Perceptual Recognition Cues in Native English Accent Variation: "Listener Accent, Perceived Accent, and Comprehension"†

Ayako Ikeno and John H.L. Hansen

The Center for Robust Speech Systems (CRSS)
Erik Jonsson School of Engineering and Computer Science
University of Texas at Dallas
Richardson, Texas 75083, USA, http://crss.utdallas.edu

ABSTRACT

There are many aspects of speech that can provide information about a particular speaker's characteristics. Accent is a linguistic trait of speaker identity. It indicates the speaker's language and social background. The goal of this study is to provide perceptual human recognition of English native accent variation for accent and dialect identification applications. To examine relationships of the listener's accent background with perceived accent and comprehension of the speech, perceptual experiments are conducted with three types of listeners – US and British native English listeners, and nonnative English listeners. The tasks are accent detection and classification, and transcription of the speech. The results from the study show that listeners' accent background significantly impacts accent perception. The results also indicate that listeners use perceptual cues differently based on the task. Our analysis also suggests that comprehensibility of the speech affects accuracy of accent detection and classification. These observations point to the complex nature of the cognitive process involved in accent perception, which is bidirectional (bottom-up and top-down processing) and multi-dimensional (speech perception, language comprehension, etc.). This suggests the importance of understanding accent variation from a cognitive perspective for further development of accent and dialect identification systems as well as speech processing algorithms in general.

1. Introduction

Accent (or dialect) is a crucial factor for speech technology in various areas including business, forensics and security, and language education, as illustrated in Fig. 1.

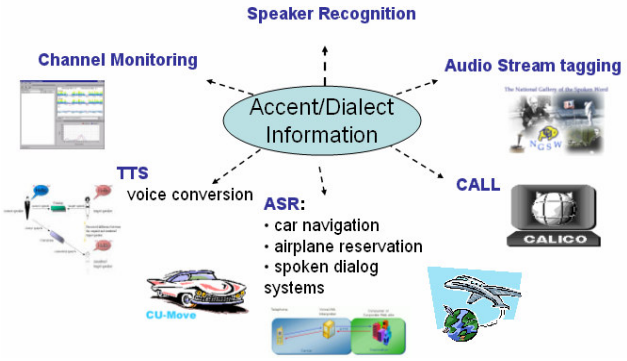

Figure 1. Applications that can benefit from Automatic Recognition of Accent and Dialect Information

For example, identification or classification of speaker accent can provide useful information for Automatic Speech Recognition and Understanding. Accent and dialect characteristics can also provide important information about speaker identity. Automatic Accent Identification and Classification, therefore, is an important part of technological application of accent characteristics for forensics and security. Investigating the cognitive aspect of accent variation is important, since factors based on how humans categorize accents provide meaningful insight and knowledge for further development of accent classification algorithms [1, 2] and speech technology.

The goal of this study is to identify speech characteristics that distinguish different accents perceptually across a variety of native and nonnative English accents. Specifically, this study will examine how the listeners' accent background affects their detection and classification accuracy of speakers' accent type, and comprehensibility of the speech.

Accent and *dialect* both refer to linguistic variation of a language. Use of these terms can be ambiguous. In this paper, we use the term *accent* as defined in Crystal[3] – "The cumulative auditory effect of those features of pronunciation which identify where a person is from regionally and socially. The linguistic literature emphasizes that the term refers to pronunciation only, is thus distinct from dialect, which refers to grammar and vocabulary as well."

In human perception, listener familiarity with a particular type of accent has been shown to affect their comprehension of the speech. In Bent and Bradlow[4], for example, it was shown that when listeners hear speech in their native accent, their comprehension is much easier. More specifically, Korean accented English is more comprehensible than Chinese accented English for Korean listeners. A variety of studies have explored issues regarding intelligibility and comprehensibility of nonnative accented speech as well as perceived nonnative accents (e.g., Flege[5], Jilka[6], Megan[7], Munro and Derwing[8]). The analysis in this paper focuses on native accent variation and addresses the following two issues: 1) relationships between listeners' native accent and perceived accent of the speech, and 2) relationships between perceived accent type and comprehensibility of the speech. Perceptual experiments test how accurately listeners with different native accents can identify an accent in question and comprehend the speech. Listener familiarity is grouped into three categories – native (US, British) and nonnative (their first language is not English).

† This work was supported by the U.S. Air Force Research Laboratory, Rome NY under contract No. FA8750-05-C-0029.

