Editorial note

Welcome to the fourteenth issue of *York Papers in Linguistics Series 2* (YPL). The issue contains papers from staff, postgraduate students and alumni from the Department of Language and Linguistic Science, University of York UK.

The editors would like to thank all the anonymous reviewers who agreed to review papers and provided invaluable comments and advice; the YPL editorial board, Mariam Dar and Aiqing Wang, for their time and assistance with finding reviewers and checking papers; Natalie Fecher and Ania Kubisz for practical and technological support; and the Department of Language and Linguistic Science academic and administrative staff for their endless patience.

Please note that publication in YPL does not prejudice future publication elsewhere.

For more information on the current issue and past YPL publications please visit [www.york.ac.uk/language/ypl/](http://www.york.ac.uk/language/ypl/) Please address all enquiries to the YPL committee at ypl@york.ac.uk.

Rebecca Woods and Ella Jeffries
York, 30th June 2015
Contents

Developing an Urdu CDI for early language acquisition
Mariam Dar, Huma Anwaar, Marilyn Vihman & Tamar Keren-Portnoy……1

Pre-school children's identification of familiar speakers and the role of accent features
Ella Jeffries………………………………………………………………………15

The role of gender-correlated sociolinguistic variables in identifying speaker-indexical information
Ania Kubisz………………………………………………………………………41

Implicit associations with Welsh in two educational contexts
Rachelle Lee………………………………………………………………………81

Accent categorisation by lay listeners: which type of "native ear" works better?
Chen Shen & Dominic Watt……………………………………………………106
Abstract

This paper reports on the development of a lexical checklist for parents to trace vocabulary advances in Pakistani children. Various cross-linguistic adaptations of the MacArthur-Bates Communicative Development Inventory (CDI) were consulted. Pilot testing was carried out using a sample of 17 children from middle-class homes, aged from 12 to 30 months. Cross-linguistic comparison reveals similarities between the vocabulary growth of the Pakistani children and children learning other languages. Plans for further pilot-testing and eventual validation are discussed.

1. Introduction

The MacArthur–Bates Communicative Development Inventory (CDI) has been used extensively for both clinical and research purposes. Various studies have testified to the effectiveness of this parent report in measuring lexical development in the early years (Dale, 1991; Dale, Bates, Reznick, & Morisset, 1989; Miller, Sedey, & Miolo, 1995; Thal, O’Hanlon, Clemmons, & Fralin, 1999). The instrument has been adapted into 38 languages (Bates, Dale & Thal, 1995), including some bilingual adaptations (Gatt, 2007; O’Toole & Fletcher, 2008; O’Toole & Fletcher, 2010; O’Toole, 2013). Cross-linguistic comparison shows similarities in the vocabulary growth of monolingual and bilingual children. The current study is a part of larger investigation exploring the development of the lexicon in children from multilingual homes in Pakistan, which has so far received little attention. The overall aim is to formulate an assessment tool for Urdu-speaking children, most of whom are learning it in a multilingual context; this can lead to advances in child language research as well as in clinical practice in Pakistan.

1.1. Significance of Urdu adaptation of CDI

Urdu is estimated to be spoken by 100 million people around the world (BBC, 2014). Apart from Pakistan, it is spoken and understood in parts of India, Bangladesh, Nepal, the Middle East, and many other countries around the world. For instance, the Urdu speaking community in the UK is comprised of about 400,000 speakers. Urdu is essentially the same language as Hindi, which is spoken by 425 million as a native language (The Constitution of India, 2007). Both Hindi and English are official languages in the major cities in India, along with 22 official regional languages throughout the country. Urdu and Hindi, despite having different names and orthography, have only minor differences in lexicon and share the same grammar. Due to these linguistic and cultural similarities, the Urdu adaptation of CDI could also be used (with modifications, notably in the spelling of Urdu/Hindi words) for multilingual children in India.
1.2. The linguistic situation in Pakistan

Pakistan is home to many different languages and cultures, which makes the development of an Urdu adaptation of the CDI for children’s early words a challenging task. Pakistan has a total population of 186 million; roughly 60 languages are spoken over the region (Akram & Mahmood, 2007). Punjab is the largest province and Punjabi the regional language with the largest number of speakers (44% of Pakistanis speak Punjabi: Akram & Mahmood, 2007). Other widely spoken regional languages include Pashto (15%), Sindhi (14%), Saraiki (10%), Balochi (4%), and Hindko (1%; see Figure 1).

It is worth noting that Urdu, the national language of Pakistan, is native to only 7% of the population, predominantly those living in urban areas. However, it is used as a lingua franca among people from different ethnic backgrounds. There is no exact figure available on the number of L2 speakers of Urdu, although we estimate it to be roughly 60%. In the villages and small towns, children are exposed to Urdu as an L2 at school when they are aged three to five years.

Another language commonly spoken in the cities is English, also an official language. The majority of the population in urban areas speak either Urdu or English, which, as the languages of the educated and professional class, carry the most prestige.

English was introduced to the sub-continent through colonization by the British Empire in 1858. The British introduced a new educational system with English as the medium of instruction to create a link between the colonisers and the colonised (Evans, 2002). Later, English found its way into government administration, the law, the military, commerce and the mass media, where it continued to dominate, even after the end of colonization. It has surpassed

Figure 1: Distribution of native language use in Pakistan (Akram & Mahmood 2007: 2; adapted from table 1).
even the national language Urdu in popularity and has become a status symbol, particularly in urban areas. “It is the gateway to success, higher education and white collar jobs” (Ghani 2003: 105). Code-switching is very common in the cities in particular, where it is fashionable to insert English words into everyday speech.

The complex linguistic situation is reflected in the education system. Currently, four different school systems co-exist: Private Elite English medium, Private non-elite English medium, Government Urdu medium and Dini Madrasas (private religious schools focused on the teaching of the Quran, its interpretation, the sayings of the Prophet Muhammad and Islamic Law: Coleman & Capstick, 2012). Private elite English medium schools, which are very expensive and thus serve only the most prosperous families, are found in the larger cities; in these schools the medium of instruction is English only. The tutors are either native speakers of English or have received their training abroad. Children are extensively trained from the start to speak English ‘like a native’. Private non-elite English-medium schools, which have local tutors, are tailored to the needs of the aspirational middle class, lower middle class and upper working class unable to afford the fees of the elite schools. These schools aim to develop reasonable proficiency in English. Children of the lower class in the urban and rural areas go to government-supported Urdu-medium schools, which are free; the Dini Madrasas, also Urdu-medium, provide religious education only. There are also some primary schools in Sindh and Khyber Pakhtunkhwa in which the medium of instruction is the regional language. Note that English is a compulsory subject in all but the Dini Madrasas.

1.3 The linguistic situation in homes

Because Urdu co-exists with English in a multilingual society there are abundant English loan words in Urdu as well as in the regional languages. A child raised in a middle-class home is typically exposed to Urdu, English and one or two regional languages. Parents encourage their children to learn and communicate in English since they regard it as the means to progress and prosperity. From a very early age infants are exposed to English through cartoons, books and rhymes. Parents and caretakers commonly insert English words into their everyday speech through code-switching. They usually switch to English when disciplining their children. Children are encouraged to learn and practice English politeness expressions such as thank you, please, sorry etc. Although children are exposed to English mostly through code-switching, we consider these children ‘bilingual’ because in most cases they produce more English than Urdu words (see Table 3). The children are fully exposed to English grammar only when they start school from the age of 3-5 years. Since they have already been exposed to English vocabulary at home, they begin to speak English with ease as soon as they start school.

However, it is important to note that exposure to a regional language is highly dependent on the region. People from rural areas or small cities or towns are proud of their regional affiliation and the regional languages are used frequently. However, regional languages do not hold much importance in big cities. Children are mostly exposed to Urdu and English and the regional language is introduced (if at all) after age two or three years because of the social stigma attached to it (speaking English/Urdu with a strong regional accent is associated with low social status). The Urdu CDI will initially be piloted and validated on a sample taken from the capital, Islamabad. For that reason, only English and Urdu will be included in the early versions of the CDI.
1.4 Research aims of the pilot study

The Urdu CDI is being developed with the aim of achieving a cultural and linguistic adaptation for Urdu-speaking children. Pilot testing was first carried out in order to test the appropriateness of the vocabulary items in the checklist in the context of a multilingual community. We wanted to make sure that we were choosing items that were relevant to the children’s cultural and linguistic background. Another goal was to determine whether the vocabulary level would be comparable to that of other children of similar ages tested using other-language adaptations of the tool.

2. Development of an Urdu CDI

Due to the lack of data on Urdu language development, the preliminary checklist was based on the authors’ own experience of living around children in the extended families typical of Pakistani homes. It was further validated by both face-to-face and Skype-interviews with mothers, all of whom come from middle-class Urdu backgrounds. In order to make the list more comprehensive, the American original (Fenson et al., 2007) and one UK adaptation of the CDI (Hamilton et al., 2010) were taken as models and all the vocabulary items included in those lists were screened for their relevance to lexical development in the Pakistani context. The Bengali adaptation (Hamadani et al., 2010) was also consulted for its cultural relevance to Pakistani languages. However, it proved less relevant, as it is very short (60 words) and was developed and tested with children from rural areas. The Maltese (Gatt, 2007) and Irish (O’Toole & Fletcher, 2008; O’Toole & Fletcher, 2010; O’Toole, 2013) adaptations were consulted for help in dealing with the multilingual aspect. A great number of lexical terms were added at this stage in order to make the list of words that children might know as complete as possible. A comprehensive checklist was developed based on these modifications and then further improved after feedback from colleagues.

2.1. Vocabulary Checklist

The initial checklist takes elements from both the “Words and Sentences” and “Words and Gestures” vocabulary inventories (CDI: WS), intended for use with infants aged 12 to 30 months. In order to reflect the linguistic and cultural features of Pakistani society the list includes 24 lexical categories (compared with 19 in the British and 22 in the American versions).

