ACCENT CATEGORISATION BY LAY LISTENERS: WHICH TYPE OF "NATIVE EAR" WORKS BETTER?

CHEN SHEN[1], DOMINIC WATT[2]
University of Sheffield[1], University of York[2]

Abstract

Listeners, be they lay or expert, can to a greater or lesser extent distinguish and correctly identify different accents of familiar languages. This ability plays a central role across a spectrum of speech perception-based activities, not the least of which are speaker profiling and comparison of the sort carried out for forensic purposes. Factors that affect listeners' abilities to identify different accents, such as their native linguistic competence and their level of familiarity with the target variety, have been reported extensively in the literature. However, we currently lack data from studies that make direct comparisons of listeners' abilities to correctly categorise foreign-accented varieties of a language they speak natively (L1 listeners) against the competence of non-native speakers of that language (L2 listeners), where the aim of the listening task is to categorise familiar and unfamiliar foreign accents of the language. The present study is designed to investigate these issues by looking at how well a listener's "native ear" allows him or her to judge the origins of non-native speakers of English. By "native ear", we mean in the present case the sort of perceptual acuity that we might associate with being an L1 English speaker versus an L1 Chinese speaker who speaks English as an L2, in the situation where speakers of these two sorts are exposed to recordings of Chinese-accented English (CE). For the experiment reported in this paper, the speakers recruited were L2 English speakers from East Asia (Japan, South Korea, and three regions of China), with the principal interest being in those in the Chinese group. A total of 42 listeners (monolingual L1 English speakers vs. Chinese L2 English speakers) completed a dual-task forced-choice experiment in which they were exposed to samples of English spoken by the East Asian talkers. The results showed that the L1 Chinese listeners were significantly better than the L1 English listeners at distinguishing the CE speakers from the Japanese and South Korean ones. However, neither of the two types of "native ear" was better than the other (or indeed in absolute terms) at correctly categorising the three subvarieties of CE according to three broadly-defined geographical/dialect regions (the North, South, and West of China). We conclude by discussing the relevance of our results to forensic cases in which an analogous kind of categorisation task might be necessitated.

1. Introduction

Variability in speech production has numerous sources: it derives at the most basic level from individual differences in speakers' anatomy and physiology (Nolan 2012), but also results from speakers' social characteristics. Place of origin, gender, age, social status, sexual orientation, and a host of other factors related to regional and social identity are all encoded to some degree in the speech signal, and listeners are generally highly skilled at detecting and reacting to features of a talker's speech that are associated with these non-linguistic parameters (for overviews of research in this area, see Foulkes et al. 2010 and Baranowski 2013).

The language being spoken is of course also relevant. A number of studies (e.g. Ladefoged & Ladefoged, 1980; Thompson, 1987; Goggin et al., 1991; Schiller & Köster, 1996) have looked at how the language(s) being used by talkers, and those that are spoken natively or non-natively by listeners, have an effect on how well listeners perform in perception experiments. Listeners, who could be monolingual, bilingual or multilingual, might be exposed to their native

language(s) spoken in the listeners' own accents, or in different ones (regional, social, non-native). Alternatively, the stimuli that subjects hear could be spoken in a different language altogether. Differences in listeners' performance may occur according to whether they have prior knowledge of the target language(s) (Goggin et al., 1991; Köster et al., 1995; Köster & Schiller, 1997). In forensically-relevant circumstances—for example, where an earwitness overhears criminals talking at the scene of a robbery—a listener might claim that s/he recognised the voice of a person who was speaking in the listener's native language, but with a heavy native or non-native accent. Conceivably, the witness might be asked to identify a suspect in a voice parade on the basis of a sample of speech in a language that the witness speaks non-natively, but where the speakers' accents are familiar or unfamiliar (say, where an L2 English speaker of Chinese origin is asked to judge whether s/he has previously heard the voice of a French person speaking English). It is currently unclear whether being a native speaker of a language (e.g. British English) versus a speaker of a non-native variety of that language (e.g. Chinese-accented English, CE) better equips listeners to categorise different sorts of accented speech correctly. In the present case, does being a "native speaker" of CE mean that one is better able to identify fellow CE speakers by their accents, or does being a native speaker of English (in the conventional sense of the term) mean that one is likely to be better than a CE speaker at distinguishing accurately between CE speakers and L2 English speakers with other East Asian accents? Moreover, are CE "native speakers" better at telling which region of China other CE speakers come from than are monolingual L1 English speakers? Put another way, which kind of "native ear" performs better at listening tasks of this kind?

Two speech perception experiments were run so as to address these questions. In the first experiment, listeners with different linguistic backgrounds (L1 English vs. L1 Chinese) were asked to categorise the accents of a variety of female East Asian speakers (native speakers of Chinese, Japanese, and Korean) reading an English text aloud. In the second experiment, the listeners were instructed to classify the CE speakers according to the region of China (North, South, West) that they believed the speakers to come from. We hypothesised that, by virtue of their being native speakers of the language of the experimental stimuli, the L1 English speakers would be as good or better at distinguishing CE speakers from Japanese or Korean ones than the CE-speaking listeners were. Conversely, we hypothesised that the CE-speaking listeners would be better able to classify the CE speakers as Northern, Southern or Western than would the L1 English speakers.

Before further details of the experimental design and hypotheses are discussed, however, we assess the context of the study.

2. Background

2.1. Identifying accents and dialects

The diversity of accents and dialects of English around the world is a phenomenon that scarcely needs any introduction, and by comparison with all other languages the synchronic and diachronic variation to be found in English is easily the most thoroughly documented. Although the properties of dialects and accents of English and other major languages are well described in terms of their production, less well understood is how listeners perceive and process the phonetic variation to which they are exposed. These aspects of the speech perception faculty are of central concern in both sociophonetics and forensic phonetics, two areas of linguistic inquiry in which it is impossible to sideline the very high levels of variability in speech

produced under normal conditions. To date, however, the vast proportion of work on speech perception has dwelt upon how listeners respond to speech stimuli under controlled laboratory conditions, rather than upon how listeners deal with speech signals in day-to-day communicative situations.

It should also be recognised that there is a good deal of variation between individuals with respect to their abilities to identify accents/dialects, or indeed individual languages. In attempting to account for these perceptual disparities, many factors should be considered: these include listeners' geographical and language backgrounds, their familiarity with the language or accent in question, and the level of linguistic expertise that listeners possess. A significant quantity of research with a focus on listeners' abilities to identify speakers based on language and accent under varying circumstances has to date been carried out. Studies of naïve listeners' abilities to identify dialects have been conducted by, for instance, Preston (1993), Purnell et al. (1999), Williams et al. (1999), and Clopper and Pisoni (2004a, b), while Köster et al. (1995) and Köster and Schiller (1997) have published findings on earwitnesses' speaker- and speech-identification abilities. Research questions in another field that centres on accent identification, Language Analysis for the Determination of Origin (LADO), have also been receiving attention from a number of scholars (Fraser, 2009; Cambier-Langeveld, 2010; Wilson and Foulkes, 2014).

2.1.1. Identifying accents and dialects in sociolinguistic settings

In Preston (1993), untrained listeners from two US states (Michigan and Indiana) were asked to assign samples of the speech of unknown talkers to one of the nine cities lying on a north-south line between Saginaw, Michigan and Dothan, Alabama. The results showed that although the listeners did not necessarily categorise speakers by city very accurately, they were still able to make reliable distinctions between dialect groups using the broader categories "North" and "South". Clopper and Pisoni (2004b), who asked their participants to categorise a set of speakers according to six US dialect regions in a forced-choice task, found that their listeners chose correctly only 30% of the time. However, a *post hoc* analysis of their data revealed that, like Preston's subjects, listeners were making correct distinctions between Western, Southern and New England speakers, rather than assigning them to these broader categories at random.

Purnell et al. (1999) demonstrate untrained listeners' skills at identifying a speaker's dialect using a matched-guise design. For their study, a tridialectal male speaker (John Baugh) capable of speaking African American Vernacular English (AAVE), Chicano English and Standard American English left answering-machine messages directed at landlords inquiring about renting apartments in a selection of localities in the San Francisco Bay Area. Purnell et al. argued, on the basis of how the landlords chose to respond to the inquiries, that the landlords had identified the caller's dialect based on relatively short samples of telephone speech. By telling the Chicano- and AAVE-guise callers that the accommodation was no longer available much more often than they told this to the Standard American English-guise caller, the landlords appeared to have discriminated against speakers of the two non-standard dialects. An additional experiment using native speakers of the three target varieties also showed that other naïve listeners (university students) correctly categorised the three dialects significantly better than chance.

These results, and those of Preston and Clopper & Pisoni, reinforce the claim that untrained listeners are able to recognise dialects and accents in a systematic way.

2.1.2. Identifying accents and dialects in forensic settings

Multilingualism in countries such as the UK means that, in situations of forensic relevance, lay and expert witnesses often have to try to identify a talker who speaks a language or dialect different from their own. Accent features are regularly used by both naïve and expert listeners in criminal investigations and trials (Foulkes & French, 2012).

Studies conducted by French (1990), Ellis (1994), and others reported in Foulkes & French (2001, 2012) have demonstrated the importance of accent identification in a variety of forensic cases. When attempting to resolve the words being spoken in poor-quality recordings, working out whether the speaker used a non-standard or foreign accent, and identifying which accent it was, may be of crucial importance to deciphering content. In speaker profiling cases, in which the police have yet to apprehend a suspect, accent cues allow expert listeners to glean some regional, social and ethnic background information about the speaker in a recording so as to help narrow down the field of potential offenders. Naïve listeners may be called to courtrooms to give testimony relating to the voice of a wrongdoer, and may offer their judgments about the regional and social background of an offender based on impressions of the accent they heard the offender using (Jessen, 2008). Under such circumstances, it is critical to be able to estimate how well naïve listeners can identify different accents, and the extent to which their judgments can thereby be relied upon.