2. Data, Listeners & Experiments

The test data represents the following native and nonnative English accents – US, Irish, British English, Welsh, and Canadian native accents, and Chinese, French, German, Japanese, Spanish, Thai, and Turkish nonnative accents. The data set is composed of single words, phrases, and sentences extracted from three corpora (CU-Accent[9], IviE[10], N-4[11]).

The number of listeners totals 33 with 11 from the US, 11 from England, and 11 from non-English speaking countries. Nonnative listeners' native languages are Chinese (1 listener), Croatian (1), German (1), Japanese (1), Korean (3), Spanish (1), Thai (2), and Tigrinya (from Ethiopia, 1).

	US	British	Nonnative
Number of Listeners	11	11	11
Age	22-34	27-43	24-36
Age of Arrival (US)	N.A.	22-40	18-33
Years of Residence in US	N.A.	1-10	2-12

Table 1. Summary of Listener Information

The listening test was conducted individually in an ASHA certified sound booth by using an interactive computer interface and a headset. Tasks are: 1) perceptual detection of native vs. nonnative English accent, 2) perceptual classification of UK accents, and 3) human transcription of the speech.

3. Human Accent Detection & Results

Based on acoustic cues without relying on grammatical and lexical characteristics, how accurately can listeners detect English accents? For the task in this section, listeners are presented with a set of native and nonnative English speech samples, consisting of 1 to 14 words extracted from spontaneously produced utterances (6 words per utterance on average), and asked to detect the type of accent (native vs. nonnative English). Listeners were informed that native accent, in this test, represents speech produced by speakers whose first language is English (e.g., speakers from the US, Canada, and the UK), whereas nonnative accent represents speech produced by speakers of English as a second or foreign language (i.e., their first language is not English). They were also asked to indicate a confidence rating on a 1-5 point scale for each selection. Listener confidence was rated as shown in Fig. 2.

1	2	3	4	5
not sure at all		somewhat sure		Absolutely sure

Figure 2. Confidence Ratings

The result in Fig. 3 illustrates native and nonnative listeners' accent detection accuracy with overall confidence (ratings 1–5). The distribution of the speaker accent categories shown here represents UK native English – Cambridge, Belfast and Cardiff.

The results show both listener-group-dependent and speaker-accent-dependent trends. The less familiar the listeners are with an accent, the lower the detection accuracy – British: 90%, US: 73%, nonnative: 55% on average. In addition, for unfamiliar listener groups (US and nonnative), Belfast accent is misperceived as a nonnative accent about half of the time (45% to 55%), which is significantly more often compared to the cases of Cambridge accent and Cardiff accent.

Further analysis based on the contextual variation (single words vs. phrases), as shown in Fig. 4, indicates that longer contexts (2 to 14 words, 7 words on average) contribute to detection with significantly higher accuracy for all three listener groups – British, US and nonnative. Especially in the case of British listeners, when speech samples with two or more words are provided, UK accents are perceived as native 100% of the time. This would suggest a benchmark necessary to achieve for automatic accent classification systems.

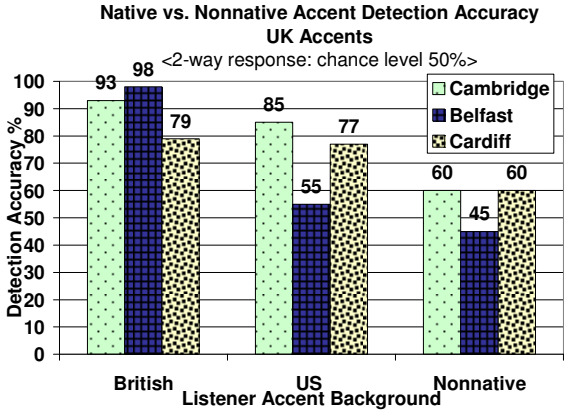

Figure 3. Native vs. nonnative accent detection accuracy: for example, Cambridge accent was correctly perceived as native by British listeners 93% of the time.

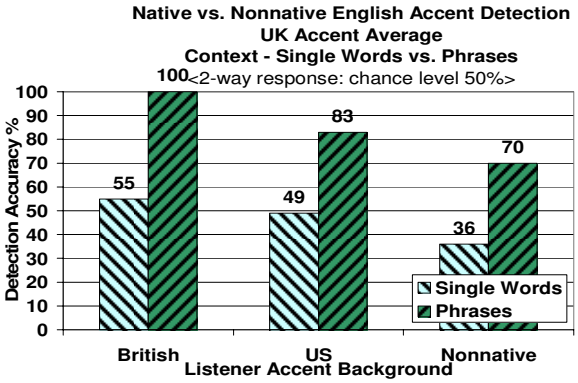

Figure 4.