It was deemed essential to reflect the bilingualism characteristic of Pakistani homes in the Urdu CDI. Most concepts were represented as doublets, by including an English term in parallel with the Urdu one (e.g., cat and /bɪl:i/, car and /ɡaːtɪ/). Singlets were also added, including a) concepts specific to Pakistani culture, which have no equivalent in English (e.g. /mendi/ ‘temporary tattoo’, /ˈrɒtɪ/ ‘a kind of bread’), b) concepts present in Urdu but replaced by English words in everyday usage (e.g. /xaːndan/ replaced by family, /gəmd/ replaced by ball) and c) concepts adopted as inventions from the West with no equivalents in native speech (e.g. computer, telephone), which we consider loan words.

For expository purposes the lexical terms can be broadly divided into noun phrases, predicates and closed class words.
2.1.1. Noun phrases

The noun phrases in the early Urdu checklist correspond to the English-version categories of ‘sound effects and animal sounds’, ‘animals’, ‘vehicles’, ‘toys’, ‘food and drink’, ‘clothing’, ‘body parts’, ‘small household items’., ‘furniture and rooms’, ‘outside things’, ‘places to go’, and ‘people’. Two new categories, ‘festivals’ and ‘religion’, were added to the existing ones to reflect the culture and life style of Pakistani people.

Religion plays an important role in the children’s environment. For example, children see adults praying on the prayer mat (/jaː naməz/ ‘praying mat’), raising their hands in a prayer (/duːz/ ‘prayer’). It is, for instance, a favourite routine of toddlers to bow down on the prayer mat in imitation of adults. Pakistani children also grow up witnessing religious and cultural festivals. For example, a wedding in Pakistan (/ʃaːdɪ/) takes place over three days, including /mendi/, the occasion of putting temporary tattoos on the bride’s hands, /bəɾaːt/, the event in which the bride is married to the groom and is seen off by her family, and /valiː maː/, the official reception party offered by the groom’s side. Other religious occasions include /iːd/, the most popular religious festival, celebrated after the holy month of Ramadan, and /tʃaːnd raːt/, the practice of moon sighting before /iːd/. These events and festivals are frequently celebrated in Pakistan, so all Pakistani children are familiar with them.

Additional culturally specific lexical items have been added to the existing categories. For example, words related to the concept of power shutoff, a common phenomenon in Pakistan that may be less often witnessed and spoken about in other societies, include such early-understood and used terms as lantern, generator, UPS and load shedding. These concepts are deeply integrated into the Pakistani way of life and every child is exposed to them. In fact, while doing the transcriptions in a recent project on 18-month-old-Pakistani children, we found that the children were already producing these words. One of the mothers in one of the video recordings actually switched off the light, to elicit electricity gone from the child.

The category ‘people’ was also extended by adding numerous kinship terms. Pakistani children are raised in a joint family system under the influence of grandparents, uncles, aunts and cousins and as a result Urdu has a complex system of kinship terms. Since respect and regard for elders is an important aspect of Pakistani society, special titles exist for almost all the maternal and paternal relationships (e.g. /naː naː/ ‘maternal grandfather’, /daː daː/ ‘paternal grandfather’). Older and younger kin sometimes also have special terms (e.g., /ˈbʰəija/ ‘older brother’, /bʰaːiː/ ‘younger brother’, /tʃətʃə/ ‘younger paternal uncle’ and /taːjə/ ‘older paternal uncle’).

2.1.2. Predicates

Predicates include verbs and adjectives. In Urdu both of these parts of speech carry number and gender markers. Since all the forms could not be listed, parents are instructed to tick the infinitive form of verbs (e.g. /kaːtəː/ ‘to cut’, /dəbaːnə/ ‘to press’). That is, explicit instructions are given to parents to mark the form even if the child does not use the infinitive but another form.
2.1.3. Closed class

These words include pronouns, question words, prepositions, quantifiers, auxiliaries. An additional category of ‘tenses’ was added, since Urdu tenses are quite complex. The tense is expressed by both an auxiliary and a main verb, each of which takes different forms according to tense and gender.

Within each category the Urdu words and their English equivalents were presented together, separated by a diagonal line (see Figure 2). Urdu words were also given in the Roman alphabet used for English; it is common practice in Pakistan to use the Roman alphabet for informal communication in Urdu over emails, text messages and Internet chat. Mothers were instructed to underline the form(s) used by the child. This was intended to minimise the length of the Urdu CDI.

<table>
<thead>
<tr>
<th>6. BODY PARTS</th>
<th>U</th>
<th>U/S</th>
<th>U</th>
<th>U/S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haath/Hands</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naal/Belly Button</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pait/Taund/Tummy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pair/Pacun/Feet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ankh/Eye</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moo/Face</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ongli/Finger</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dant/Teeth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zaban/Tongue</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thori/Chin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bummies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Sample of items listed under ‘body parts’ (‘U’ means understands only and ‘U/S’ means both understands and says).

3. Methodology

3.1. Participants

The data was collected in two phases. Eight Urdu-speaking mothers with children in the age range of 12 – 30 months participated in the first pilot. Gender was not controlled for. Only full-term children from middle-class homes with no hearing disabilities were included. All the children were reportedly spoken to primarily in Urdu within the home setting. The families
come from the two most important cities of Pakistan, Islamabad and Peshawar. Friends and acquaintances were contacted for the study. All had completed postgraduate education. Nine Urdu-speaking mothers participated in the second pilot. The same criteria were used for sample selection as in the first pilot. One child from the first pilot (18 months of age at that time) also took part in the second, at 30 months of age. All the families were settled in Islamabad, the capital city of Pakistan and Urdu was mainly spoken in all households. All parents had completed postgraduate education (see Table 1).

<table>
<thead>
<tr>
<th>Infant age in months</th>
<th>Gender</th>
<th>Siblings</th>
<th>Birth order</th>
<th>Mother’s profession</th>
<th>Father’s profession</th>
<th>% Urdu spoken</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Pilot Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 16</td>
<td>F</td>
<td>0</td>
<td>1</td>
<td>auditor</td>
<td>banker</td>
<td>45%</td>
</tr>
<tr>
<td>2 18</td>
<td>M</td>
<td>0</td>
<td>1</td>
<td>house-wife</td>
<td>banker</td>
<td>50%</td>
</tr>
<tr>
<td>3 24</td>
<td>F</td>
<td>1</td>
<td>2</td>
<td>linguist</td>
<td>engineer</td>
<td>65%</td>
</tr>
<tr>
<td>4 24</td>
<td>F</td>
<td>1</td>
<td>2</td>
<td>IT</td>
<td>IT</td>
<td>50%</td>
</tr>
<tr>
<td>5 24</td>
<td>M</td>
<td>0</td>
<td>1</td>
<td>lecturer</td>
<td>lecturer</td>
<td>55%</td>
</tr>
<tr>
<td>6 30</td>
<td>F</td>
<td>0</td>
<td>1</td>
<td>lecturer</td>
<td>lecturer</td>
<td>45%</td>
</tr>
<tr>
<td>7 30</td>
<td>M</td>
<td>1</td>
<td>2</td>
<td>house-wife</td>
<td>businessman</td>
<td>60%</td>
</tr>
<tr>
<td>8 30</td>
<td>M</td>
<td>1</td>
<td>2</td>
<td>teacher</td>
<td>armed forces</td>
<td>60%</td>
</tr>
<tr>
<td>Second Pilot Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 12</td>
<td>F</td>
<td>1</td>
<td>2</td>
<td>House-wife</td>
<td>banker</td>
<td>50%</td>
</tr>
<tr>
<td>2 13</td>
<td>F</td>
<td>0</td>
<td>1</td>
<td>house-wife</td>
<td>Engineer</td>
<td>69%</td>
</tr>
<tr>
<td>3 17</td>
<td>M</td>
<td>1</td>
<td>1</td>
<td>School teacher</td>
<td>Businessman</td>
<td>65%</td>
</tr>
<tr>
<td>4 18</td>
<td>M</td>
<td>0</td>
<td>1</td>
<td>House wife</td>
<td>IT</td>
<td>70%</td>
</tr>
<tr>
<td>5 22</td>
<td>M</td>
<td>0</td>
<td>1</td>
<td>lecturer</td>
<td>Sales</td>
<td>45%</td>
</tr>
<tr>
<td>6 24</td>
<td>M</td>
<td>0</td>
<td>1</td>
<td>lecturer</td>
<td>lecturer</td>
<td>55%</td>
</tr>
<tr>
<td>7 28</td>
<td>F</td>
<td>0</td>
<td>1</td>
<td>HR</td>
<td>businessman</td>
<td>60%</td>
</tr>
<tr>
<td>8 28</td>
<td>M</td>
<td>1</td>
<td>2</td>
<td>teacher</td>
<td>armed forces</td>
<td>65%</td>
</tr>
<tr>
<td>9 30</td>
<td>M</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Background of participants

3.2. Data Collection

In both the first and the second pilot the Urdu CDI was submitted to the parents electronically and returned within two weeks of submission. Background information was collected by phone prior to submission, at which time mothers were also familiarised with the checklist. Detailed instructions were sent along with the CDI a few days later. Mothers were asked to make a note of the time it took them to fill in the Urdu CDI and also to rate the difficulty level. There was a comment box after each category to allow parents to add words not mentioned in any given category. The checklist data, including additional lexemes contributed by the parents, were coded manually.
4. Analysis

4.1. Challenges in assessing lexicon size for multilingual children

Assessing lexicon size in multilingual children is challenging due to the likely overlap of lexical knowledge in two (or more) languages. We wanted to compare the lexical level of Pakistani children with the lexical levels of monolingual and bilingual children learning other languages at similar ages. That would be helpful in determining how the language development reported through the tool relates to general learning trends in children all over the world. In previous adaptations of the MacArthur Communicative Development Inventories (CDI), the Total Vocabulary (TV) score – the total number of words a child knows – was computed to assess lexical growth in monolinguals; in bilinguals the total would logically be the words known in two languages. However, measuring lexical development in multilinguals is more complex than in monolinguals. In monolinguals the number of concepts can be taken to be roughly equal to the number of words, whereas in bilinguals or multilinguals one concept can be mapped to two or more words. The approach generally taken is to calculate the total number of concepts mapped onto two (or more) languages; Total Conceptual Vocabulary (TCV) is arrived at by counting only once the different labels for a single concept in two (or more) languages (Pearson et al., 1993). However, this measure may also be misleading, in that the translation equivalents might not be semantically equivalent for the child. Volterra and Taeckner (1978) showed that a bilingual German-Italian child was using doublets to talk about separate concepts (barco for sailboats and boat for other boats). In order to analyze the extent to which lexical knowledge in one language overlaps with lexical knowledge in the other(s) a separate comparison between Total Vocabulary (TV) and Total Conceptual Vocabulary (TCV) is advisable. This means that the vocabulary size for each of the languages has to be determined separately. This not only demands more time but also raises the issue of determining the source language for phonetically similar e.g., onomatopoeic forms like ‘crow call’, caw (English) and kaẽ-kaẽ (Urdu).