2.1.3. Dialect and accent identification in LADO fieldwork

An area in which the ability to correctly and consistently categorise different accents of a language is of particular importance is that of language analysis for the determination of origin (LADO). Linguistic analysis of an asylum claimant's speech and language patterns may be carried out by the immigration authorities where there are grounds for scepticism over the claimant's purported place of origin. LADO professionals may attempt to verify whether an asylum seeker is genuinely from the region that he or she claims to be from, and/or may (as per speaker profiling) try to identify the claimant's actual place of origin or of socialisation (Patrick, 2012; Wilson & Foulkes, 2014).

The issue of whether native speakers of the language in question should give input during the LADO process has proven over the last decade or so to be a matter of especially vigorous dispute. Cambier-Langeveld & Samson (2007) mention the contributions a "native ear" can make in LADO work. They contend that the language competence and experience of a native speaker is required when attempting to ascertain whether the claimant's speech exemplifies the linguistic structures used by genuine speakers from the claimed community of origin. Cambier-Langeveld & Samson do, however, stress that native speakers involved in the analysis should be supervised by trained linguists. In the first example of LADO-oriented experimental research, Wilson (2009) investigated the performance of four listener groups (native speakers of Ghanaian English with no linguistic training; British undergraduate linguistics students; phoneticians (academics and postgraduate students with experience of forensic phonetics); and practicing LADO analysts), in a speech perception task differentiating Ghanaian English from the fairly similar Nigerian English. The results of her study show that native Ghanaians performed the task with the highest level of accuracy (86%), and of the four listener groups reported the highest level of self-confidence in their own judgments (see also Foulkes & Wilson, 2011).

Native speakers, then, have been shown empirically to be very competent at distinguishing their own accent variety from other, similar ones. By testing the extent to which a "native ear"

makes one better able to identify linguistic variety(s), the value of conducting dialect and accent identification tasks in the LADO context with native speaker involvement is made clearer. Although a consensus appears to be emerging regarding the precise role that native speakers ought to fulfil in the LADO process (Eades et al., 2003; Fraser 2009, 2011; Wilson, 2009; Cambier-Langeveld, 2010; Wilson and Foulkes, 2014), so far little attention has been paid to how asylum seekers themselves might be able to corroborate the accounts they offer to the authorities by demonstrating their competence at categorising regional and social varieties of the language(s) they claim to speak natively.

2.2. Native language competence vs. accent familiarity

The studies listed in §2.1 argue strongly for the critical role that the language(s) or dialect(s) spoken by speakers and listeners play in speech and accent identification tasks. To gain a better understanding of the effects that native language competence has on accent categorisation accuracy, we first need to clarify what "native language competence" actually refers to. Davies (2004:438) defines a native speaker as one who "owns" the language by dint of birth and "by virtue of being a native user", and a native-like speaker as an "an exceptional learner" who attains a level of communicative competence comparable to that of a native speaker by birth. Davies (2004:433) also points out that "a native speaker is expected to "know" another native speaker in part via intuition and partially because the individuals in question make common use of a characteristic and systematic set of linguistic indicators grounded in shared cultural knowledge. The amount of experience a listener has had with a linguistic variety through contact with one or more native speakers of that variety is a principal determinant of how readily the listener can distinguish between the target variety and some other.

Fraser (2011:126) makes the following generalisation in respect of this perceptual faculty:

> Native speakers are, under certain circumstances and within certain limits, good at identifying fellow native speakers of their own language variety, just as no one doubts that earwitnesses are, under certain circumstances and within certain limits, good at identifying voices of speakers they know.

Although the boundary between the definitions of "native speaker" is not necessarily a sharp one, for our present purposes we will make a simple binary distinction between people who speak a language as their L1 as native speakers, and those who speak a non-native variety of the language as their L2. The latter group can in a sense be thought of as "native speakers" of their non-native variety. While a number of studies have demonstrated the capacities of different listeners to recognise their native languages and dialects and/or those languages and dialects that they are familiar with, to our knowledge no study has yet asked whether native speakers of a language can identify foreign-accented variants of their L1 better, equally well, or worse than "native speakers" of the non-native variety of the language. In the following sections, we review studies that have investigated native language competence, on the one hand, and language variety familiarity, on the other.

2.2.1. Studies of native language competence

Several studies of voice identification suggest that knowledge of the target language influences speech recognition results. In one of the earliest inquiries of this type, Bush (1967) asked untrained American English-speaking listeners to identify the countries of origin of American, British and Indian speakers on the basis of a set of samples of the speech of these three groups. The findings of Bush's study show that L1 English-speaking listeners could identify the

speakers' nationalities with over 90% accuracy in a three-alternative forced-choice categorisation task using real words, nonsense words, and sentences. She also found that the Indian (L2 English) speakers were most reliably identified by all the listener groups, probably due to the salient foreign-accented features of Indian English. Similarly, Goldstein et al. (1981) report an experiment in which American English-speaking listeners were able to correctly identify the ethnicity of white American, black American, and Taiwanese English speakers approximately equally well (85%, 82%, and 81%, respectively).

A more recent study by Perrachione et al. (2009) explored native versus non-native listeners' abilities when they were asked to identify individual talkers. They compared L1 English and L2 English (L1 Mandarin Chinese) listeners' talker identification accuracy, and found that both listener groups were better at identifying voices of talkers speaking the listeners' mother tongues. They also note that although the L1 Chinese listeners all had functional English language skills, being students or university researchers who had lived in the US for some time, they were still outperformed by the L1 English-speaking participants when asked to identify individual voices speaking in English. These results show the advantage of being a native speaker over a non-native speaker when identifying individual talkers, despite the non-native speakers' functional language skills in the test language.

Schlichting & Sullivan (1997) and Neuhauser & Simpson (2007) gauged listeners' native linguistic competence by asking participants to attempt to distinguish between genuine and imitated speech in their mother tongues. More specifically, Schlichting & Sullivan (2007) used a voice parade experiment to show that L1 Swedish listeners could distinguish between the authentic voice of a well-known Swedish politician and a competent imitation of it by a professional impersonator. Similarly, in Neuhauser & Simpson's (2007) study, L1 German-speaking listeners were able to identify imitated accents of their mother tongue. These findings reinforce the proposal that native speakers possess competence in recognising and discriminating voices spoken in their L1.

2.2.2. Studies of accent familiarity

Dialectologists have examined the effect of accent familiarity on listeners' perceptions of subvarieties of languages. One basic assumption has been that listeners who have more familiarity with variety X should be better at identifying it than are listeners with a lower level of familiarity with the variety (Kerswill & Williams, 2002; Clopper & Pisoni, 2006).

In one study investigating dialect categorisation, Williams et al. (1999) recorded two L1 English speakers from each of six regions of Wales, and two L1 speakers of British Received Pronunciation (RP), narrating personal stories. They then conducted a dialect perception experiment, playing short segments of the recordings to different groups of listeners from each of the same six Welsh regions, and asking them to categorise each talker according to one of eight categories (viz., the six regions of Wales, RP, and "don't know"). They found that although the overall accuracy of the listeners' responses was not high (30%), the performance of each listener group when classifying the two speakers from their own Welsh variety was substantially better (45%). This supports Preston's (1993) findings, which showed that listeners were better at identifying accent varieties from regions close to their own home regions than accents from further away.

Preston (1993) shows, furthermore, that the amount of exposure listeners have had to a target linguistic variety also influences their accent identification abilities. Preston conducted a perceptual dialectology experiment using non-native listeners, the results of which demonstrate that the listeners who had greater experience of the target accent were better at identifying the

dialect than were those with less experience. Clopper & Pisoni's (2004a) study of two groups of individuals (one made up of people who had either stayed in their home state all their lives – the "homebodies" – and the other of people from armed forces families – the "army brats") revealed that listeners from the second group, who had lived in various parts of the US, were better at accurately identifying different varieties of American English than were those from the first group.

2.3. Research questions and aims

The aforementioned studies demonstrate how variable individuals are with respect to their abilities in accent categorisation, and show that native language competence and level of familiarity with a subvariety of a language spoken non-natively both affect listeners' accent categorisation capabilities. However, because little research with a focus on direct comparison between these two types of linguistic competence has yet been carried out, it is not altogether straightforward to formulate hypotheses concerning which sort of competence confers the greater advantage when it comes to accurate accent categorisation of the kind investigated in the present study.

We therefore address the latter issue experimentally, by collecting responses from L1 and L2 English-speaking listeners using a forced-choice accent categorisation task. The research questions motivating the study are as follows. How accurately can listeners with these differing linguistic backgrounds identify and correctly label L2 accents of English? Which group performs better at this task? Since in earlier studies listeners were shown to be competent at identifying accents similar to their own – or otherwise ones that they are familiar with – is it the case that our L2 English listeners (Chinese nationals) will be better able than the L1 listeners to classify CE-accented English by the region of origin of the speaker?

We considered it reasonable to suppose that the L1 English-speaking listeners would be no less competent at accent categorisation than the CE-speaking listeners when exposed to recordings of Chinese, Japanese and South Korean speakers talking in English. However, we hypothesised that the CE-speaking listeners would perform better than the English-speaking listeners at the second task (Chinese regional accent categorisation) because of their higher sensitivity to the phonetic cues associated with a talker coming from the north, south or west of China.