4. Human Accent Classification & Results

In the accent classification task, three UK native English accents were used in the test data set: Cambridge-British English (Accent 1), Belfast-Irish (Accent 2) and Cardiff-Welsh (Accent 3). An approximately 60-second-long audio file per accent was provided for listener training and reference. Listeners were not informed where those accents originated.

They listened to a set of test audio files, consisting of 1 to 27 words extracted from spontaneously produced utterances (9 words per utterance on average), and classified each as Accent 1, Accent 2, or Accent 3. They were also asked to provide a confidence rating (1-5) for each selection (previously shown in Fig. 2).

For the UK native English accent classification task, US and nonnative listeners' accuracy is both significantly lower compared to British native listeners' accuracy, as shown in Fig. 5. This trend is especially clear in the case of Belfast accent, where British native listeners classified it correctly 91% of the time while US and nonnative listeners were able to do so only 66% and 38% of the time respectively.

This result also indicates that Welsh accent is often unidentifiable by all three listener groups. It is suggested that the listeners have either less familiarity with Welsh accented English or that speech production cues are not sufficiently distinct to convey this accent to the listener pool.

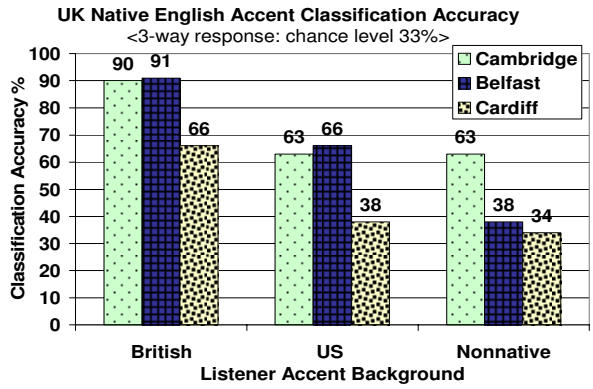

Figure 5

To investigate this further, a pairwise comparison of accent confusability was performed. Fig. 6 illustrates results, where Cardiff accent is often misperceived as Cambridge accent by all three listener groups. For all three listener groups, Cambridge accent and Belfast accent are the least confusable. On the other hand, Cambridge accent is not misperceived as Cardiff accent as often by any of the three listener groups, as shown in Fig. 7. This indicates that confusability between two accents is not mutually equal. As illustrated in Fig. 8, Belfast accent is mistaken as Cambridge accent only 3% to 14% of the time, which is significantly lower compared to the cases of being correctly perceived as Belfast accent (38% to 91%) and being misperceived as Cardiff accent (6% to 48%).

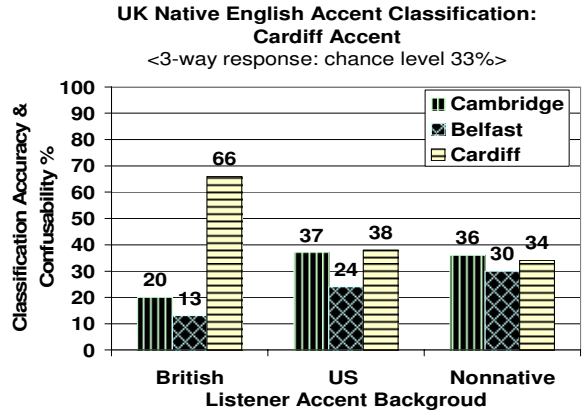

Figure 6

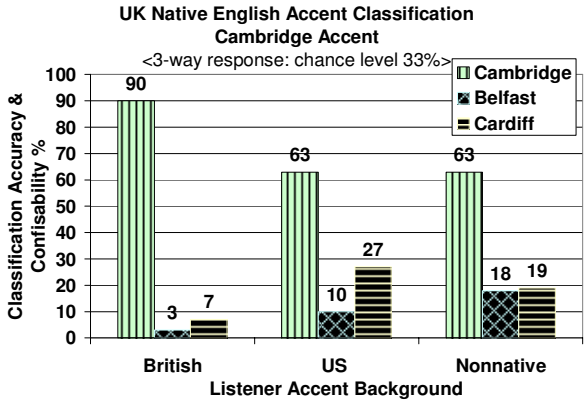

Figure 7

One of the clearest and important trends observed here is that certain types of accents are more confusable than others for all three listener groups. This suggests that certain accents are perceptually more clearly identifiable than others even if they are unfamiliar to the listeners, although the degree may vary.