4.2. Analysis for the current study

The analysis could not be carried out as planned for the first pilot. In the checklist, Urdu words and their English equivalents were presented on the same line, separated by a diagonal line. Parents were instructed to underline the form(s) used by the child, which placed an additional demand on them: They had to tick the box as well underline the words. In the event, only three mothers followed the instructions properly, which made it impossible to judge the children’s Total Vocabulary (TV). Only the Total Conceptual Vocabulary (TCV) score, in which each column was counted only once even though a child could potentially produce/comprehend two different forms for that concept, was calculated.

In the second pilot, the format of the checklist was modified to resolve the issue identified in the first one. The Urdu words and their English equivalents were given on separate lines and Urdu words were written in Urdu orthography (see Figure 3). This allowed the calculation of Total Vocabulary (TV) along with Total Conceptual Vocabulary (TCV). Also, in the open class categories only words reported by at least one parent from the first pilot were retained in the second. Additional lexemes contributed by parents were also incorporated. All closed-class items were retained, irrespective of whether or not they were scored by any parents.
Figure 3: Sample of revised Urdu checklist items listed under ‘people’. The figure shows how two kinds of vocabulary items, doublets and singlets, are represented in the CDI.

Table 2 summarises the descriptive statistics for the seventeen children whose lexical development was recorded in the pilot study. Depending on the age of the child and the level of expressive vocabulary, the Urdu CDI took from 30 to 45 minutes to complete. The difficulty level was rated as medium by the parents. As the table shows, the sample size is very small in each group, ranging from 1 – 3 children.

It is also important to mention the Total Vocabulary (TV) scores in English and Urdu separately. As discussed in the earlier section, Urdu and English are the most often used languages in the major cities in Pakistan. Although children are too young to be formally exposed to English at school, the English learning that occurs through exposure to TV, books and code-switching from adults plays an important role in the children’s development. Table 3 reports the vocabulary scores of children from the second pilot.
## First Pilot Test

<table>
<thead>
<tr>
<th>Child ID</th>
<th>Age in Months</th>
<th>Total Conceptual Vocabulary (TCV)</th>
<th>Receptive-only Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>125</td>
<td>292</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>544</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>162</td>
<td>124</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>430</td>
<td>119</td>
</tr>
<tr>
<td>5</td>
<td>24</td>
<td>251</td>
<td>132</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>281</td>
<td>125</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>350</td>
<td>540</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>407</td>
<td>148</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>398</td>
<td>371</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>385</td>
<td>353</td>
</tr>
</tbody>
</table>

## Second Pilot Test

<table>
<thead>
<tr>
<th>Child ID</th>
<th>Age (Months)</th>
<th>Total Vocabulary (TV)</th>
<th>Total Conceptual Vocabulary (TCV)</th>
<th>Receptive-only Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>68</td>
<td>67</td>
<td>92</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>144</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>27</td>
<td>23</td>
<td>87</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
<td>109</td>
<td>102</td>
<td>207</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>324</td>
<td>270</td>
<td>93</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>285</td>
<td>227</td>
<td>62</td>
</tr>
<tr>
<td>7</td>
<td>28</td>
<td>468</td>
<td>385</td>
<td>93</td>
</tr>
<tr>
<td>8</td>
<td>28</td>
<td>705</td>
<td>574</td>
<td>180</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>586</td>
<td>479</td>
<td>136</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>818</td>
<td>627</td>
<td>129</td>
</tr>
</tbody>
</table>

### Table 2: Vocabulary level reported for sampled children

<table>
<thead>
<tr>
<th>Child ID</th>
<th>Age (Months)</th>
<th>Total Vocabulary (TV) in English</th>
<th>Total Vocabulary (TV) in Urdu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>38</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
<td>43</td>
<td>56</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>169</td>
<td>155</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>140</td>
<td>145</td>
</tr>
<tr>
<td>7</td>
<td>28</td>
<td>269</td>
<td>199</td>
</tr>
<tr>
<td>8</td>
<td>28</td>
<td>400</td>
<td>305</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>574</td>
<td>244</td>
</tr>
</tbody>
</table>

### Table 3: Vocabulary level in English and Urdu reported for sampled children
It is significant to mention that one of the aims was to reduce the size of the vocabulary of Urdu CDI. As was pointed above, only those words in the open class categories were retained, after both the pilots, which were reported by at least one of the parents. The words not reported by any of the parents were taken out. This was done to make the word list compact and comparable to word lists in other languages. To start with, the CDI had a word count of 2356 which has been reduced to 1290 words after the two pilots. Table 4 provides the word count of some other monolingual and bilingual CDIs for comparison.

<table>
<thead>
<tr>
<th>Monolingual</th>
<th>Vocabulary count</th>
<th>Bilingual</th>
<th>Vocabulary count</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>680</td>
<td>Irish</td>
<td>826</td>
</tr>
<tr>
<td>Italian</td>
<td>670</td>
<td>Maltese</td>
<td>1681</td>
</tr>
<tr>
<td>Danish</td>
<td>725</td>
<td>Urdu</td>
<td>1290</td>
</tr>
</tbody>
</table>

Table 4: Comparison of the word count of monolingual and bilingual versions of the CDI.

A comparison of the Pakistani children’s vocabulary level with that of American, Italian, Danish, Maltese and Irish children is presented in Figure 4 (Total Vocabulary scores (TCV) of Irish, Maltese and Urdu children were compared). The results can be viewed in terms of the vocabulary growth typical of monolingual as compared with bilingual children. First, the vocabulary development of the Urdu children is in line with that of monolingual children in both pilots. For instance, at 24 months Urdu-speaking children knew an average of 281 (first pilot) and 227 (second pilot) words; this compares with 307 for American (Dale & Fenson, 1996), 292 for Italian (Caselli et al., 2001) and 262 for Danish children (Anderson et al., 2006). The vocabulary sizes were also in line with those reported in the studies involving bilingual children: At 24 and 30 months the Urdu-speaking children knew an average of 281 and 385 words (in the first pilot), which is comparable to Irish (240 and 440 words at similar ages) and Maltese (356 words at 30 months). Similarly, in the second pilot, Urdu speaking children knew an average of 68 and 227 words at 16-18 and 24 months, which is comparable to Irish (81 and 240 at similar ages) and Maltese (58 words at 16-18 months). At 16-18 months (first pilot) and 30 months (second pilot) the TCV scores from the Urdu CDI considerably surpass the mean scores from either monolingual or bilingual children in the other studies. This is because one child with an unusually high TCV score (512) at 16-18 months in the first pilot also took part in the second pilot at 30 months of age.
Figure 4: Comparison of mean expressive vocabulary scores by age from CDI adaptations. Sources: American (Dale & Fenson, 1996), Irish (O’Toole & Fletcher, 2010), Maltese (Gatt, 2007), Italian (Caselli et al., 2001), Danish (Anderson et al., 2006).

5. Discussion and conclusions

The basic aim of the study was to measure the lexical development of Pakistani children using an Urdu CDI, with a view to designing a larger study on the basis of the outcome. Both monolingual and bilingual adaptations of the MacArthur Communicative Inventory (CDI) were consulted. The preliminary Urdu checklist was pilot tested on eight children from middle class homes and the Total Conceptual Vocabulary (TCV) score was established. In the second pilot testing, data was collected from nine children from middle-class homes. Both Total Vocabulary (TV) and Total Conceptual Vocabulary (TCV) scores were calculated. The sample size is still too small to permit any conclusions about general trends in lexical development in Pakistan. However, these two stages of pilot testing have helped to refine the CDI and increase its appropriateness for use with the bilingual children of major cities of Pakistan. The results from the CDI are also comparable to those obtained from monolingual and bilingual children who are learning other languages. The research has now moved to the next stage, validation, which started in January 2015. The instructions have been prepared in both languages, as recommended by O’Toole (2013). A sample of 20-25 Urdu-learning children are being selected for the initial validation process. Care will be taken to include at least five age groups, with no fewer than two children in each group. Parental reports will be supplemented by video-recordings of a conversational sample (of approximately 30 minutes) involving each parent and child and using toys and picture books.

