations between friends with the same social and linguistic back-

grounds that in conversations between strangers, or in formal, read speech. The latter in particular must be taken into account when analysing the ABI corpus, where the majority of the material is read speech.

4. CORPUS DESIGN

The first stage of the ABI project, and the main topic of this paper, was the collection of a corpus of speech that is representative of the accents of the British Isles. With this in mind, the main issues for the design of the corpus were:

- What constitutes an accent?
- Which accents should be recorded?
- Which subject should be recorded?
- What material should each subject speak?

Initially the key difficulty was the definition of what constitutes an ‘accent’. For a native British English speaking listener it is normally quite easy to say whether or not a person has a particular type of accent. However different judges may not agree that that person has a ‘good’ or ‘strong’ accent, or the precise definition of that accent. For example, in the Birmingham recordings, some apparently suitable subjects declined to take part because they considered themselves to have ‘black country’, rather than Birmingham accents. This appears to be a common phenomenon. In most of the areas considered there were accent variations that contain acoustic parameters of the overall accent but are very distinctive to locals, who may claim they are completely different accents.

This problem was circumvented by focusing on specific regions associated with accents, rather than attempting to define the accents themselves explicitly. Fourteen regions associated with different accents were identified. For each of these regions a town or city was chosen whose population, we believed, would speak with a variant of the required accent (table 1). These locations represent both urban and rural accents. Of course, local accents will vary between different areas in a particular city, and this fact was pointed out on many occasions by subjects who took part in the recordings. For example, the subjects in Hull distinguished between accents from the East and West of the city.

Standard Southern English	Ulster (Belfast)
Midlands (Birmingham)	NE England (Newcastle)
Wales (Denbeigh)	Scotland (Glasgow)
Scottish Highlands (Elgin)	Inner London
Republic of Ireland (Dublin)	NW England (Liverpool)
East Yorkshire (Hull)	East Anglia (Lowestoft)
Lancashire (Burnley)	West Country (Truro)

Table 1: : *Regions of the British Isles (together with the corresponding towns or cities) where the ABI recordings were made.*

In each location twenty people were recorded, ten female and ten male. In order to capture local accents only, subjects were required to have been born in that location and to have lived there all of their lives. All recordings for a particular accent were made on location in the chosen town or city.

The texts to be recorded were designed by 20/20 Speech Limited. The objective was to elicit accent specific phenomena and also to provide examples of typical application words, sentences and phrases. The recording consisted of readings of 20 prompt files, and each subject was asked to read every file. The prompt texts are roughly divided into 2 categories; short and long phrases.

The short phrases included, for example;

- game commands (e.g: “change view”, “grab image”, “toggle source”, “select left”)
- ‘careful’ words used to elicit vowel sounds (e.g: “hide”, “hoid” (to rhyme with ‘void’), “hoed” (to rhyme with showed), “howd” (to rhyme with ‘loud’)
- letters and the international radio operator’s alphabet (e.g: “G P Y O”, “golf”, “papa”, “yankee”, “oscar”)
- digit sequences (e.g: ‘four zero nine one”)
- short phrases (e.g: “it’s so sweet”, “while we were away”, “thin as a wafer”, “has a watch”, “roll of wire”)

These prompts were generally presented to the subjects in sub-lists, so that if an error was made the subject was required to repeat only the sub-list and not the whole list.

The long phrases contained:

- A short ‘accent diagnostic’ story (“When a sailor in a small craft...”)
- Equipment specific commands (e.g: “climate control seventy one degrees”, “navigation select route home”)
- SCRIBE¹ sentences (e.g: “Gary attacked the project with extra determination”, “I itemise all accounts in my agency”)

When an error was made here, the subject was asked to repeat the sentence. The final screen asked a set of questions. The objective was to elicit some spontaneous speech and also to gather information about the subject, such as age and height.

5. RECORDING PROTOCOL

Two weeks before each recording session, the Public Relations Office in the University of Birmingham² sent out press releases to all of the local media in the target location. Because of the high level of general interest in local accents in the British Isles, this normally resulted in a number of newspaper, radio and television features. Each of these features included a request for suitable subjects to call a ‘freephone’ message service to volunteer to take part in the data collection. All calls were acknowledged, and the most suitable candidates were chosen from the messages which they left.

¹ The SCRIBE sentences are an Anglicised version of the TIMIT sentences

² The authors would like to acknowledge the contribution of Kate Chapple, of the University of Birmingham Public Relations Office, who facilitated the recruitment of subjects by bringing the ABI project to the attention of local media in the target locations

In general this strategy for recruiting subjects was very successful, and in some regions over 100 volunteers phoned in shortly after the initial media coverage. However, in other regions there were very few volunteers (in one region only two people contacted the 'freephone' number, and one of these only phoned to point out the futility of trying to capture the full range of that particular regional accent by making recordings in just one location. In another location there were no calls and the recording session was abandoned). We believe that this probably reflects local differences in social attitudes to regional accent.