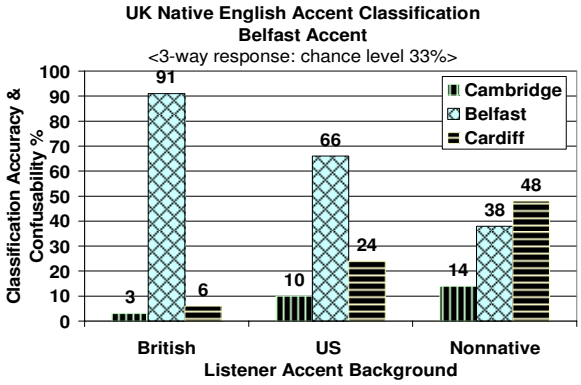

Figure 8

In classification, longer context (phrases, 10 words per utterance on average) contributes to classification with significantly higher accuracy for familiar (British) listeners, as illustrated in Fig. 9. Unlike the detection task, unfamiliar listeners (US and nonnative) do not benefit from longer context.

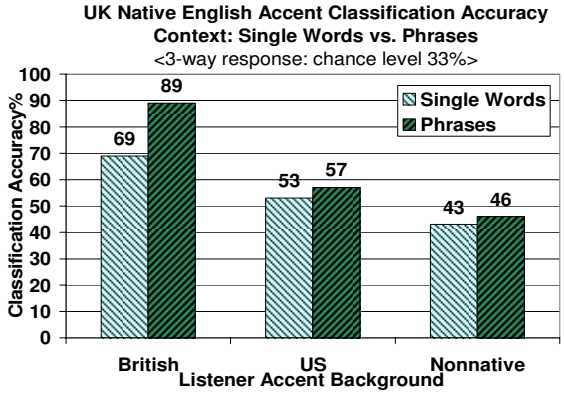

Figure 9

5. Human Transcription & Results

In this third task, speech samples were orthographically transcribed by listeners. This again would establish a useful benchmark to achieve for automatic ASR based transcription systems. Transcriptions were scored based on their word error rates in a manner consistent with WER in automatic speech recognition. We note here that the transcription accuracy was calculated as 100% minus WER.

The overall results are shown in Fig. 10. Transcription accuracy is affected by the listeners' nativeness to the language rather than their accent background. Both British and US native English listeners comprehended the speech similarly (average of 78% and 82%) in comparison to nonnative listeners (48%).

Further analysis of transcription accuracy indicates that higher comprehensibility of the speech does not mean higher accuracy in detection or classification rate, especially in the cases of native (British and US) listeners. For example, Cardiff accent is clearly more comprehensible than Belfast accent although its classification accuracy is lower, as illustrated

earlier. However, comprehensibility of the speech impacts detection and classification accuracy in that more comprehensible speech tends not to be confused with less comprehensible speech (e.g., Cardiff accent is less often misperceived as Belfast accent than as Cambridge accent) but confused with similarly comprehensible speech (e.g., Cardiff accent is more often misperceived as Cambridge accent than as Belfast accent).

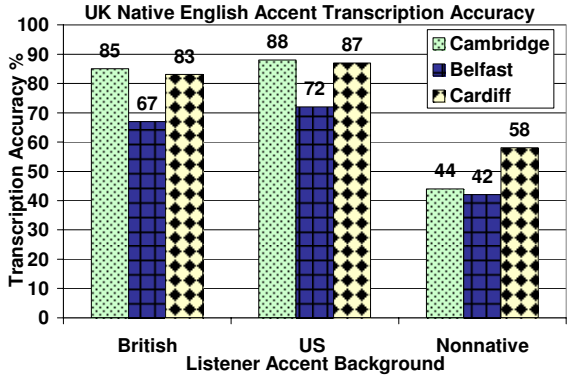

Figure 10

Both US and British native listeners seemed to had some difficulty in understanding speech with Belfast accent. This is also contrastive to results from the classification task since Belfast accent was the most clearly identifiable accent in comparison to Cambridge accent and Cardiff accent.