As in the pilot studies reported in this paper, only words reported by at least one parent will be retained in the open class categories in subsequent studies. Additional lexemes contributed by parents will also be incorporated. All closed-class items will be retained, irrespective of whether or not they have been checked by any parents.
References


Mariam Dar, Huma Anwaar, Marilyn Vihman, Tamar Keren-Portnoy
Department of Language and Linguistic Science
University of York
Heslington
York
email: md738@york.ac.uk
Abstract

This paper explores pre-school children’s abilities in recognising and identifying familiar speakers. Two experiments were run with a group of 22 nursery children from York, in the north of England, aged 2.4-4.9 years. The first experiment investigates the children’s ability to identify six known nursery school teachers from short audio stimuli. The results of this experiment showed an improvement in the identification ability of the children correlating with age, greater exposure to the teachers, and a longer stimulus length. The pitch and voice quality of the speakers’ voices were also found to play a role in how easily identified the individuals were. The second experiment considers the role of regional accent features in the recognition of one particular speaker from Yorkshire when compared to other, unfamiliar speakers with different regional accents of British English. This experiment presents more mixed results and different results between the sexes. Among the boys, the results showed an age-related improvement. Girls with parents from outside of Yorkshire performed better than those with parents from Yorkshire. Overall, the children were more likely to misidentify an unfamiliar speaker as the familiar speaker if the speakers had a similar regional accent. Furthermore, a disguise in accent by the familiar speaker resulted, as expected, in more children being unable to identify her. The results are discussed in terms of memory models of speaker recognition.

1. Introduction

Studies show that infants 4–5 months old demonstrate a familiarity and preference not just for the voice of their mother (whose speech they are able to process faster; cf. Purhonen et al. 2005), but also more generally for their own accent over other regional accents (Butler et al. 2011). Furthermore, when it comes to word learning, infants have been found to only recognise familiar words in their local accent, before being able to extrapolate this across other regional accents as they grow older (cf. Schmale et al., 2010) whose study shows an improvement in this ability between 9-month-olds and 12-month-olds and Best and Kitamura (2012) who show an improvement between 15-month-olds and 19-month-olds). These studies therefore indicate the role of accent in speech perception and speaker recognition from an early stage in a child’s linguistic and cognitive development.

However, beyond the age of infancy (if we take this roughly to mean post-18 months), very little has been done to investigate the further development of both speaker and accent recognition and the link between these perceptual skills in early childhood. While a certain level of familiar speaker recognition amongst pre-school children has been established (Bartholomeus 1973; Spence et al., 2002), an issue of central concern, which these studies do not address, is which particular aspects of the voice the children are using in order to recognise and identify a certain speaker. Both these studies use a large range of speakers (around 20 in each case) and utterances of 12 words or 4 seconds long. Therefore, each child is likely to have
heard a varied range of voice qualities, pitches and accent differences (amongst other speaker-specific distinctions) during these experiments. However these differences are not explored in either study and therefore there is no way of knowing whether any or all of these aspects of the voice were important in the children being able to identify the speakers.

The present study focuses on analysing speaker differences in addressing the question of pre-school children’s ability to identify familiar speakers and the role of regionally-based phonetic accent features in this process. Two related experiments were carried out in order to investigate the different aspects involved in familiar speaker recognition. First, experiment 1 was carried out in order to establish the level at which the pre-schoolers were able to identify familiar speakers. In this experiment, named the ‘Identification experiment’, the children were asked to identify familiar speakers in a voice-face match procedure, and the results address the following questions:

1) Are pre-school children from the age of 2.5 years able to identify familiar nursery teachers from short audio stimuli?
2) Does this ability improve with age throughout the pre-school years?
3) Is this ability affected by the amount of exposure the child has had to the speakers?
4) Does this ability improve when the children hear a longer length audio stimulus?
5) Are particular speakers, with more idiosyncratic features of voice such as their pitch or voice quality, more recognisable than others?

With the children’s ability to identify familiar speakers established from the results of experiment 1, experiment 2 leads on from this to look at the role of particular phonetic accent features in speaker recognition. In this experiment, named the ‘Recognition experiment’, children were asked to give yes/no answers as to whether the speaker was recognised as a particular nursery teacher or not. Following on from the Identification experiment, the Recognition experiment addresses the questions:

6) Are the same children able to recognise one particular familiar nursery teacher (with a strong regional accent) from single word stimuli?
7) Is this ability influenced by external social factors, such as the child’s age/sex or their exposure to different languages and/or accents at home?
8) Are the children able to distinguish other, unfamiliar speakers from this familiar speaker based on the same single word stimuli?
9) Are some unfamiliar speakers with different regional accents more easily distinguished from the familiar speaker than others?
10) Does a disguised accent result in a reduced ability to recognise the familiar speaker?

2. Previous work

Speaker identification relies on both segmental and suprasegmental linguistic cues such as a speaker’s pronunciation, pitch, prosody and voice quality (cf. Nolan, 1983). The extent to which these different cues are used depends on factors such as the wider context in which the speech is heard, the familiarity of the listener with the speaker, and the presence of idiosyncrasies in the voice of the speaker. Depending on how familiar the speaker is to the listener, social information associated with particular linguistic cues can be relied upon in the
identification process. A speaker may be instantly recognised as male or female, old or young, middle class or from a particular part of the country. This may, in turn, help a listener to identify a particular speaker who they know from that particular social group. Furthermore, this kind of ‘personal’ information has been found to be intrinsically linked to linguistic information. Nygaard and Pisoni (1998) found that ‘talker-specific information’ helped listeners in word learning tasks. Listeners performed better in intelligibility tests involving novel words when the words were spoken by speakers they had been familiarised with during the testing procedure. Therefore, there seems to be a strong likelihood that listeners store social information alongside linguistic information in the speaker encoding process. Identification of a more long-term familiar speaker, however, relies more heavily on processes of pattern recognition (Van Lancker & Kreiman, 1985, 1987). This kind of recognition and identification has been shown to activate different regions of the brain compared with unfamiliar speaker discrimination (Van Lancker & Kreiman, 1985, 1987; see Ohman, 2013 for an overview of such studies). Therefore it is important not to conflate findings from studies on familiar speaker recognition and those from speaker identification more generally. Overall, though, more exposure, more attention and/or recent exposure to a particular speaker results in a higher level of accuracy in speaker identification (Watt, 2010). Therefore it is unsurprising that it seems on balance a lot easier to identify speakers we are more familiar with.

In terms of the role that a speaker’s accent plays in the identification process, Nolan (1983) suggests that an accent cannot be properly defined outside of the speaker who uses it. He proposes that ‘personal’ and ‘accentual’ information in speech are intertwined with one another. Therefore an individual’s use of accent features depends on the style of speech they are employing. Thus, the more familiar we are with a speaker, the more likely we are to know their stylistic range. This tendency was investigated by Ladefoged and Ladefoged (1980), who tested one of the authors’ ability to identify familiar speakers from a sample of 29 friends, family and acquaintances. Overall he was able to identify 83% of the speakers with 30 second samples of speech but this was reduced to 66% when hearing just a single sentence, and to 31% when the single word ‘hello’ was heard. One surprising finding in this study was that Ladefoged failed to correctly identify his own mother from both the one-word and one-sentence samples. On reflection, Ladefoged highlights the role of accent here, as he struggled to differentiate his mother’s voice from those of the other similarly-accented RP speaking women in the sample. In this case, then, accent had a negative impact on his ability to identify a speaker relative to the other possible candidates in the task. Accent can have the opposite effect, however, and can make a speaker noticeable if they have a different accent to others in a group. In the same study, Ladefoged specifically refers to hearing ‘North Country vowels’ and then being able to identify the speaker in the sample he knew came from the North. Also, Foulkes and Barron (2000) found that within a group of 10 young male friends of diverse backgrounds, the two most consistently identified were those with the strongest accents (Tyneside and London) relative to the group. These findings suggest that accent is the most salient feature of a speaker’s voice in the recognition process.

Sjöström et al. (2009) present further evidence for the important role of accent in their study of a bidialectal speaker who was not recognised across his two Swedish dialects. Participants were familiarised with the speaker in one of his two dialects by listening to a reading passage. They were then asked to identify the speaker from a voice line-up, with four foil voices alongside the target heard saying phrase-length utterances. When the target speaker shifted dialects, listeners found it much more difficult to identify him than when he was heard speaking in the same dialect as the reading passage. Features of a speaker’s dialect or accent, therefore, do appear to be playing a role in the identification and recognition process. However, the questions
of which specific features and to what extent they contribute to identification have yet to be fully addressed.

Few studies have looked at young children’s abilities to identify familiar speakers. Bartholomewes (1973) tested 4-year-old nursery children in a voice-face matching task involving other children and staff at the nursery. During the testing procedure the children heard a speech sample and were asked to choose the corresponding picture of the speaker from a choice of 21. Each child took part in 17-19 trials altogether and overall the children scored just under 60% correct. Spence et al. (2002) found that 3-5-year-old children scored above chance in matching a cartoon voice with its picture. The children listened to 20 voice samples and chose from a closed set of 6 pictures after each stimulus. They found an improvement in this ability between 3-year-olds (61% correct overall) and 4-year-olds (81% correct overall). Also, they found that there was a significant improvement in line with the level of familiarity as the children were more likely to match the voice with its picture when the character was more familiar to them. Although these studies show a certain level of familiar speaker identification by young children, they do not further investigate what aspects of a familiar speaker’s voice the children are using in the identification/recognition process. Whether the speaker’s accent may play a role, like it has been shown to for adults, is still a question for further investigation.