3. Methodology

A sequence of two forced-choice perception experiments were conducted (a) to explore the effects on accent categorisation accuracy of native and non-native competence in English, and (b) to test for the effects of familiarity with Chinese-accented English.

3.1. Speech materials

Samples of the speech of 25 speakers were used for the perception experiment. This set was comprised of recordings of 15 speakers of Chinese-accented English. These speakers were further divided equally into three broad Chinese geographical regions in the second task (see §3.1.1). The other speakers were 5 Japanese-accented and 5 Korean-accented L2 English speakers (henceforth labelled the JE and KE groups, respectively). The majority (n = 18) of the speech samples (15 CE, 1 JE and 2 KE) were obtained from an online corpus (International

Dialects of English Archive; IDEA, 1997)[1], while the remaining 7 samples (4 JE and 3 KE, code-named Japan-R1, -R2, -R3, -R4 and Korea-R1, -R2, -R3, respectively) were recorded by the first author specifically for this study.

So as to control for variation correlating with speaker gender, age, and education level and so forth, only samples of the speech of young educated female speakers of English as a second language were used for the experiment.

The speech samples chosen were a mixture of read and spontaneous speech. For the read speech, a short section of "Comma Gets a Cure", a phonetically-balanced text passage containing Wells' (1982) lexical set keywords (Honorof et al., 2000), was used. This allowed the examination of the speakers' pronunciations in a fixed set of contexts. The spontaneous speech, by contrast, is mostly in the form of narratives in which the speakers described their hometown and life experience. This material was used so as to provide more naturalistic speech which would potentially contain a greater number of salient idiosyncratic accent features. Spontaneous speech was also thought to serve the interests of forensic realism better than read speech would have done.

The reason for choosing CE as the target language is that L2 CE speakers account for a large proportion of the world's L2 English speakers and learners (Kirkpatrick, 2007). CE can be heard in every urban area in the UK (Gye, 2014). JE and KE speakers were used as foils mainly because, although Chinese, Japanese and Korean are very distinct from one another, they nonetheless share a certain number of linguistic properties. Owing to the relatively short geographic distances involved and the intimate historical contacts between the three countries, cross-linguistic transfer of features at different linguistic levels (not least at the phonetic and phonological levels) has taken place over a period of many centuries (Curnow, 2001). It is understandable, therefore, if listeners confuse one of the three accents with the other two, even if they are fairly familiar with what CE, JE and KE tend to sound like.

3.1.1. Characteristics of the IDEA samples

The 15 CE samples selected from the IDEA archive were all recorded in the Chinese city of Suzhou. All of the speakers were students at Suzhou University at the time they were recorded, but they originated from different parts of China. They all learned English principally from CE teachers at school, and had had limited prior access to native English speakers. Their CE accent features could therefore be described as fairly marked. The CE speakers were selected with respect to the broad geographical/dialect divisions chosen for the experiment, namely Western, Northern, and Southern. Accents of Mandarin Chinese are quite distinct across these broad dialect zones (Ramsey, 1987), and it was considered plausible that these differences in L1 pronunciation would be reflected in the speakers' pronunciation of English (Major, 2001). Table 1 shows the files chosen from the IDEA archive.

[1] http://www.dialectsarchive.com/ (accessed 20th April 2015).

Western	Xinjiang Province	Sichuan Province	Gansu Province
	Xinjiang 1 *Xinjiang 5*	*Sichuan 2* *Sichuan 3*	*Gansu 3*
Northern	Liaoning Province	Heilongjiang Province	Hebei Province
	Liaoning 1 *Liaoning 3*	*Heilongjiang 2* *Heilongjiang 4*	*Hebei 3*
Southern	Zhejiang Province	Guangdong Province	Jiangsu Province
	Zhejiang1 *Zhejiang 3*	*Guangdong 2* *Guangdong 3*	*Jiangsu 41*

Table 1: Details of CE files selected from the IDEA archive. Province names followed by numbers (e.g. *Xinjiang 1*) denote the individual IDEA file chosen for inclusion in the experiment.

The one JE and two KE samples chosen from IDEA were Japan 5, South Korea 3, and South Korea 5. These three speakers had resided in English-speaking countries for various lengths of time during their adult lives, but their spoken English is noticeably foreign-accented, as we would expect of individuals who have learned a second language in adulthood (Major, 2001).

3.1.2. Characteristics of the recorded speech samples

Because IDEA does not contain a sufficient number of JE and KE samples to be used in the present study, an additional 4 JE and 3 KE samples were recorded by the first author following the protocol used by IDEA so as to make sure all the samples were of a uniform character. The recordings were made in a sound-treated recording facility at the University of York using a Neumann U87i P48 condenser microphone fed into a TAC Scorpion 16-8-2 mixing desk. Audio capture was via an M-Audio 24/96 soundcard and Adobe Audition CS5.5 software. The sampling rate of the recordings was 44,100Hz. The speakers were all postgraduate university students. Although their length of residence in an English-speaking country varied, the English of all seven subjects was still markedly accented.

3.2 Pilot Study

A pilot study was conducted prior to running the actual experiment so as to examine the workability and effectiveness of the experimental design. 5 participants (3 Chinese and 2 English speakers) completed the pilot study. A few minor amendments were made to the design to ensure that the materials (PowerPoint slides and audio recordings) were displayed in the most suitable way. Feedback from participants also suggested that the length of each speech sample (c. 15 seconds), plus a 3-second break between each sample, was enough to allow them to make a judgment, and that the duration of the experiment (approximately 35 minutes) was acceptable to them.

3.3. Listeners

A total of 45 participants (22 native Chinese and 23 native English; mean age = 24, SD = 2) was recruited from the student body at the University of York. None of the participants reported any speech or hearing disorders. All L1 English-speaking listeners claimed to have had exposure to CE before, although their amount of experience varied. All of the L1 Chinese-speaking listeners had attained a sufficient level of ability in English to be able to pursue degree-level studies in the UK. No other requirements, e.g. gender or region of origin, were considered at the recruitment stage. Participants' linguistic backgrounds (including whether they had had training in linguistics and phonetics), the English listeners' self-rated familiarity with CE, and the Chinese listeners' region of origin in China, were recorded for post hoc analysis, however.

3.4. Listening tasks

The 25 speech samples were edited into 50 short excerpts based on speech style (i.e. 25 samples apiece of spontaneous and read speech). The edited read-speech excerpts each contain 44 words and are 14-16 seconds in length after disfluencies such as pauses greater than 100ms in length, repetitions, and self-corrections were removed. The spontaneous speech was also edited into 14-16 second long excerpts, with disfluencies and all identifying information removed. This sample length was chosen such that there would be enough material for a range of accent cues to be available to listeners, while keeping the experiment acceptably short. The final stimuli were normalised for amplitude using Sony SoundForge v. 10.0 so as ensure that all audio stimuli were equal in volume when presented to the listeners.

Instructions on how to complete the two listening tasks were displayed as PowerPoint slides on a projector screen, and were simultaneously talked through by the researcher. Each of the speech samples was then presented via its own PowerPoint slide which also showed an arbitrary speaker and test block number (see further below). The slides and recordings were set to play automatically, with a 3-second gap between each speaker, and a 5-second break accompanied by a notification tone at the start of each block. Participants were also encouraged not to think too hard about their judgments and to make their decisions quickly. The experiment was composed of the tasks described in the following sections.

3.4.1. Task 1

Task 1 was the CE identification task. The 50 edited speech excerpts were divided into 10 blocks containing samples for 5 speakers per block, following the suggestion of Clopper et al. (2011). The participants were informed that within each block the speech samples represented a mixture of CE, JE and KE, and that there would be at least one CE, one JE, and one KE sample in each block, to prevent (for example) participants from simply labelling all five samples per block as CE. Samples were played in randomised order so as to eliminate any ordering effects. The first five blocks were read speech (part 1), with the remaining five blocks being spontaneous speech (part 2).

Listeners were given paper answer sheets on which they were asked to circle the number corresponding to the speaker(s) in each block that they thought were Chinese. They were also asked to state, on a scale from 1 to 5, their level of confidence concerning their decisions about their chosen speaker(s) being Chinese and the unchosen one(s) being non-Chinese (here, 1 indicates "not at all confident" while 5 means "very confident"). In addition, participants were

encouraged to make notes in spaces provided on the response sheet, and at the end of the task were asked to provide comments about how they made their choices. There was a practice block before both parts 1 and 2 to familiarise participants with the task.

3.4.2. Task 2

Task 2 was the CE categorisation task. This was designed to test whether listeners could further subcategorise CE speakers according to three broadly-defined Chinese geographical/dialect regions (North, South, and West), based on their L2 English speech. The 30 previously edited CE speech excerpts from Task 1 were divided into 6 blocks of 5 speakers, incorporating a mixture of speakers from the three regions. The samples were presented in randomised order, with the first three blocks being read speech (part 1), and the last three being spontaneous speech (part 2).

By way of giving the participants some training in hearing differences between Chinese-accented English from the different regions of China, three short (10-second) speech samples each containing up to one read sentence spoken by all 5 speakers from each region were played to both the L1 English and the L1 Chinese listeners prior to the categorisation task itself. Participants were also given a map showing the Chinese regions in question. Each sample was repeated twice, with a 2-second pause in between.

Participants were asked to sort the speakers according to the three regions by writing the initial letter of that region ("N" for North, "S" for South and "W" for West) for each speaker on their response sheets. Participants were allowed to refer to the map and the notes they made during the training session if they wished. They were again asked to rate their confidence in making their decisions following the same method used in Task 1, and were encouraged to make notes concerning the specific province they believed the speaker to originate from, if they felt they could discriminate more finely than the three broad geographical zones. No practice block was thought necessary for Task 2, as the task was very similar in nature to Task 1.