According to these observations, it is suggested that the native listeners classified more comprehensible speech as Cambridge accent and less comprehensible speech as Belfast. That can partially explain why Cardiff accent was often confused with Cambridge accent where Belfast accent was not. Cambridge and Cardiff accents are equally well understood by the listeners. The detection result suggests, in addition, that US listener tended to misperceive less comprehensible accent, Belfast accent in this case, as nonnative accent.

In essence, more comprehensible speech does not mean better classification accuracy. However, in both classification and detection tasks, the listeners seem to have used the degree of comprehensibility as an indicator of their own familiarity with the given accent. In this sense, the relationships between detection or classification accuracy and transcription accuracy are associated.

6. Discussion

In the area of automatic accent classification and automatic transcription via ASR engines, it is important to establish human performance in order to assess the significance of automatic systems. The two main observations from our results are that the listeners' familiarity with the language and their familiarity with the accent both impact accent perception. Listeners' familiarity with the language, which is English in this case, affects the comprehensibility of the speech more severely than accent perception. This trend is clearly observed in the differences between nonnative listeners' and native (both US and British) listeners' performance on the given tasks. For example, nonnative listeners' classification accuracy on Cambridge accent is very similar to American listeners' (63% and 64% respectively). However, their transcriptions are only half of the time correct (44%) compared to those of American listeners' (88%).

On the other hand, listeners' familiarity with a particular accent may affect accent perception rather than comprehensibility, as can be seen in differences between US and British native listeners' performance. US listeners are able to understand what was said by UK accented native speakers. However, in the detection task, for example, the speech can be mistaken as having nonnative accent (i.e., speech produced by speakers of English as a second language). This tendency is even stronger using speech with a Belfast accent (45% detection accuracy). These trends indicate that different types of speech characteristics provide cues for listeners to understand the speech and to distinguish accent type variations. Furthermore, the results suggest that comprehensibility of the speech provides cues for accent classification.

7. Conclusion

Our goal in this study has been to provide perceptual assessment of accent variations for accent identification applications. This is important in order to establish benchmark human performance to compare with automatic systems. The results showed that listeners' accent background impacts accent detection and classification accuracy. Longer context (single words vs. phrases) was also shown to contribute to higher accuracy for all three listener groups in the case of detection, and for familiar (British) listeners in the case of classification. It is also observed that better comprehensibility of the speech does not lead to higher accuracy in accent detection and classification. However, listeners may be using the level of comprehensibility as an indicator of their familiarity with a particular accent, which affect accent perception. Furthermore, some accents are indicated to be perceptually more distinct and identifiable than others regardless of the listeners' accent background and comprehensibility of the speech. This suggests that the listeners are relying on different types of cues to understand the speech and recognize its accent type, but comprehensibility of the speech also affects accent perception. These observations point to the importance of understanding cognitive aspects of accent variation, which will contribute to further development of speech technology, including automatic accent identification and speech recognition.

REFERENCES

[1] Arslan, L., Hansen, J.H.L (1996). Language Accent Classification in American English, *Speech Communications*,18(4), pp.353-367.

[2] Yanguas, L.R., O'Leary, G.C., Zissman, M.A., Incorporating Linguistic Knowledge into Automatic Dialect Identification of Spanish, *ICSLP-98.*

[3] Crystal D. (2003). A dictionary of linguistics and phonetics. Malden, MA: Blackwell Pub.

[4] Bent, T., Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Acoust. Soc. of America, 114(3),1600-1610.*

[5] Flege, J.E. (1988). Factors affecting degree of perceived foreign accent in English sentences. J. Acoust. Soc. Amer., v.84, p.70-77.

[6] Jilka M. (2000).The contribution of intonation to the perception of foreign accent. Stuttgart:PhD thesis, University of Stuttgart.

[7] Megan H.S. (1998) The perception of foreign-accented speech, Journal of Phonetics, vol.26, pp.281-400.

[8] Munro M.J., Derwing T.M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. Language Learning vol.49, pp.285-310(26).

[9] Angkititrakul P. and Hansen J.H.L. (2004). Advances in Phone-Based Modeling for Automatic Accent Classification. Accepted to *IEEE Trans. Speech & Audio Proc., March 2004.*

[10] Grabe, E., Post, B and Nolan, F. (2001). The IViE Corpus. Department of Linguistics, University of Cambridge.

[11] Lawson A.D., Harris D.M. and Grieco J.J. (2003). Effect of foreign accent on speech recognition in the NATO N-4 corpus, *Interspeech/Eurospeech-2003 pg.1505-1508, Geneva, Swiss.*