The extent of the awareness of accent differences by children in this age group is a contentious issue. Nathan et al. (1998) found an improvement from a group of 4-year-olds to a group of 7-year-olds (all from London) in a task testing their lexical comprehension of accents. This was based on the children’s own reproductions of words they heard in a Glaswegian accent. Conversely, Floccia et al. (2009) found that 5-year-olds were unable to group speakers into a local vs. non-local group based on their accent. Children were asked to listen to sentence length utterances by speakers with an accent from their own region (West Country) and an ‘alien’ region (Ireland). They were then instructed to divide these speakers into a blue or a red group, depending on whether the speaker was from their home town (Plymouth) or was an alien from somewhere else. The 5-year-olds scored around chance level (just over 56%) in this task. However, some methodological issues with this study means that the ability of 5-year-olds to distinguish and group speakers on accent-based criteria cannot be completely ruled out. The stimuli heard by the children consisted of full-length sentences and therefore there is no way of knowing which particular linguistic cues the children were using to group the speakers. The length of the stimuli meant that the children were listening for content as well as segmental information, making the task cognitively demanding in terms of the processing and storing of information in their short term memory (Beck, 2014).

In an attempt to address these criticisms, Beck (2014) used a different methodology in her study of 5-7-year-olds. She looked at the ability of children in this age group to discriminate between a local and non-local accent. The children took part in an ABX discrimination task in which they heard a token word, either in their native Philadelphian or a General Southern (US) accent. They then heard two more tokens of the same word (one from each accent) and were asked to choose which one sounded most like the first. The children scored above chance, with 64% correct answers overall. Beck’s study shows a discriminatory ability of at least some of these young children. How this ability may play a role in identifying familiar speakers is a further question that the present study aims to address. The question is whether children use an ability to differentiate a familiar accent from an unfamiliar one in the process of identifying a familiar speaker, as adults have been shown to do to a certain extent.
3. Identification experiment

3.1 Methodology

3.1.1. Participants

22 children aged between 2 and 4 years (2;4 to 4;10) all attendees at the same nursery in York, in the north of England, were tested. Three children were excluded due to them not responding adequately\(^1\) during the experiment, leaving 19 altogether.

3.1.2 Background questionnaire

The children’s parents were asked to answer some background questions about the languages and accents their child was exposed to at home, the length of time the parents had lived in the city of York, and how many days/hours a week their child attended the nursery.

3.1.3 Speakers and stimuli

Stimuli for the experiment were taken from all six\(^2\) nursery teachers. The teachers all worked regularly at the nursery and so were in contact with all of the children. They were all females, aged 21-48; five had lived in York all their lives or came from elsewhere in Yorkshire (Knaresborough, Leeds). The exception was one teacher (Jane\(^3\)) who had moved around the country quite a lot, living in the South of England and the Midlands for different periods of time.

All the teachers were recorded in a quiet room at the nursery using a Zoom H4n recorder which was set to record at a 32bit 96kHz sampling rate. All speakers were recorded producing the same stimuli, which consisted of a story passage and 15 short phrases which they were asked to read as naturally as possible in a relaxed, informal style. The story passage consisted of 177 words from the start of the children’s book *The Gruffalo* (Donaldson and Scheffler, 1999), a book regularly read to the children at the nursery and written in rhyming couplets such as (1).

\[1\] A mouse took a stroll through the deep dark wood.
   A fox saw the mouse and the mouse looked good.

The 15 phrases that were also recorded were chosen as examples of the kind of phrases the nursery teachers use on a daily basis at the nursery. They ranged from 1 to 14 syllables (see Table 1). This was in order to be able to investigate the potential effect that the length of the stimulus might have on the children’s ability to identify the speaker (research question 4).

---

\(^1\) These three children gave no response to the stimuli even after several efforts from the experimenter to help them understand and engage with the task. Therefore their experiments were stopped after attempting the first few stimuli.

\(^2\) Recording of a seventh nursery teacher was also taken but her sample was discarded after pilot testing the experiment. She left the nursery part-way through the experimental design process and children in the pilot study failed to recognise her, even from her picture.

\(^3\) Pseudonyms are used throughout
3.1.4 Experimental design

Audacity software (Audacity Team, 2012) was used to edit the sound files into the stimuli. The Gruffalo passage was divided into seven separate and continuous stimuli, with one stimulus taken from each nursery teacher (apart from Jane who had two stimuli). Also, from the phrases recorded, one phrase was taken from each nursery teacher (apart from Leanne) to be used as stimuli (see Table 1).

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Phrase</th>
<th>Nursery teacher</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gruffalo 1</td>
<td>Jane</td>
</tr>
<tr>
<td>2</td>
<td>Gruffalo 2</td>
<td>Wynne</td>
</tr>
<tr>
<td>3</td>
<td>Gruffalo 3</td>
<td>Leanne</td>
</tr>
<tr>
<td>4</td>
<td>Gruffalo 4</td>
<td>Kristina</td>
</tr>
<tr>
<td>5</td>
<td>Gruffalo 5</td>
<td>Alice</td>
</tr>
<tr>
<td>6</td>
<td>Gruffalo 6</td>
<td>Claire</td>
</tr>
<tr>
<td>7</td>
<td>Gruffalo 7</td>
<td>Jane</td>
</tr>
<tr>
<td>8</td>
<td>Hello, Nursery?</td>
<td>Leanne</td>
</tr>
<tr>
<td>9</td>
<td>Line up to wash your hands</td>
<td>Alice</td>
</tr>
<tr>
<td>10</td>
<td>Daddy’s here to collect you</td>
<td>Claire</td>
</tr>
<tr>
<td>11</td>
<td>There’s a good girl</td>
<td>Jane</td>
</tr>
<tr>
<td>12</td>
<td>Where’s teddy?</td>
<td>Wynne</td>
</tr>
<tr>
<td>13</td>
<td>No!</td>
<td>Alice</td>
</tr>
<tr>
<td>14</td>
<td>Bye!</td>
<td>Leanne</td>
</tr>
</tbody>
</table>

Table 1: Stimuli and speakers for the Identification experiment

The experiment was constructed using the psychology software Psychopy (Peirce, 2007) and was designed in order to be run on the experimenter’s laptop. Each audio stimulus was presented separately while a picture relating to its content was displayed on screen. This was in order to keep the task entertaining for the children, therefore retaining their attention. Also, a picture of all the nursery workers was displayed at the bottom of the screen throughout the experiment (see Figure 1). The experiment was designed to be manually controlled by the experimenter and each new visual stimulus was displayed when a key was pressed on the keyboard, with the audio starting after one second each time. The responses (including any null responses) were logged by the pressing of a corresponding answer key. The stimuli were presented in the same order for each child as they followed a chronological order.

---

4 Jane’s second Gruffalo passage and Leanne’s extra phrase replaced the missing seventh nursery teacher’s stimuli.
5 The original list of nursery phrase stimuli was shortened (from 15 to 7) as piloting the experiment revealed that the children found it difficult to concentrate for the full length of time.
3.1.5 Experimental procedure

The children were tested individually in a quiet corner of the nursery, on the experimenter’s laptop. Headphones designed for use by children (JVC HA-KD5-Y –E) were worn by the children for the duration of the experiment. The experimenter also wore headphones (SONY MDRXB400W) in order to monitor the experiment. Additionally, the experimenter had a microphone to talk through which was connected to the computer via an amplifier mixer. This allowed the experimenter to be heard by the child in order to give instructions and prompt him or her for a response if necessary. The use of this audio equipment also helped to minimize distraction from the rest of the room.

Each child was asked to sit at the computer with the experimenter and was given the instructions ‘You are going to hear your teachers reading a story and I would like you to point at the picture of who you think is talking.’ As the child and the experimenter were wearing headphones, this allowed the child to indicate a response without having to speak. Also, this ensured that s/he was matching voices and faces, a much easier task than having to name the speaker outright (as found by Bartholomeus, 1973).

Before the experiment started, the experimenter checked that the child recognised all of the nursery teachers by pointing at their pictures and asking the child to name them. This ensured that all of the children reached a baseline level of familiarity with the speakers. The experiment was started and the first audio clip played. If the child did not initiate pointing at one of the nursery teachers, the experimenter prompted them: ‘Who was that speaking? Can you point to their picture?’

The child’s responses were recorded on to the computer by the experimenter and then the child was asked to do the same again with the rest of the stimuli. At the end of this task, the headphones were removed in order to give further instructions for the Recognition experiment (see section 4 below).
3.2 Identification experiment results

3.2.1 Results overall

Table 2 displays each child’s overall number of correct answers for the Identification experiment. Background information pertaining to the amount of time the children have spent at the nursery as well as the languages spoken at home is also shown in the table.

Overall, the children’s mean score was 8.53 out of 14 (SD = 3.64), equating to 60.9% accuracy (chance = 16.7%).

<table>
<thead>
<tr>
<th>Child</th>
<th>Age (years)</th>
<th>Correct answers (/14)</th>
<th>Correct answers (%)</th>
<th>Years at nursery</th>
<th>Hours a week at nursery</th>
<th>Languages spoken at home</th>
</tr>
</thead>
<tbody>
<tr>
<td>F9</td>
<td>3.8</td>
<td>14</td>
<td>100.0</td>
<td>3</td>
<td>45</td>
<td>2</td>
</tr>
<tr>
<td>M7</td>
<td>4</td>
<td>13</td>
<td>92.9</td>
<td>2</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>F10</td>
<td>4.9</td>
<td>12</td>
<td>85.7</td>
<td>3</td>
<td>36</td>
<td>2</td>
</tr>
<tr>
<td>F12</td>
<td>3.8</td>
<td>12</td>
<td>85.7</td>
<td>4</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>F8</td>
<td>4.8</td>
<td>12</td>
<td>85.7</td>
<td>1</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>F11</td>
<td>2.4</td>
<td>10</td>
<td>71.4</td>
<td>2</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>F13</td>
<td>2.8</td>
<td>10</td>
<td>71.4</td>
<td>4</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>F2</td>
<td>3.8</td>
<td>10</td>
<td>71.4</td>
<td>2</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>F6</td>
<td>4.6</td>
<td>9</td>
<td>64.3</td>
<td>3</td>
<td>32</td>
<td>2</td>
</tr>
<tr>
<td>M1</td>
<td>3.6</td>
<td>9</td>
<td>64.3</td>
<td>2</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>M4</td>
<td>3.6</td>
<td>9</td>
<td>64.3</td>
<td>3</td>
<td>32</td>
<td>2</td>
</tr>
<tr>
<td>F3</td>
<td>3</td>
<td>8</td>
<td>57.1</td>
<td>2</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>M5</td>
<td>4.3</td>
<td>7</td>
<td>50.0</td>
<td>3</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>F4</td>
<td>4.5</td>
<td>6</td>
<td>42.9</td>
<td>2.5</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>M8</td>
<td>4</td>
<td>6</td>
<td>42.9</td>
<td>2</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>M6</td>
<td>2.7</td>
<td>3</td>
<td>21.4</td>
<td>2</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>M2</td>
<td>2.5</td>
<td>2</td>
<td>14.3</td>
<td>1.5</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>F7</td>
<td>3.3</td>
<td>1</td>
<td>7.1</td>
<td>1.5</td>
<td>27</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 2: Identification experiment results (scores in rank order) and background information for each child: F=Female, M=Male