Participants were given a short break between Tasks 1 and 2, which they used to fill in a linguistic background questionnaire. No feedback about the accuracy of their responses was provided during any phase of the experiment.

4. Results

4.1. Overall performance

Answers for three of the 45 participants were excluded from the final data analysis due to their having made irreversible mistakes in their questionnaires: two did not select the CE speakers in Task 1, and the other one wrote the wrong initial letter for one of the three regions in Task 2. Overall, 3,360 valid responses (2,100 from Task 1 and 1,260 from Task 2), together with the corresponding self-rated confidence scores (1-5), were collected from the remaining 42 questionnaires. 20 of these were completed by L1 Chinese listeners, with the rest being those for the L1 English listeners. 6 of 2,100 responses in Task 1 were then eliminated because two listeners identified one of the speakers as individuals that they knew in person. All of the answers were then coded according to a binary distinction between correct (correctly categorised CE and non-CE in Task 1, and correctly identified region in Task 2) and incorrect. The overall accuracy of the performance of the two listener groups, where accuracy is defined

as the percentage of correct answers among all the responses in the two tasks, is shown in Figure 1.

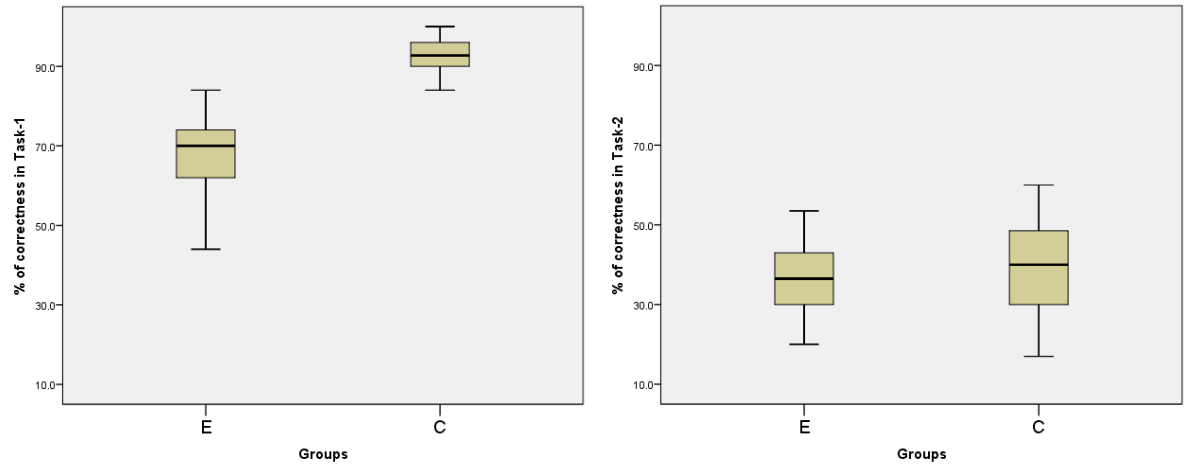

Figure 1: Overall performance of the L1 English (E) and L1 Chinese (C) listener groups in Tasks 1 (left panel) and 2 (right panel).

Initial observation of these results reveals that in Task 1, the L1 Chinese listener group (C) performed much better than the L1 English group (E). However, the performance of the C and E groups in Task 2 was rather similar, and was markedly worse than their performance in Task 1, showing that both listener groups could distinguish CE from JE and KE better than they could categorise CE according to the three Chinese geographical/dialect regions.

Overall performance in Task 1 is well above chance level (50%) for the L1 English listeners ($\bar{x}$ = 68.2%), and near ceiling for the L1 Chinese listeners ($\bar{x}$ = 92.9%). In Task 2, by contrast, the listeners' performance is barely above chance (33.3%) among the L1 English listeners ($\bar{x}$ = 34.9%), and only slightly higher among the L1 Chinese listeners ($\bar{x}$ = 39.4%). The data (the distribution of which is found to be normal using the Shapiro-Wilk W test) were subjected to a two-tailed *t*-test, revealing a significant difference between the two listener groups in Task 1 ($t(28.4) = 10.7$, $p < .001$), but no significant effect in Task 2 ($t(40) = 1.34$, $p = .194$).

4.2. Effects of different factors on performance

4.2.1. Linguistic/phonetic training

To investigate the effect of linguistic/phonetic training to at least first-degree level on listeners' ability to categorise accents, the two listener groups were each split into two, yielding a total of four groups: trained English (ET), untrained English (EU), trained Chinese (CT), and untrained Chinese (CU).

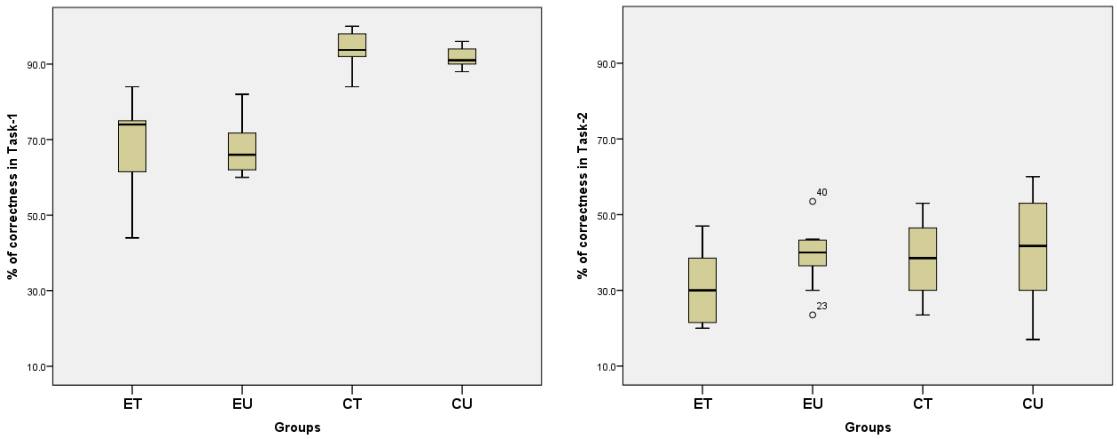

Figure 2: Overall accent categorisation performance across the four listener groups in Tasks 1 (left panel) and 2 (right panel). ET = linguistically trained English listeners; EU = untrained English listeners; CT = trained Chinese listeners; CU = untrained Chinese listeners.

The results of ANOVA testing show a highly significant effect of listener group in Task 1 ($F(3,38) = 34.259$, $p < .001$), but no such effect in Task 2 ($F(3,38) = 1.654$, $p = .193$). *Post hoc* tests reveal that the significant effect of group in Task 1 is dependent on the native language of the listener (ET ~ CT, $p < .001$; EU ~ CU, $p < .001$), rather than on training (ET ~ EU, $p = .570$; CT ~ CU, $p \approx 1$). These results indicate that the linguistic background of listeners was the main factor predicting their CE categorisation accuracy, while training in linguistics/phonetics played a less significant role in the same task.

To examine the performance of the four listener groups in more detail, consider the descriptive statistics shown in Table 2.

Group	Mean %	SE % & SD %		Min %	Max %	IQR
ET	68.3	3.81	12.6	44	84	18
EU	68.1	2.08	6.91	60	82	11.5
CT	94	1.52	4.81	84	100	6.5
CU	91.8	0.96	3.05	88	96	5

Group	Mean %	SE % & SD %		Min %	Max %	IQR
ET	30.9	2.93	9.73	20	47	20
EU	38.8	2.35	7.8	24	54	7
CT	38.4	3.13	9.9	24	53	18
CU	40.4	4.65	14.7	17	60	26

Table 2: Descriptive statistics relating to the accent categorisation proficiency of the four listener groups in Task 1 (upper panel), and Task 2 (lower panel). SE = standard error; SD = standard deviation; Min = minimum; Max = maximum; IQR = interquartile range.

It can be seen in Table 2 that the CT listener group has the highest mean (94%) in Task 1, closely followed by CU (91.8%). Both these listener groups' low SD and IQR values indicate that their within-group performances are more consistent than those of the ET and EU groups. In Task 2, although CU shows the highest mean (40.4%), its SD (14.7%) and IQR (26) are also the highest. This might stem from the fact that listeners' performance in this task is not very

consistent. CT and EU have similar means (38.8% and 38.4% respectively), with EU returning a smaller SD (7.8%) and IQR (7). The ET group displays the lowest mean (30.9%), which is below chance level.

4.2.2. Speech style

To investigate whether speech style plays a role in the categorisation task, the responses of the two listener groups subdivided by speech style in the two tasks were examined. The results are shown in Figure 3.

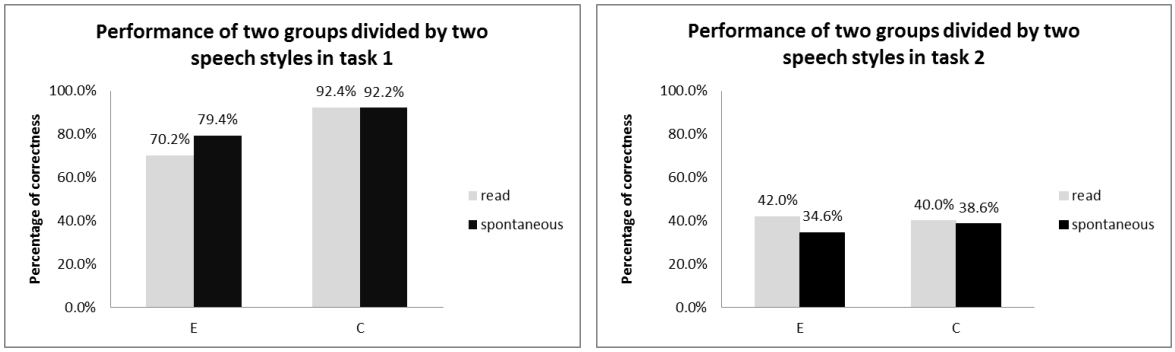

Figure 3: The performance of the two listener groups according to speech style in Tasks 1 and 2 (E = L1 English listeners; C = L1 Chinese listeners).