In those locations where the 'media campaign' was successful the standard of the recordings tended to be close to the original goals of the project, since we were able to select the subjects with reference to an example of their speech. Some people who replied to the advertisements did not have a particularly strong accent or were not proficient at reading, and these subjects were not used. In areas where the media campaign was less successful we were forced to be less selective, though volunteers with no discernible accent were not recruited. In these cases the volunteers from the 'freephone' route were supplemented by volunteers recruited 'on location' over the days of the recordings.

The preferred recording location was a room in a public library. This presented a problem for some regions, as a town had to be chosen that represented the desired accent but also had a large enough population to support a library and provide the required number of qualified volunteers. Libraries were the preferred recording locations because of their quiet nature, the availability of a room that could be rented, and the availability of a pool of suitable potential recording subjects to replace any who failed to attend, or to fill other gaps in the recording schedule.

It was also noted that in locations where a library was not used the profile of subjects was different to that encountered in library-based recordings, for example in terms of levels of literacy. Even when a room in a library was used, there was variation in the background noise levels present in the room (for example, in some libraries the 'quiet' room houses the server for the library's local computer network).

6. EQUIPMENT

This section describes the equipment used to record that ABI corpus.

6.1. Hardware

The subjects were recorded directly on to a computer hard disk. A laptop was used for mobility. The additional hardware which was used comprised an Emkay head mounted microphone and a Telex desk mounted microphone (for near and far field recording, respectively), an Edirol UA-5 USB sound card interface (this removed any device specific factors associated with individual laptop soundcards), and an Emkay VR3294 Battery Box (to provide bias voltage for the microphones). In addition, a digital camera was used to take a picture of each subject for future reference.

6.2. Software

The list of prompts described above was incorporated into the ABI prompting and recording software³. This software controlled each recording session by displaying prompts, recording speech at a sample rate 22050 samples per second and 16 bit resolution, and saving the speech files in a logical directory structure.

6.3. Procedure

As each subject entered the recording room or area, a short explanation of what was expected of them was given. They were told that they would simply be required to read text from a computer screen. Subjects who did not appear to be confident in their reading abilities were reassured that the task was not difficult and that help (human prompting) would be given if necessary. Each subject was asked to sign a consent form, giving the University of Birmingham permission to use the data recorded and the images taken by the digital camera.

Subjects were seated in front of the laptop and the headset microphone placed on their head with the microphone angled about 2cm away from the right corner of their mouth. The desk-mounted microphone was placed to the left of the laptop. Every effort was made to ensure that all subjects were approximately the same distance from the far field microphone.

The software was controlled by the researcher in charge of the session, rather than by the subject. The appropriate set of prompt texts was loaded. The first file to be recorded was an introductory passage that asked people to say their name and read some phrases; this file was used to adjust the recording levels for both microphones using the gain controls on the Edirol UA-5. The objective was to achieve peak recording levels between -25 to -30 dB. This initial session was recorded repeatedly until the researcher was happy with the levels. The subject was then recorded reading all of the material. After the session the levels were checked again to ensure that the recorded signals had not been 'clipped'. Each subject was paid £15 for taking part in the project.

The corpus was recorded over a period of approximately 4 months and comprises approximately 95 hours of read speech. All of the data has been annotated at the word or phrase level by 20/20 Speech Limited.

The original objective was to restrict the recordings to people between the ages of 18 and 50, so that effects due to very young or old speakers are not included in the corpus. The graph in figure 1 shows the actual age distribution of male and female subjects. As can be seen from the graph, the corpus includes speech from a significant number of subjects aged 60 or over, particularly male subjects. This reflects the relative availability of older male subjects during normal working hours. The figure also shows a significant number of subjects aged less than 20, but in this age group female subjects dominate.

³The ABI prompting and recording software was designed and implemented by Paul Dixon, Department of Electronic, Electrical and Computer Engineering, The University of Birmingham

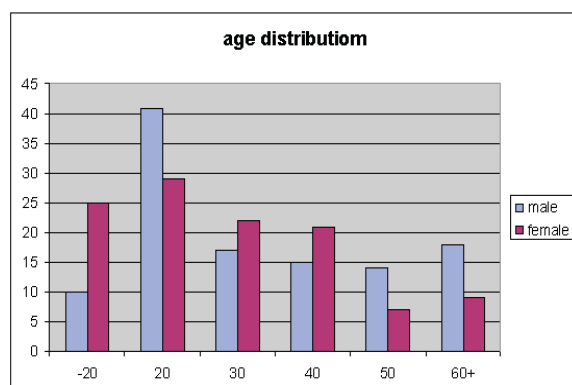


Figure 1: *Distribution of subjects according to age*

7. LESSONS LEARNT

Despite our best efforts, there is some inconsistency between different parts of the corpus.

The insistence that a potential subject should have lived in a location for all of his or her life was largely successful in obtaining subjects with local accents, however there were certainly a small number of subjects who satisfied this ‘residency’ criterion but did not exhibit a noticeably strong accent. In some cases a subject would consider himself or herself to satisfy the residency criterion, but it would become apparent later, in the spontaneous, conversational section at the end of the recording session, that he or she did not.