3.2.2 Statistical analysis of Identification experiment results

A stepwise backward regression method was used in a binary, mixed effects logistic model, run in R using the lme4 library (R Core Team, 2013). A logistic model is used when there are two possible outcomes, in this case a correct or incorrect answer, in order to show the likelihood that the predictor variables entered into the model predict a particular outcome (Baayen, 2008). A mixed effects model includes random variables as well as fixed predictor variables in order to account for individual variation. In this model, each individual’s responses are represented by their own coefficient, ensuring that the potential for one particular individual’s results to warp the overall results is greatly reduced (Drager, 2011).
The children’s responses, correct or incorrect, were turned into a factor so that each response from each individual was used in the model analysis. The dependent variable was therefore a binomial factor distinguishing between correct and incorrect answers, with the default set to correct answers. This means that the model presents the log odds of a correct answer. Independent, fixed predictor variables of stimulus length, years at nursery and hours at nursery were included in the model. All of these were measured and entered as continuous variables. Additionally, child was treated as a random factor in the model. Two-way interactions between all the fixed predictors were also included. Age was not included as a fixed variable as it was found to correlate highly with the children’s years spent at nursery. Therefore the effects of these predictors cancelled each other out when entered into the model together. As the children’s years spent at nursery was a more explanatory predictor in pre-testing (as judged by both a lower AIC and BIC which tests the likelihood of a model being closer to the truth based on both goodness-of-fit and the complexity of the model), this was included in the model at the expense of age.

Only the interaction between stimuli length and hours spent at the nursery did not significantly affect the predictive power of the model. Therefore this was the only interaction removed from the final model (see Table 3).

|                                    | Estimate | Std. Error | z value | Pr(>|z|) | Sig. |
|------------------------------------|----------|------------|---------|----------|------|
| (Intercept)                        | 5.810    | 3.294      | 1.764   | 0.078    | .    |
| STIMULUS LENGTH                    | -0.494   | 0.265      | -1.865  | 0.062    | .    |
| YEARS SPENT AT NURSERY            | -2.709   | 1.341      | -2.020  | 0.043    | *    |
| HOURS A WEEK AT NURSERY           | -0.269   | 0.144      | -1.862  | 0.063    | .    |
| STIMULUS LENGTH*YEARS              | 0.334    | 0.124      | 2.706   | 0.007    | **   |
| YEARS*HOURS                        | 0.116    | 0.057      | 2.049   | 0.041    | *    |

Table 3: Mixed effects logistic regression model fit to the data for the Identification experiment
(Significance level: ‘.’ = 0.1, ‘*’= 0.05, ‘**’ = 0.01)

Table 3 shows the effect of the dependent variables on the log odds of the children giving a correct answer. The variables of stimulus length and years at nursery enter into a significant interaction, as do years at nursery and hours spent at nursery every week. Figure 2 plots the significant interaction between the length of the stimulus and the number of years that the children have been attending the nursery. As the majority of the children had attended the nursery for 2, 3 or 4 years, these are the years plotted. This shows that while both a longer stimulus and more years at nursery result in a higher probability of a correct answer from the children, these are not independent of one another. As the number of years the children have been attending the nursery increases, the length of the stimulus has a stronger effect on their predicted ability. In other words, for the lowest performing younger children who have been attending the nursery for fewer years, the length of the stimulus has some effect. This effect is strengthened for the older children, who have been attending the nursery for more years and who perform better overall.
Pre-school children’s identification of familiar speakers and the role of accent features

Figure 2: The interaction between stimulus length and the years the child has been attending the nursery.

Figure 3 plots the significant interaction between the number of years that the children have been attending the nursery and the number of hours a week that they attend the nursery. The line plots the predicted probability of a correct answer. For example, 0 means an incorrect answer is predicted and 1 that a correct answer is predicted, dependent on the variables of the child’s years at nursery and the hours a week they spend at nursery. Those who have attended the nursery for longer are more likely to give a correct answer overall. Of the children who have been attending the nursery for 3 or 4 years, those who spend more hours per week at the nursery have a higher probability of achieving a correct answer in this task. The children who have been attending the nursery for 2 years however are not showing the same pattern of improvement, possibly due to a couple of low-scoring individuals driving this effect (cf. M2 and M6 in Table 2).

Figure 3: The interaction between the number of hours the children attend the nursery a week and the number of years they have been attending the nursery altogether.
3.2.3 Speaker effects

The rate of identification of each individual speaker was further analysed. This was in order to address research question (5).

Certain voices were more readily recognised by the children than others. Table 4 shows the number of children who correctly identified each speaker for each of their stimuli.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Number of children correct (/19)</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>Jane</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>Jane</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>Kristina</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>Alice</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>Leanne</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
<td>Claire</td>
</tr>
<tr>
<td>7</td>
<td>13</td>
<td>Claire</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>Jane</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>Leanne</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>Alice</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>Leanne</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>Wynne</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
<td>Wynne</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>Alice</td>
</tr>
</tbody>
</table>

Table 4: Number of children with correct answers for each speaker

Excluding Kristina, as she had only one stimulus token as opposed to two or three from all the others, Jane was the most often recognised with an average of 14.67 correct responses. Wynne was the least well recognised with an average of 6.5 correct responses. As Jane’s voice stood out as being much lower in pitch than the others, this was taken as the basis for further investigation. The teachers’ fundamental frequencies were measured to explore the effect of this feature in more detail across the whole group of speakers (cf. Foulkes and Barron, 2000 and Compton, 1963 for previous work showing f0 to be one of the most readily distinguishable characteristics of a voice). F0 was measured by calculating each speaker’s mean fundamental frequency across all of their stimuli used in the experiment.

The box plot in Figure 4 shows each speaker’s average f0 (in Hz). From this figure it is clear that Jane’s average f0 is conspicuous. Whereas Jane’s f0 falls near the typical male average of 120Hz, the other speakers are around the female average of 225Hz (cf. Fry, 1979: 68). Although Leanne's pitch range is also noticeably different to the others, as it is above rather than below the average pitch for a woman, it is arguably less conspicuous as her voice is obviously female sounding.
Jane also has a very distinctive voice quality; which under Laver’s VPA protocol might be classified as a deep, harsh, whispery, creaky voice (Laver, 1968: 48). Therefore altogether it appears that Jane has a relatively idiosyncratic f0 and voice quality, which can both be predicted to mark her out as easier to identify than the other speakers.

3.2.4 Discussion of Identification experiment results

In addressing the research questions (1)–(5), results from the Identification experiment show that pre-school children are able to identify six individual familiar speakers with a high rate of accuracy. Overall, they performed well above chance level (with only two children below chance) and at a very similar level to the children in Spence et al.’s (2002) study. Also, similarly to Spence et al., there appears to be a developmental effect, as the older children outperform the younger children. However, this improvement seems to be mostly affected by the amount of exposure the children have had to the speakers. This exposure is both in terms of the years that they have been attending the nursery combined with their regular attendance at the nursery on a weekly basis. Not surprisingly, both of these exposure criteria have an effect, in line with the obvious prediction that more exposure to a speaker results in better recognition rates (e.g. Watt, 2010). Spence et al. showed this to be true for children in terms of how familiar they appeared to be with cartoon characters (e.g. could they name the characters without prompting). The present experiment, however, shows this a little more rigorously though the measurement of exposure.

The children’s ability to identify the correct familiar speaker improves with a longer audio stimulus. This confirms earlier findings from adults (Ladefoged and Ladefoged, 1980; Schweinberger et al., 1997). In turn, this finding suggests that better identification relies on a number of different cues, as a longer stimulus is also likely to provide more acoustic indications pertaining to the identity of the speaker. In terms of these cues, f0 and voice quality were found to play a role in the success of the best identified teacher. Jane’s overall f0 and voice quality were very distinct from the other five nursery teachers, demonstrating the importance of idiosyncrasies in a speaker’s voice in the identification process.

The next question this research addresses is whether accent is also a factor in recognising a familiar speaker from shorter stimuli.
4. Recognition experiment

4.1 Methodology

4.1.1 Participants
The same children from the Identification experiment were tested. The youngest two children were excluded due to not responding to this second experiment. This left 17 children (aged 2.7-4.9 years) altogether.

4.1.2 Stimuli and speakers
Stimuli used for the experiment were taken from recordings of one of the nursery teachers (Alice) and nine other speakers, unknown to the child participants. These nine speakers were a mixture of females (7) and males (2), aged 20-29 with different regional accents. Alice was chosen as she was the teacher judged impressionistically to have the strongest Yorkshire accent. For example, she had a consistently monophthongal vowel quality in the GOAT and FACE lexical sets (Wells, 1982). Therefore, being the most distinctly ‘Yorkshire’, her accent provided the clearest benchmark for comparison with the other broad accents of the unfamiliar speakers. Also, she was of a comparable age (21 years) to the unfamiliar speakers recorded. The speakers were all recorded reading the same story passage and phrases as those in the Identification experiment.