Although in Task 1 there seems to be a trend whereby L1 English-speaking participants perform better when listening to spontaneous speech than they do when they are exposed to read-speech stimuli, a chi-square test shows that the difference is not statistically significant ($\chi^2(1) = 1.5$, $p = .203$). Interestingly, in Task 2 an opposite trend is found, but the difference is still not significant ($\chi^2(1) = 0.57$, $p = .450$).

4.2.3. Self-rated confidence and accuracy

As listeners all rated how certain they felt about their responses using a confidence score ranging from 1 to 5, we investigated next whether any correlations could be found between the confidence scores and the accuracy of the listeners' responses in the two tasks. Figure 4 shows the relationships between these variables.

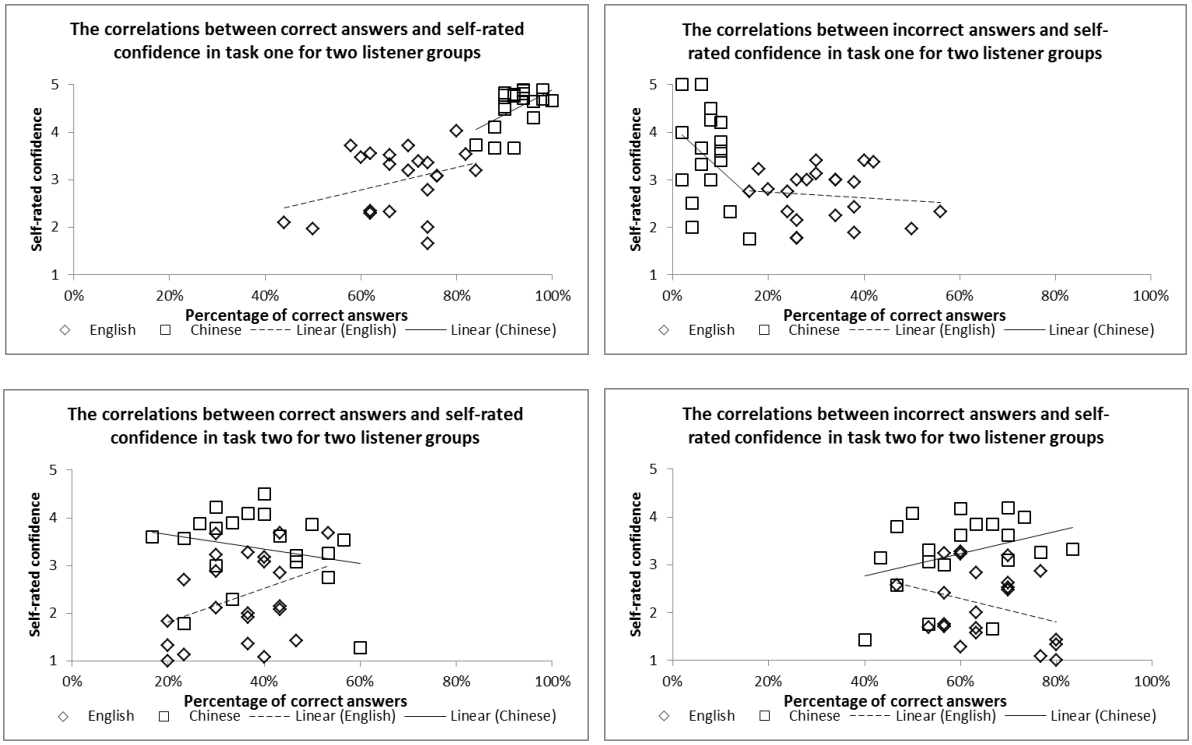

Figure 4: Correlations between the two listener groups' correct/incorrect answers and their self-rated confidence scores in Task 1 (upper panels) and 2 (lower panels).

In order to test the relationships between these variables, a series of linear regression tests (Pearson's *r*) were conducted. In Task 1, the positive correlation between confidence and accuracy is of moderate strength,[2] but is not significant for the L1 English listeners ($r(20) = .340$, $p = .121$), while a stronger, significant correlation is found for the L1 Chinese listeners ($r(18) = .541$, $p = .014$). Confidence ratings for incorrect judgments result in a small, non-significant correlation among the L1 English listeners ($r(20) = .111$, $p = .622$), and a stronger but nonetheless non-significant one for the L1 Chinese listeners ($r(17) = .370$, $p = .119$).

In Task 2, the correlation between confidence and correctness is of medium strength but non-significant for both groups ($r(20) = .365$, $p = .094$ (L1 English), and $r(18) = .298$, $p = .202$ (L1 Chinese)). The same is true of the correlation between confidence and incorrectness (English: $r(20) = .307$, $p = .165$), with a smaller, non-significant correlation for Chinese ($r(18) = .240$, $p = .309$).

4.2.4. Other factors

In order to assess whether L1 English listeners' self-rated familiarity with CE is connected with their ability to categorise the target accents, this group's familiarity and performance scores were tested. A correlation, albeit not a strong one, between the two factors was found ($r(20) = 0.42$, $p = .049$). The performances of some individual listeners reinforced the notion that there is a close relationship between accuracy and familiarity (for further discussion, see §5.0).

To examine whether the L1 Chinese listeners' places of origin influenced their abilities to categorise speakers into Chinese regions based on the speakers' L2 English speech, the 20 L1

[2] The scale used follows Cohen (1988), who suggests the following correspondences: small: r =.10 to .29; medium: r =.30 to .49; large: r =.50 to 1.0.

Chinese listeners were subdivided into three groups based on their residential history in one of the three broadly-defined geographical/dialect regions in China.

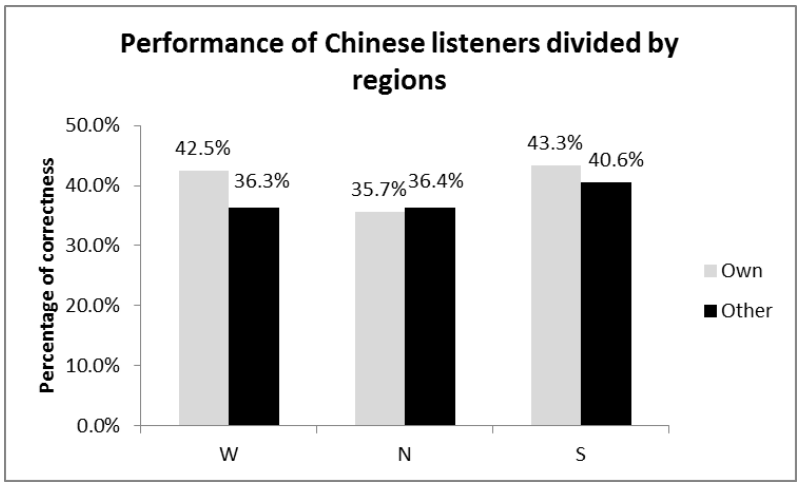

Figure 5: The performance of Chinese listeners split by region (Own = listener and speaker are from the same region; Other = listener and speaker are from different regions). W = West; N = North; S = South.

Figure 5 shows that listeners from the West and South regions of China tend to categorise speakers from their own region better than they classified speakers from the two other regions, although this difference was found not to be statistically significant ($\chi^2(1) = .040$, $p = .841$).

Listeners / Speakers	West (n = 4)	North (n = 7)	South (n = 9)
West	**42.5**	34.3	38.5
North	35.0	**35.7**	43.3
South	37.5	38.5	**43.3**

Table 4: Mean percentage accuracy of Chinese listeners categorising speakers according to three Chinese regions in Task 2.

A breakdown of the respective Task 2 responses of the three listener groups yields more complex findings. Although the Western group's score for correctly classifying fellow Western speakers (42.5%) is much higher than this group's scores for classifying Northern (35.0%) and Southern (37.5%) speakers, the difference is found not to be significant ($\chi^2(1) = 1.632$, $p = .201$). Southern listeners, on the other hand, performed significantly above chance in their classifications of both Southern and Northern speakers ($\chi^2(1) = 4.346$, $p = .037$). It also appears that, with the exception of the Western listener group members themselves, all listeners categorise Southern speakers more accurately than they do Western ones. We discuss these results further in section 5.

It is also the case that there are individual speakers whose accents are most and least correctly categorised by the two listener groups (Tables 5 and 6). The categorisation rates for these speakers lead us to consider the particular set of linguistic and phonetic cues that listeners use when categorising CE and its regional subvarieties, and JE and KE. We explore this issue more fully in §5.2, below.