In locations where there were a large number of volunteers, the ‘freephone’ phone messages could be screened to remove subjects with little evidence of a local accent. The sessions in which there were a large number of such volunteers and a suitable room in a library was secured are probably the most successful. The press releases were published in local papers or radio, whose readers or listeners were usually long term residents of the area. When people had to be recruited ‘on location’ the ‘quality’ of the accents was generally not as good.

In some areas, such as Northern Ireland and Birmingham, we were able to call on local knowledge to obtain any additional subjects that were required, and to identify suitable recording facilities. We believe that these sessions are comparable with those with a large number of ‘freephone’ volunteers.

Libraries are a valuable resource, as they contain literate local people with time on their hands who can fill gaps in the recording schedules. In cases where a room in a library was not available, a room in another accessible public place, such as a Community Centre, was used. With hindsight, the noise levels in such rooms were generally relatively high, as these buildings are foci for many different activities, most of which involve noise. Also, the literacy skills of subjects recruited on location in these places could not be guaranteed.

Subjectively, the most variable sessions are ones which include a mixture of a small number of ‘freephone’ volunteers and subjects recruited on location, and where the

location was not in a library.

For some accents, relatively small towns were targeted as it was felt that there might be less migration and that there would be a population who had lived in the town for generations. However, this was not always a successful strategy. In Denbeigh in Wales, very few of the subjects recorded have what most British English listeners would consider to be a strong Welsh accent. On reflection we believe that this is due to the relatively close proximity of Liverpool. Also, a disproportionate number of volunteers in some of the smaller towns were outside the target age bracket.

A similar problem is evident in the London recordings, where, subjectively, few of the subjects have strong accents.

The fact that the majority of the corpus comprises read speech, prompted by written prompts, also caused some difficulties. It has already been noted that in some locations the reading proficiency of some of the volunteers caused problems. However, even for literate subjects the accents themselves caused some difficulties. For example, subjects were asked to say “hured” (to rhyme with “cured”) and “heard”, but for many of the subjects in Lowestoft there was no distinction between the sounds of these two words, and both were pronounced “heard”.

If, in the future, the goal is to create a more consistent set of recordings for different accents, then we believe that it would be important for all regions to be approached in a consistent manner. All subjects should be recruited in advance through the ‘press release’ and ‘freephone’ mechanism, and rooms should be of a consistent type (for example all in libraries). The room should also be inspected prior to the recording sessions and another room arranged in cases where the first is not suitable. If any of these criteria could not be met, then another location should be chosen.

A further issue is that in some areas younger people appeared to be trying to lose their accents. Contrary to popular beliefs people in some areas do not appear to be proud of their local accents, and would not agree that they had the accent named in the press release. One solution would be to use a general phrase such as ‘local accent’ rather than something more specific, such as ‘Cornish accent’.

8. CONCLUSIONS

This paper describes the design and recording of the ABI (Accents of the British Isles) speech corpus. In particular it describes the contents of the corpus, and the procedures for selecting recording locations and soliciting volunteers, are described. The paper gives an accurate account of the strengths and limitations of the corpus, and of the lessons which have been learnt. In addition to its potential utility for speech technology, the corpus is a unique ‘snapshot’ of the range of accents spoken in the British Isles at the start of the 21st century.

A second phase of recordings are planned for the future, which will include those regions and accents which are not covered the existing ABI corpus.

BIBLIOGRAPHIE

- [1] S. Fitt, and S. Isard. «Sunthesis of regional english using a keyword lexicon». In *Proc. Eurospeech '99, Budapest, Hungary*.
- [2] E. Grabe, (2002, April). «Variation adds to prosodic typology». In *Proc. Speech Prosody 2002 Conference, Aix-en-Provence*.
- [3] M. Huckvale, (2004). «Accdist: a metric for comparing speaker's accents». In *Proc. Int. Conf. on Spoken Lang. Proc., Juju Island, Korea*.
- [4] A. Hughes, and P. Trudgill (1996). *English accents and dialects*. Arnold.
- [5] C. Teixeira, I. Trancoso, and A. Serralheiro (1997). «Recognition of non-native accents». In *Proc. Eurospeech '97, Rhodes, Greece*.
- [6] J.C Wells, (1982a). *Accents of English, Volume 1*. Cambridge University Press.
- [7] J.C Wells, (1982b). *Accents of English, Volume 2*. Cambridge University Press.
- [8] Q. Yan, S. Vaseghi. « A comparative analysis of uk and us english accents in recognition and synthesis ». In *Proc. IEEE-ICASSP, Orlando, FL*.
- [9] Q. Yan, S. Vaseghi. «Analysis, modelling and synthesis of formants of british, american and australian accents». In *Proc. IEEE-ICASSP, Hong Kong*.

ABI, recording session, recordings, Cambridge University Press, Eurospeech '99, S. Fitt, S. Isard, Northern Ireland, speech corpora, E. Grabe, Paris, C. Teixeira, A. Serralheiro, English accents, London Cockney, BIBLIOGRAPHIE, Budapest, Hungary, systematic study, public library, background noise levels, regional accent, volunteers, subjects, British Isles, regional accents, British accents, recording, speech corpus, automatic speech recognition, The University of Birmingham, Electrical and Computer Engineering

The accents of the British Isles (ABI) corpus