4.1.3 Experimental design
Eight words were chosen from the recordings to use as stimuli. These words captured differences between the Yorkshire accent and the accents of the other speakers recorded. Seven of these were based on vowel quality differences and one based on the rhotic/non-rhotic accent distinction. These words were chosen based on the main differences between the regional accents of the unfamiliar speakers compared to Alice. The decision of which words to use was also largely based on the quality and clarity of recordings once the individual words were extracted and heard in isolation.

For each word, three tokens were extracted as stimuli: one token from Alice and one each from two unfamiliar speakers with different accents. Seven of the other speakers had distinctly different regional accents compared with Alice (two Standard Southern British English, one North American (Californian), one North East of England, two Northern Irish, one Scottish) while two of the speakers had similar accents to Alice’s. These two were another Yorkshire speaker and a speaker from Lancashire, which, as another area of the central north, shares many phonetic based accent features with Yorkshire (Hughes et al., 2012). Stimuli taken from these two speakers were therefore used in order to investigate whether the similarity in their pronunciation to Alice’s would cause the children to misidentify these speakers as Alice.

Additionally, Alice was recorded saying two of the stimuli for a second time (gruffalo and coat). She was asked to disguise her accent by pronouncing these words with an accent from the South of England, using the diphthong [əʊ] as opposed to the monophthong [o:]. These recordings were taken in order to investigate whether Alice would be miscategorised as an unfamiliar speaker when using the phonetic realisation of a different accent from her own. Table 5 shows the full list of stimuli and each speaker’s phonetic realisation of the vowel
quality/consonant distinction under consideration. Alice’s disguised tokens are indicated with *.

<table>
<thead>
<tr>
<th>Word</th>
<th>Associated lexical set/ feature of accent</th>
<th>Alice: Phonetic realisation</th>
<th>Unfamiliar speaker 1: Accent and phonetic realisation</th>
<th>Unfamiliar speaker 2: Accent and phonetic realisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>gruffalo</td>
<td>GOAT [ɔː]</td>
<td>SSBE M [ɔː]</td>
<td>Lancashire F [ɔː]</td>
<td></td>
</tr>
<tr>
<td>never mind</td>
<td>rhoticity no /t/</td>
<td>Northern Irish F /t/ realisation</td>
<td>American F /t/ realisation</td>
<td></td>
</tr>
<tr>
<td>know</td>
<td>GOAT [ɔː]</td>
<td>SSBE F [ɔː]</td>
<td>North East F [ɔː]</td>
<td></td>
</tr>
<tr>
<td>mouse</td>
<td>MOUTH [ɒʊ]</td>
<td>Northern Irish F [au]</td>
<td>Scottish F [au]</td>
<td></td>
</tr>
<tr>
<td>food</td>
<td>GOOSE [ʊː]</td>
<td>Northern Irish F [u]</td>
<td>Scottish F [u]</td>
<td></td>
</tr>
<tr>
<td>coat</td>
<td>GOAT [ɔː] *</td>
<td>SSBE F [ɔː]</td>
<td>Yorkshire F [ɔː]</td>
<td></td>
</tr>
<tr>
<td>good</td>
<td>FOOT [o]</td>
<td>Northern Irish F [u]</td>
<td>Scottish F [u]</td>
<td></td>
</tr>
<tr>
<td>gruffalo</td>
<td>GOAT [ɔː] *</td>
<td>SSBE F [ɔː]</td>
<td>Yorkshire F [ɔː]</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Stimuli used in the Recognition experiment, along with each speaker’s phonetic realisation of the vowel/consonant distinction. SSBE= Standard Southern British English, M = male, F= female

4.1.4 Experimental procedure

After finishing the Identification experiment and in preparation for the Recognition experiment, the children were told, ‘Now you are going to hear some more people speaking. It might be Alice that you hear, or it might be someone that you don’t know. Each time you hear a voice, I would like you to tell me whether you think it is Alice talking or not. If you think it is Alice, say ‘Alice’ but if you think it is someone else that you don’t know, say ‘No’, okay? Let’s have a go.’ The headphones were then again placed over the child’s head.

A picture of Alice (with the headline ‘Is it Alice?’) appeared on screen to remind the children who they were listening out for. Then the screen was changed to a picture of a teddy bear and the child heard the first stimulus. They were asked ‘Was that Alice speaking?’ and their response was logged by the experimenter. Next, a different teddy bear appeared on screen and a different speaker was heard saying the same stimulus. Again, the children were asked if this was Alice speaking. The process was repeated a third time with a third teddy bear and a third speaker (see Figure 5). This whole process was then repeated for each set of three stimuli. Two versions of the experiment were created, with different orders of the speakers in each set of three; 8 children took version 1 of the experiment and 9 different children took version 2. Children were rewarded with stickers for playing the game.

---

6 No trial runs were included in the experiment as piloting the experiment found that the children were able to understand the task but struggled to concentrate if the task was any longer than the final version used for the experiment.
4.2 Recognition experiment results

4.2.1. Results overall

Table 6 presents each child’s results along with relevant background information. Overall, the children’s mean score was 21.53 out of 33 (SD = 6.06), equating to 65.1% accuracy (chance = 50%).

<table>
<thead>
<tr>
<th>Child</th>
<th>Age (years)</th>
<th>Correct answers (/33)</th>
<th>Correct answers (%)</th>
<th>Years at nursery</th>
<th>Hours a week at nursery</th>
<th>Languages spoken at home</th>
<th>Yorkshire parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>F9</td>
<td>3.8</td>
<td>30</td>
<td>90.9</td>
<td>3</td>
<td>45</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>F2</td>
<td>3.8</td>
<td>29</td>
<td>87.9</td>
<td>4</td>
<td>24</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>F13</td>
<td>2.8</td>
<td>28</td>
<td>84.8</td>
<td>2</td>
<td>16</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>F10</td>
<td>4.9</td>
<td>27</td>
<td>81.8</td>
<td>2</td>
<td>15</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>F3</td>
<td>3.0</td>
<td>27</td>
<td>81.8</td>
<td>2</td>
<td>22</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>M7</td>
<td>4.0</td>
<td>27</td>
<td>81.8</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>F8</td>
<td>4.8</td>
<td>26</td>
<td>78.8</td>
<td>4</td>
<td>17</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>F12</td>
<td>3.8</td>
<td>24</td>
<td>72.7</td>
<td>3</td>
<td>36</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>F6</td>
<td>4.6</td>
<td>21</td>
<td>63.6</td>
<td>2</td>
<td>10</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>M5</td>
<td>4.3</td>
<td>20</td>
<td>60.6</td>
<td>3</td>
<td>15</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>M1</td>
<td>3.6</td>
<td>17</td>
<td>51.5</td>
<td>3</td>
<td>32</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>M8</td>
<td>4.0</td>
<td>17</td>
<td>51.5</td>
<td>2</td>
<td>24</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>F4</td>
<td>4.5</td>
<td>16</td>
<td>48.5</td>
<td>2.5</td>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>M3</td>
<td>3.6</td>
<td>16</td>
<td>48.5</td>
<td>2</td>
<td>20</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>M4</td>
<td>4.3</td>
<td>16</td>
<td>48.5</td>
<td>2</td>
<td>9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>F7</td>
<td>3.3</td>
<td>14</td>
<td>42.4</td>
<td>1.5</td>
<td>27</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>M6</td>
<td>2.7</td>
<td>11</td>
<td>33.3</td>
<td>2</td>
<td>17</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 6: recognition experiment results and background information for each child
A post hoc comparison showed no difference between the two versions of the experiment, therefore the results from both versions were combined for the analysis. A significant positive correlation was found between the children’s scores for the Identification experiment and those for the Recognition experiment ($r = 0.641$, $p < 0.01$), indicating that those who performed well in the first experiment were more likely to score well in the second.

In addressing research question (6), the overall results show that the children’s performance was above chance and therefore to some degree they were able to recognise Alice and distinguish the other unfamiliar speakers from Alice. Questions (8)-(10) will be further investigated in section 4.2.3 which will take a closer look at the cases in which Alice was missed and unfamiliar speakers were misidentified as Alice. First, question (7) (repeated below) is addressed in the next section.

(7) Is the children’s ability to recognise one familiar nursery teacher influenced by external social factors, such as their age/sex or their exposure to different languages/accents at home?

4.2.2 Statistical analysis

As with the Identification experiment, statistical analyses were carried out in R. A stepwise backward regression method was used in a binary, mixed effects model. Again, the dependent variable was a binomial factor distinguishing between correct and incorrect answers, with the default set to correct answers. Main effects predictors of age, sex and Yorkshire parentage were included in the model. Age was measured and entered into the model as a continuous variable. For the Yorkshire parent predictor, the default was set to no Yorkshire parents and for the sex predictor, the default was set to female. Therefore the coefficient in the ‘estimate’ column shows the effect of each of the default dependent variables on the log odds of the children scoring a correct answer. Two-way interactions between all the main effect predictors were also included. Only the interaction between age and Yorkshire parents was not found to significantly alter the predicting power of the model. Therefore this was the only interaction removed from the final model (see Table 7 below). As the interactions between the predictor variables are found to be significant, the main effects are interpreted as part of these interactions rather than individually.

|                          | Estimate | Std. Error | z value | Pr(>|z|)   | Sig. |
|--------------------------|----------|------------|---------|------------|------|
| (Intercept)              | 4.41     | 1.31       | 3.37    | 0.0007***  | ***  |
| AGE(Years)               | -0.70    | 0.30       | -2.37   | 0.02*      | *    |
| SEX(Male)                | -6.65    | 2.02       | -3.3    | 0.00097*** | ***  |
| YORKSHIREPARENT(True)   | -1.39    | 0.43       | -3.22   | 0.0013**   | **   |
| AGE(Years):SEX(Male)     | 1.27     | 0.53       | 2.4     | 0.02*      | *    |
| SEX(Male): YORKSHIREPARENT(True) | 1.71     | 0.66       | 2.61    | 0.009**    | **   |

Table 7: Mixed effects logistic regression model fit to the data for the Recognition experiment

(Significance level: ‘*’= 0.05, ‘**’ = 0.01, ‘***’ = 0.001)

Table 7 shows that the final model finds two significant interactions. Figure 6 plots the significant interaction between the children’s sex and whether they have any parents from Yorkshire. The model shows that girls generally outperform boys but this is also dependent on whether they have parents from Yorkshire. The model predicts that for girls if they have at

7 Due to the limited number of participants in the experiment, only three predictors were included in the model. The predictor pertaining to languages spoken at home was left out of the model as this did not show a strong or significant correlation with the children’s correct answers. Additionally the children who came from bilingual homes varied in terms of their exposure and fluency with the other languages themselves.
least one parent from Yorkshire, the likelihood of getting a correct answer decreases. For boys the opposite is true but this is over a much smaller range and their probability of getting a correct answer stays below 0.6.