	Most correctly categorised	Least correctly categorised
L1 English listeners	Liaoning 3 (86%)	Jiangsu 41 (25%)
L1 Chinese listeners	Liaoning 1 (100%) Zhejiang 3 (100%)	Jiangsu 41 (60%)

Table 5: The most and least correctly categorised CE speakers (speakers labelled by IDEA filenames; see §3.1)

	Most correctly categorised	Least correctly categorised
L1 English listeners	Korea-R2 (86%)	Korea-R3 (57%)
L1 Chinese listeners	Japan-R2 (100%) Japan-R3 (100%)	Korea-R3 (80%)

Table 6: The most correctly and incorrectly categorised non-CE speakers (speakers labelled by country of origin plus identifying code number)

5. Discussion

5.1. Listener groups in the two tasks

Overall, both L1 Chinese and L1 English listeners performed well above chance in the CE categorisation task (Task 1), while their accuracy rates fell close to chance level in the CE categorisation task (Task 2). In Task 1, the L1 Chinese listeners performed significantly better than did the L1 English listeners, demonstrating the advantage afforded by having a "native ear" for the target linguistic variety when trying to distinguish between one's own and one or more other non-native accented variety/-ies of an L2 language. In Task 2, however, CE "native ears" were only slightly better than English "native ears", with both groups performing just above chance. This gives an indication of the difficulty of the CE regional categorisation task for both listener groups. We therefore suggest that the different linguistic backgrounds of the two listener groups, together with the somewhat different nature of the two tasks, could account for the disparity in participants' performance between the tasks.

5.1.1. L1 English and L1 Chinese listeners in Task 1

Based on the results obtained from Task 1, the L1 English listeners' overall performance, although well above chance, was still significantly worse than that of the L1 Chinese listeners. One might argue that because English is the first language of the L1 English listeners they would have more richly-specified phonological representations of the many subvarieties of the language they will likely have hitherto been exposed to (Major, 2001). Having access to this kind of "internal catalogue" of accents and dialects would thus make it easier for them to identify and categorise varieties of English than it is for non-native speakers. But as English is typologically very different and geographically very distant from Chinese, and as none of the English participants had learned Chinese or had been exposed to CE speakers for any significant period of time, they might not have a great deal of knowledge about what CE sounds like to start with. In spite of the clear influence that the CE speakers' L1 Chinese accent has on their L2 English pronunciation in the samples used in this experiment, it could nevertheless be

difficult for the L1 English listeners to make correct judgments and generalisations by relying just on the short CE samples they heard during the experiment.

On the other hand, almost all Chinese listeners, regardless of their years of staying and studying in an English-speaking country and their scores for self-rated English fluency, could recognise a CE accent equally well. However, it did appear in their comments that only a few of them correctly identified the other two East Asian varieties as JE and KE. This suggests that there might be an accent prototype for CE with which all Chinese ELF (English as a Foreign Language) learners are familiar, thanks to a shared first language and perhaps quite closely similar life histories. Consequently, the L1 Chinese listeners are more consistently able to identify the fellow CE speakers with whom they have this prototype in common.

The performance of the L1 Chinese listeners was better than those of the L1 English listeners, arguably because the task was easier for the Chinese listeners. Although English is not their first language, the CE the Chinese listeners were asked to identify is the non-native variety of English that they speak natively. Assuming a realistic degree of prior exposure to CE, they could already have developed a detailed understanding of its properties on the phonological level, as well as of lexical and grammatical features typically used by CE speakers (Wilcox, 1978; Major, 2001). Another possible reason for the superior performance of the L1 Chinese listener group is that the instructions for Task 1 did not ask listeners to name the other two non-native English varieties. For the L1 Chinese listeners, the task became one of identifying the CE accent that they were already familiar with, while for the L1 English listeners, not all of whom were familiar with CE beforehand, active comparison of the linguistic properties of the CE and non-CE samples was required. Further research in which listeners are asked to specify what they believe to be the first language of the non-native speakers could provide a useful insight into the abilities of different "native ears" in identification tasks of the present sort. Notwithstanding the fact that the overall performance of L1 English "native ears" was worse than CE "native ears" in Task 1, there was still one English listener who outperformed one of the Chinese listeners. Feedback collected from the questionnaire and a post-experiment interview with this participant reveal that her extensive exposure to CE accents prior to this experiment, and her personal experience with CE, might have been of considerable assistance.

The one Chinese listener who was outperformed by an L1 English speaker was found to have less familiarity with CE, since she had lived and studied outside Mainland China, i.e. in a "non-native" CE speaking environment, for more than 17 years. The amount of exposure she had had to CE during these past years was thus rather limited, and it again illustrates the important roles played in accent identification tasks by adequate exposure to and familiarity with the target linguistic variety.

5.1.2. L1 English and L1 Chinese listeners in Task 2

The question addressed in Task 2 was whether listeners were able to categorise the same CE speakers whose samples were used in Task 1 into three broad Chinese regions, this time relying on subtler phonetic/phonological differences in CE. The results from §4.0 show that all listeners performed at around chance level. Although the Chinese listeners' more competent "native ear" in Task 1 performed slightly better in the accent categorisation exercise in Task 2, the difference was found to be not significant when tested.

Once more, this task seems to have been more difficult for the L1 English listeners. As none of this group had lived in China for longer than one month, and they lacked anything but the sketchiest knowledge of Chinese geography and dialects, their possible exposure to CE speakers from various Chinese regions was, consequently, very limited. Although a map and

some short training materials illustrating the region and the regional CE speech were provided, they seemed to have been insufficient for listeners to absorb much information.

Interestingly, the more competent CE "native ear" in Task 1 was no longer better than the L1 English "native ear" in the accent categorisation task, a situation we might account for in several ways. Firstly, it is well known that the English taught as a foreign language in Mainland China has long been a standard version of the language, and that students who have learned English at school have mostly been taught by CE-speaking teachers following a standard syllabus using standardised textbooks and learning materials (Wu, 2001; Hu, 2002). Therefore, teachers' L2 English realisations may be similar to those of each other, which also recalls the idea of the prototypical CE shared by all EFL learners in China that was discussed above. Furthermore, although the accents they have in their L1 Chinese may differ owing to the high dialectal diversity in China (Rose, 2002), EFL teachers in China all have to speak Mandarin Chinese natively or to a native-like level. Thus, the regional/dialectal phonetic and phonological influence of the same L1 language (Mandarin Chinese) might not be as strong as different L1 languages (Japanese or Korean) on L2 realisation. Although the present study did not make such direct comparisons, if the explanation proposed here is true we would expect both listener groups to be able to categorise these non-native accents more accurately by country than by regions within one country.

As we saw in §2.1.1, Preston (1993) and Clopper and Pisoni (2004a) demonstrated that listeners recognised speakers from their own or nearby regions more accurately when the speech they heard was in their L1, thanks to the linguistic common ground between the listeners and the speakers (see also Baker et al., 2009). In order to examine whether this generalisation also applies in L2 speech, the performance of the L1 Chinese listeners in Task 2 was further investigated.

Although the size of the available dataset was rather small, the results obtained show that L1 Chinese listeners identified CE speakers from their own broadly-defined regions no better than they did those from other regions (Figure 5). However, one thing to note here is that these three regions are very broadly-defined geographical/dialect regions. Within each region there are, unsurprisingly, a number of distinct Chinese dialects. L1 Chinese listeners coming from the same broad region as the speaker does not entail that the listener and the speaker speak the same regional dialect. For example, the Western region is very sizeable, and the five speakers selected to represent the "West" dialect were actually from two distinct Chinese dialect regions. This intraregional variation might be partially responsible for the relatively poor categorisation of the Western speakers. Listeners seemed correspondingly better at categorising Southern speakers, which is probably due to the smaller size of the Southern dialect region and the perhaps more distinct Southern dialect features in their speech. Southern dialects can be argued to deviate more from standard Mandarin Chinese than do the Northern and Western dialects, and it might follow that the L2 speech of Southern speakers contains a greater number of salient dialectal features transferred from their L1s than do the accents of the North and West. Further investigation of whether listeners can identify more narrowly-matched regional L2 accents could offer us better insights here.

5.1.3. Additional factors

In the current study, listeners' knowledge of linguistics/phonetics did appear to have an influence upon the accuracy of their accent categorisation, but not to a very significant level. Those listeners in both language groups who had received some training in one or both of discipline only performed marginally better in Task 1 than those who had had no such training,

and there was no evidence at all of an advantage of this type in the results of Task 2. This would suggest that training in linguistics/phonetics has only a minimal effect on the performance of individuals undertaking accent identification tasks of the sort administered in the present study. However, there is one possible exception in the form of the L1 English listener who had had previous training in forensic phonetics. This individual outperformed one of the phonetically-trained L1 Chinese listeners. A follow-up experiment in which the type and quantity of phonetic training is more tightly controlled might provide a more detailed picture of the benefits that (forensic) phonetic training has on accent categorisation accuracy.

Another factor to be considered is speech style. Although the L1 Chinese listeners performed more consistently than the L1 English listeners with respect to categorisation of samples in both speech styles, the patterns shown by the latter listener group were more complex. In Task 1, the L1 English listeners' identification accuracy increased by 9.2% for spontaneous speech relative to read speech. It is possible that the CE speakers self-monitored their pronunciations less when speaking spontaneously as compared to when they read aloud, and so it seems plausible that in their spontaneous utterances there might be a relatively greater number of CE-characteristic cues available to listeners. Another possibility we might wish to take into account is rapid foreign-accent adaptation (Clarke & Garrett, 2004). The spontaneous speech samples were heard by listeners in the second part of Task 1, and so the L1 English listeners would have had the chance to familiarise themselves with the properties of the CE accent through listening to it in the form of read speech in the first part of the task. We might suppose, then, that their ability to distinguish CE from JE and KE could thereby have been boosted.

In Task 2, however, the L1 English listeners' performance actually fell by 7.4% when the stimuli were samples of spontaneous speech, further complicating the interpretation of the results. One possible explanation for this finding is that the listeners were able to make direct comparisons of the linguistic/phonetic features of the read speech of the different speakers because the lexical content was identical in each case, whereas with spontaneous, unscripted speech the same comparison task is made more difficult. The complex patterns in the L1 English listeners' results might also reflect the possibility that they used cues during the accent categorisation decision process that were different from those utilised by the L1 Chinese listeners. Further research on the topic may yield results that provide empirical support for these suggestions.

Additionally, owing to the complexity of the design of the current study, it is rather difficult to draw explicit conclusions about the relationship between the listeners' performance and their confidence self-ratings. The fact that the L1 Chinese listeners expressed a high level of confidence when giving correct answers lends extra weight to the idea that their "native ear" competence equips them exceptionally well for categorisation tasks of the kind described in this paper. Modifications to the length and technical quality of the speech samples – i.e., making them shorter and noisier – could be a way of gauging how much more competent this type of "native ear" would remain under conditions more closely resembling those encountered in everyday life, including in forensically-realistic scenarios.

5.2. Features of CE, JE and KE

Clopper & Pisoni (2004b) report that dialect categorisation performance in their study was better for some listeners than others, even where only one dialect was being considered. This finding may reflect the fact that individual talkers differ from one another by exhibiting different dialect-specific properties that are salient to listeners to different degrees. Comparison of each individual speaker's speech production patterns against how well accurately they were

classified by different listener groups can, thus, give indications of which features of CE, JE and KE are noticed more readily by different listeners. A lack of space in the present paper unfortunately prevents further discussion of these issues. Instead, the linguistic and phonetic features noted by listeners, and those found in the speech of the speakers who had been most consistently correctly and incorrectly categorised by the two listener groups, were analysed with reference to the CE, JE and KE features listed in other studies and reports (Deterding, 2006; Zhang & Yin, 2009; ENGLISH Speak Like A Native, 2013).

One of the most frequently-noted features used by L1 English listeners to differentiate (non-)native accents of English is the realisation of /r/. /r/ alternates with /n/ in some Chinese dialects, but (in the form of the alveolar tap [ɾ]) is used interchangeably with [l] by both JE and KE speakers. Thus, the realisation of /r/ could potentially be a cue distinguishing CE speakers from JE and KE speakers. Another feature is the realisation of /f/, which is a phoneme in Mandarin Chinese as well as in English. As a consequence of the phonological properties of their L1s, however, JE and KE speakers often realise English /f/ as [ɸ] and [p], respectively. /f/'s voiced counterpart /v/ also has high potential as a distinguishing feature of the three accents, as CE speakers tend to realise /v/ as [w], while JE speakers use [β] or [b]. KE speakers have a tendency to use [b] for /v/.

As for vowels, Chinese speakers do not differentiate short and long vowels, and CE is said to be more syllable-timed than JE and KE. Additionally, CE speakers very often insert a word-final /ə/ in words ending with a consonant. JE and KE speakers, on the other hand, tend to realise the vowel with a more retracted tongue and a more open jaw. Qualities such as [ʌ] or [ɔ] are typical. Listeners may base their judgments on these prosodic and vocalic features alongside the consonantal cues described above.

All of the aforementioned features were noted by listeners in the two groups, and the individual speaker categorisation rate displayed in §4.2.4 appears to tally with whether these features were present or absent in the speakers' speech (cf. Hill, 2007). On the other hand, judgments made solely on the basis of a handful of features should be treated with caution, given that listeners might not know all the various features of an accent, and that certain features can be commonly found in more than one accent.

Overall, the fact that almost all participants in this present study were able to make correct categorisation judgments at a rate higher than chance confirms that they do possess some knowledge and awareness of the linguistic/phonetic properties that characterise CE, JE and KE, even if they have never had any explicit instruction about these accent varieties.

6. Conclusion

6.1. Summary of the study

In responding to the research questions raised in §2.3, the present study has taken steps to investigate these issues, and has provided some leads for further research in accent identification and categorisation. The significant differences found between the categorisation results of the two listener groups confirms that familiarity with the target variety plays an important role in accent identification. The study also addressed the problem of the "native-language effect" from a different angle, by demonstrating that individuals with "native ears" for the target variety can be highly competent at identifying the accents used by their own kind, so to speak, irrespective of whether the language being spoken is their L1 or their L2.

It is true that certain aspects of the experimental design in the present study may have influenced the respective performances of the two groups. The methods used in the two tasks, i.e. where participants are played each short speech excerpt only once and are not provided with detailed training materials, requires listeners to make their judgments by relying chiefly on their pre-existing knowledge of the linguistic varieties being tested. By their nature these tasks might, therefore, be more difficult for L1 English listeners than they are for L1 Chinese listeners. Although this might mean deviating from the initial aim of making direct comparisons between L1 English and L1 Chinese listeners, we think it is still valid to infer that listeners who themselves speak the target linguistic variety are better at identifying that accent, even when the language in question is their second language, than are others. We believe that it is probably therefore legitimate to include accent categorisation tests, similar to Task 1 in the present study, in the battery of linguistic assessments imposed on asylum claimants in the LADO context. For instance, quantifying the extent to which asylum seekers can correctly identify accents/dialects of their claimed native language(s) could offer an additional way of testing the veracity of their accounts. Obviously, due to the complex and highly sensitive nature of LADO cases and the potential consequences of repatriating individuals to countries in which they are at genuine risk of persecution, the reliability of any such tests should be evaluated extremely carefully.

6.2. Suggestions for future work

As mentioned in section 5, there are multiple ways in which this study could be extended. Other than all the previously mentioned aspects, there would be value in reproducing the experiment using speakers of languages other than those represented here. As listeners' familiarity with an accent appears to be pivotal in how well they identify that accent, it would be interesting to examine whether native competence in the target language plus "acquired" familiarity with the target linguistic variety improves listeners' performance. It is also interesting to speculate about whether an L1 English speaker's ability to imitate a CE accent predicts how well a speaker of this sort can correctly categorise CE speech samples in listening tests, relative to a "native" CE speaker.

It would be equally informative to test L1 Chinese listeners' capabilities in terms of how they categorise samples of L2 Chinese spoken by L1 English speakers from the UK, US and Australia. Considering that speakers from these three countries share the same native language but speak it with rather different accents, the question of whether their L1 accents influence their L2 realisation, and whether native Chinese listeners are able to detect those differences, could justifiably be pursued further.

We believe that the results of the present study have interesting implications for the fields of sociophonetics and speech perception. The finding that a "native ear" for the target linguistic variety seems to be better than a non-native one when the listener is tasked with trying to categorise a linguistic variety with which they are familiar has a bearing on the question of the reliability of certain listeners' or earwitnesses' accent identification judgments. In the present context, the small scale of the experiment and the limited number of participants involved prevent us from generalising about the findings very far, but it seems clear to us that further investigation of the area is warranted.

References

BAKER, WENDY, EDDINGTON, DAVID, & NAY, LYNDSEY. 2009. "Dialect identification: The effects of region of origin and amount of experience". *American Speech* 84(1): 48–71.

BARANOWSKI, MACIEJ. 2013. 'Sociophonetics'. In Robert Bayley, Richard Cameron and Ceil Lucas (eds), *The Oxford Handbook of Sociolinguistics*. Oxford: Oxford University Press. pp. 403–424.

BUSH, CLARA N. 1967. "Some acoustic parameters of speech and their relationships to the perception of dialect differences'. *TESOL Quarterly* 1(3): 20–30.

CAMBIER-LANGEVELD, TINA. 2010. "The role of linguists and native speakers in language analysis for the determination of speaker origin". *International Journal of Speech, Language and the Law* 17(1): 67–93.

CAMBIER-LANGEVELD, TINA & SAMSON, ANNE-MARIEKE. 2007. "Language analysis: How to include both linguistic expertise and native competence, and why". *Workshop on Language Analysis in the Determination of National Origin of Refugees*, Amsterdam, the Netherlands, 21st June 2007.

CLARKE, CONSTANCE M. & GARRETT, MERRILL F. 2004. "Rapid adaptation to foreign–accented English". *The Journal of the Acoustical Society of America* 116(6): 3647–3658.

CLOPPER, CYNTHIA & PISONI, DAVID. 2004a. "Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization". *Language Variation and Change* 16(1): 31–48.

CLOPPER, CYNTHIA & PISONI, DAVID. 2004b. "Some acoustic cues for the perceptual categorization of American English regional dialects'. *Journal of Phonetics* 32(1): 111–140.

CLOPPER, CYNTHIA & PISONI, DAVID. 2006. "Effects of region on origin and geographic mobility on perceptual dialect categorization". *Language Variation and Change* 18: 193–221.

CLOPPER, CYNTHIA, HAY, JENNIFER & PLICHTA, BARTEK. 2011. "Experimental speech perception and perceptual dialectology". In Marianna Di Paolo & Malcah Yaeger-Dror (eds). *Sociophonetics: A Student's Guide.* London: Routledge. pp. 149–162.

COHEN, JACOB. 1988. *Statistical Power Analysis for the Behavioral Sciences* (1st edn.). Hillsdale, NJ: Lawrence Erlbaum Associates.

CURNOW, TIMOTHY J. 2001. "What language features can be "borrowed""? In Alexandra Y. Aikhenvald & Robert M.W. Dixon (eds), *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics.* Oxford: Oxford University Press. pp. 412–436.

DAVIES, ALAN. 2004. "The native speaker in applied linguistics'. In Alan Davies & Catherine Elder (eds), *The Handbook of Applied Linguistics.* New York: Blackwell. pp. 431–450.

DETERDING, DAVID. 2006. "The pronunciation of English by speakers from China". *English World-Wide* 27(2): 175–198.

EADES, DIANA, FRASER, HELEN, SIEGEL, JEFF, MCNAMARA, TIM & BAKER, BRETT. 2003. "Linguistic identification in the determination of nationality: A preliminary report". *Language Policy* 2(2): 179–199.

ELLIS, STANLEY. 1994. "The Yorkshire Ripper Enquiry: Part I". *Forensic Linguistics* 1(2): 197–206.

ENGLISH SPEAK LIKE A NATIVE. 2013. "Chinese pronunciation Problems'. [Online] Available at: http://englishspeaklikenative.com/resources/common–pronunciation–problems/chinese–pronunciation–problems/ [Last accessed 16th September 2014].

ENGLISH SPEAK LIKE A NATIVE. 2013. "Japanese pronunciation problems'. [Online] Available at: http://englishspeaklikenative.com/resources/common–pronunciation–problems/japanese–pronunciation–problems/ [Last accessed 16th September 2014].

ENGLISH SPEAK LIKE A NATIVE. 2013. "Korean pronunciation problems'. [Online] Available at: http://englishspeaklikenative.com/resources/common–pronunciation–problems/korean–learners/ [Last accessed 16th September 2014].

FOULKES, PAUL & FRENCH, J. PETER. 2001. "Forensic phonetics and sociolinguistics'. In Rajend Mesthrie (ed), *Concise Encyclopedia of Sociolinguistics.* Amsterdam: Pergamon, pp. 329–332.

FOULKES, PAUL & FRENCH, J. PETER. 2012. "Forensic speaker comparison: A linguistic-acoustic perspective." In Lawrence M. Solan & Peter M. Tiersma (eds), *The Oxford Handbook of Language and Law*. Oxford: Oxford University Press, pp. 418–421.

FOULKES, PAUL & WILSON, KIM. 2011. "Language analysis for the determination of origin: An empirical study". *Proceedings of the 17th International Congress of Phonetic Sciences,* Hong Kong, August 2011.

FOULKES, PAUL, SCOBBIE, JAMES M. & WATT, DOMINIC. 2010. "Sociophonetics'. In William J. Hardcastle, John Laver and Fiona E. Gibbon (eds), *The Handbook of Phonetic Sciences* (2nd edn). Oxford: Wiley Blackwell, pp. 703–754.

FRASER, HELEN. 2009. "The role of "educated native speakers" in providing language analysis for the determination of the origin of asylum seekers'. *International Journal of Speech, Language and the Law* 16(1): 113–138.

FRASER, HELEN. 2011. "The role of linguists and native speakers in language analysis for the determination of speaker origin: A response to Tina Cambier-Langeveld". *International Journal of Speech, Language and the Law* 18(1): 121–130.

FRENCH, J. PETER. 1990. "Analytic procedures for the determination of disputed utterances'. In Hannes Kniffka (ed), *Texte zu Theorie und Praxis Forensischer Linguistik.* Tübingen: Niemeyer Verlag, pp. 201–213.

GOGGIN, JUDITH P., CHARLES P. THOMPSON, GERHARD STRUBE & LIZA R. SIMENTAL. 1991. "The role of language familiarity in voice identification". *Memory & Cognition* 19(5): 448–458.

GOLDSTEIN, ALVIN G., KNIGHT, PAUL, BAILIS, KAREN & CONOVER, JERRY. 1981. "Recognition memory for accented and unaccented voices'. *Bulletin of the Psychonomic Society* 17(5): 217–220.

GYE, HUGO. 2014. "Almost as many Chinese graduate students in England as British ones'. [Online]. Daily Mail Online. Available at: http://www.dailymail.co.uk/news/article–2594935/There–Chinese–students–postgraduate–courses–English–universities–British–students.html [Last accessed: 15th September 2014].

HILL, DAVID R. 2007. "Speaker classification concepts: Past, present and future". In Christian Müller (ed), *Speaker Classification I: Lecture Notes in Computer Science.* Berlin: Springer Verlag, pp. 21–46.

HONOROF, DOUGLAS N., MCCULLOUGH, JILL AND SOMERVILLE, BARBARA. 2000. "Comma Gets a Cure". Diagnostic passage. [Online] Available at: http://www.dialectsarchive.com/comma-gets-a-cure [Last accessed 26th April 2015].

HU, GUANGWEI. 2002. "Recent important developments in secondary English-language teaching in the People's Republic of China". *Language, Culture and Curriculum* 15(1): 30–49.

IDEA. 1997. "International Dialects of English Archive. Asia: China". [Online] Available at: http://www.dialectsarchive.com/china [Last accessed 26th August 2014].

IDEA. 1997. "International Dialects of English Archive. Asia: Japan". [Online] Available at: http://www.dialectsarchive.com/japan [Last accessed 26th August 2014].

IDEA. 1997. "International Dialects of English Archive. Asia: South Korea". [Online] Available at: http://www.dialectsarchive.com/south–korea [Last accessed 26th August 2014].

JESSEN, MICHAEL. 2008. 'Forensic phonetics'. *Language and Linguistics Compass* 2(4): 671–711.

KERSWILL, PAUL & WILLIAMS, ANN. 2002. "Dialect recognition and speech community focusing in new and old towns in England: The effects of dialect levelling, demography and social networks'. In Daniel Long & Dennis R. Preston (eds), *A Handbook of Perceptual Dialectology, Vol. 2*. Amsterdam: John Benjamins. pp. 178–207.

KIRKPATRICK, ANDY. 2007. *World Englishes: Implications for International Communication and English Language Teaching.* Cambridge: Cambridge University Press.

KÖSTER, OLAF & SCHILLER, NIELS O. 1997. "Different influences of the native language of a listener on speaker recognition". *Forensic Linguistics* 4(1): 18–28.

KÖSTER, OLAF, SCHILLER, NIELS O. & KUŃZEL, HERMANN J. 1995. "The influence of native language background on speaker recognition". In Elenius, Kjell & Branderud, Peter (eds), *Proceedings of the 13th International Congress of Phonetic Sciences,* Stockholm, pp. 306–309.

LADEFOGED, PETER & LADEFOGED, JENNY. 1980. "The ability of listeners to identify voices'. *UCLA Working Papers in Phonetics* 49: 43–51.

MAJOR, ROY C. 2001. *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah, NJ: Lawrence Erlbaum Associates.

NEUHAUSER, SARA & SIMPSON, ADRIAN P. 2007. "Imitated or authentic? Listeners' judgements of foreign accents'. *Proceedings of the 16th International Congress of Phonetic Sciences,* Saarbrücken, pp. 1805–1808.

NOLAN, FRANCIS. 2012. "Degrees of freedom in speech production: An argument for native speakers in LADO". *International Journal of Speech, Language and the Law* 19(2): 263–289.

PATRICK, PETER L. 2012. "Language analysis for determination of origin". In Lawrence M. Solan & Peter M. Tiersma (eds), *The Oxford Handbook of Language and Law*. Oxford: Oxford University Press, pp. 533–546.

PERRACHIONE, TYLER K., PIERREHUMBERT, JANET B. & WONG, PATRICK C.M. 2009. "Differential neural contributions to native-and foreign-language talker identification." *Journal of Experimental Psychology: Human Perception and Performance* 35(6): 1950–1960.

PRESTON, DENNIS R. 1993. "Folk dialectology". In Dennis R. Preston (ed), *American Dialect Research.* Philadelphia: John Benjamins, pp. 333–378.

PURNELL, THOMAS, IDSARDI, WILLIAM & BAUGH, JOHN. 1999. "Perceptual and phonetic experiments on American English dialect identification". *Journal of Language and Social Psychology* 18(1): 10–30.

RAMSEY, S. ROBERT. 1987. *The Languages of China.* Princeton, NJ: Princeton University Press.

ROSE, PHILIP. 2002. *Forensic Speaker Identification.* London: Taylor & Francis.

SCHILLER, NIELS O. & KÖSTER, OLAF. 1996. "Evaluation of a foreign speaker in forensic phonetics: A report". *Forensic Linguistics* 3(1): 176–185.

SCHLICHTING, FRANK & SULLIVAN, KIRK P.H. 1997. "The imitated voice: a problem for voice line-ups'? *Forensic Linguistics* 4(1): 148–165.

THOMPSON, CHARLES P. 1987. "A language effect in voice identification". *Applied Cognitive Psychology* 1(2): 121–131.

WELLS, JOHN C. 1982. *Accents of English* (3 vols.). Cambridge: Cambridge University Press.

WILCOX, GEORGE K. 1978. "The effect of accent on listening comprehension: A Singapore study". *English Language Teaching Journal* 32: 118–127.

WILLIAMS, ANGIE, GARRETT, PETER & COUPLAND, NIK. 1999. "Dialect recognition". In Preston, Dennis R. (ed), *Handbook of Perceptual Dialectology.* Philadelphia: John Benjamins, pp. 345–358.

WILSON, KIM. 2009. *Language Analysis for the Determination of Origin: Native Speakers vs. Trained Linguists.* Unpublished MSc dissertation, University of York.

WILSON, KIM & FOULKES, PAUL. 2014. "Borders, variation, and identity: Language analysis for the determination of origin (LADO)". In: Dominic Watt & Carmen Llamas (eds), *Language, Borders and Identity.* Edinburgh: Edinburgh University Press, pp. 218–229.

WU, YI'AN. 2001. "English language teaching in China: Trends and challenges'. *TESOL Quarterly* 35(1): 191–194.

ZHANG, FACHUN & YIN, PENGPENG. 2009. "A Study of pronunciation problems of English learners in China". *Asian Social Science* 5(6): 141–146.

Chen Shen
School of English & Department of Human Communication Sciences
University of Sheffield
1 Upper Hanover Street
Sheffield
S3 7RA
email: irenechenshen@gmail.com

Dominic Watt
Department of Language and Linguistic Science
University of York
Heslington
York
YO10 5DD
email: dominic.watt@york.ac.uk