![Figure 6: Interaction of sex and Yorkshire parents](image)

Figure 6 shows the interaction of sex and Yorkshire parents. Older boys are predicted by the model to perform better than younger boys. A higher performance is maintained for the girls, although there is a slight drop over the ages. This is probably caused by a few individuals as there is a particularly high scoring 2-year-old and 3-year-old (F13 and F3) and a low scoring 4-year-old (F4) (see Table 6). Interpreting Figures 6 and 7 together it is clear that while sex is an important predictor it is not something that can be analysed independently of other factors. With a small sample size it is impossible to be conclusive but it appears that girls are more advanced generally and while age affects the boys’ results, exposure to other accents has more of an effect on the girls’ results.

![Figure 7: Interaction of age and sex](image)
4.2.3 Misses vs. false alarms

In further analysing the results from the Recognition experiment, the seven children who scored below chance level (50%) were excluded⁸. This leaves the results of ten children to examine further.

As Table 8 shows, children were more likely to miss the familiar speaker (miss) than wrongly identify an unfamiliar speaker (false alarm).

<table>
<thead>
<tr>
<th>Error type</th>
<th>Number of tokens</th>
<th>Percentage of tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miss</td>
<td>44/110</td>
<td>40%</td>
</tr>
<tr>
<td>False alarm</td>
<td>28/220</td>
<td>12.7%</td>
</tr>
</tbody>
</table>

Table 8: Number and percentage of error types for the Recognition experiment

4.2.3.1 False alarms

Table 8 shows that the overall false alarm rate is low. In answering research question (8) this indicates that, in general, the children are able to distinguish the unfamiliar speakers from Alice. However, in addressing question (9), it is noticeable that the speakers with the highest number of false alarms are the Yorkshire and Lancashire female speakers.

<table>
<thead>
<tr>
<th>Lexical set</th>
<th>Word</th>
<th>Unfamiliar speaker 1: phonetic realisation</th>
<th>No. of false alarms</th>
<th>Unfamiliar speaker 2: phonetic realisation</th>
<th>No. of false alarms</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOAT</td>
<td>graffalo</td>
<td>SSBE M [əʊ]</td>
<td>1</td>
<td>Lancs F [əʊ]</td>
<td>3</td>
</tr>
<tr>
<td>rhoticity</td>
<td>never mind</td>
<td>N Irish F [ɪ]</td>
<td>2</td>
<td>American F [ɪ]</td>
<td>1</td>
</tr>
<tr>
<td>GOAT</td>
<td>know</td>
<td>SSBE F [ʊ]</td>
<td>1</td>
<td>North East F [oː]</td>
<td>0</td>
</tr>
<tr>
<td>LOT</td>
<td>fox</td>
<td>Scottish F [ʊ]</td>
<td>1</td>
<td>American F [uː]</td>
<td>1</td>
</tr>
<tr>
<td>MOUTH</td>
<td>mouse</td>
<td>N Irish F [au]</td>
<td>1</td>
<td>Scottish F [au]</td>
<td>1</td>
</tr>
<tr>
<td>GOOSE</td>
<td>food</td>
<td>N Irish F [u]</td>
<td>2</td>
<td>Scottish F [u]</td>
<td>0</td>
</tr>
<tr>
<td>GOAT</td>
<td>coat</td>
<td>SSBE F [əʊ]</td>
<td>2</td>
<td>Yorkshire F [oː]</td>
<td>3</td>
</tr>
<tr>
<td>FOOT</td>
<td>good</td>
<td>N Irish F [u]</td>
<td>2</td>
<td>Scottish F [u]</td>
<td>1</td>
</tr>
<tr>
<td>GOAT</td>
<td>graffalo</td>
<td>Yorks F [oː]</td>
<td>3</td>
<td>SSBE F [əʊ]</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 9: False alarms for each unfamiliar speaker

In relation to question (7), this suggests that the children were more likely to misidentify an unfamiliar speaker as Alice if they heard a female speaker with a similar regional accent and

---

⁸ As this part of the analysis was based on individual responses and due to the relatively low number of children who took part in the Recognition experiment, this was a decision made by the experimenter, after careful consideration of each child’s performance and whether their contribution could justifiably be seen as showing an understanding of the task.
vowel pronunciation. Although the Lancashire female pronounced the GOAT vowel with a diphthong rather than a monophthong, this diphthong has less formant movement in the offglide than the diphthongs pronounced by the other unfamiliar speakers. This is evident by comparing the Euclidean distance between the F1 and F2 formants of all the stimuli featuring a GOAT vowel. The Euclidean distance between the nucleus value and the offglide value was calculated using the equation in (11):

$$11) \sqrt{(F1_{nucleus} - F1_{offglide})^2 + (F2_{nucleus} - F2_{offglide})^2}$$

The vowels were hand measured at the 20% and 80% vowel duration points in Praat (see Figure 8). Figure 8 shows that the SSBE speakers had a high level of offglide movement, suggestive of highly diphthongal [əʊ] vowels. The familiar speaker had a much lower level of movement, indicating a monophthongal vowel. As well as monophthongal, this vowel is a fronted GOAT vowel, sounding more like [ɵ:] than its backed equivalent [o:] (see Watt & Tillotson, 2001 and Haddican et al., 2013 for accounts of GOAT fronting in Yorkshire English). The unfamiliar Yorkshire speaker’s GOAT vowels were similarly monophthongal and fronted and while the Lancashire speaker’s GOAT vowel showed more movement, this was still less than that produced by the SSBE speakers. The North East speaker also produced a monophthongal with little movement, but this was not a fronted GOAT vowel, rather the back [o:] vowel. Therefore, this distinction meant that she was not mistaken for the familiar speaker. The high amount of diphthongal movement of the familiar speaker’s disguised GOAT tokens most likely accounts for why she is not correctly identified from these tokens by most of the children (this is further explored below).

Figure 8: Euclidean distance between the F1 and F2 nucleus and offglide values of GOAT vowels

---

9 See Haddican et al. (2013) for a discussion of the social awareness of GOAT fronting in York
4.2.3.2 Misses

Table 10 shows the number of misses for each word spoken by the familiar speaker. The words for which a disguised accent was used are indicated with *. Apart from good,10 these words result in the most number of misses. Whereas none of the children miss Alice when she used a monophthongal GOAT vowel (in gruffalo), the diphthongal quality of her GOAT vowel in both the disguised words resulted in many misses. Therefore, an accent disguise appears, as predicted, to inhibit recognition for most of these children.

<table>
<thead>
<tr>
<th>Word</th>
<th>Feature</th>
<th>Pronunciation</th>
<th>Misses (/10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>good</td>
<td>FOOT</td>
<td>[ʊ]</td>
<td>9</td>
</tr>
<tr>
<td>*coat</td>
<td>GOAT</td>
<td>[əʊ]</td>
<td>8</td>
</tr>
<tr>
<td>*gruffalo</td>
<td>GOAT</td>
<td>[əʊ]</td>
<td>7</td>
</tr>
<tr>
<td>food</td>
<td>GOOSE</td>
<td>[u:]</td>
<td>6</td>
</tr>
<tr>
<td>know</td>
<td>GOAT</td>
<td>[o:]</td>
<td>4</td>
</tr>
<tr>
<td>mouse</td>
<td>MOUTH</td>
<td>[au]</td>
<td>3</td>
</tr>
<tr>
<td>fox</td>
<td>LOT</td>
<td>[ɒ]</td>
<td>1</td>
</tr>
<tr>
<td>never mind</td>
<td>RHOTIC</td>
<td>No /r/</td>
<td>1</td>
</tr>
<tr>
<td>gruffalo</td>
<td>GOAT</td>
<td>[o:]</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 10: Misses for each word from the familiar speaker

4.2.4 Discussion of Recognition experiment results

In addressing research questions (6)-(10), results from the Recognition experiment show that pre-school children are mixed in their abilities in what was expected to be a difficult and unusual task for them. Some children were able to recognise the familiar speaker with a strong Yorkshire accent from single word stimuli while others were not. This ability is somewhat affected by external social factors. The girls generally perform better than the boys but the sexes are affected differently by external factors. The boys improve with age, indicating an important developmental effect. This development could be in terms of the understanding of the task, their level of concentration and/or their amount of exposure to the familiar speaker. Girls who have neither parent from the region performed better than those with at least one Yorkshire parent. This suggests that those who have had more exposure to different accents are better able to distinguish different pronunciations from each other. Both these findings between the sexes and whether these abilities develop differently would need to be tested with a larger dataset consisting of a greater range of ages.

The children who perform above chance level overall are good at distinguishing unfamiliar speakers from the familiar speaker, as shown by the low false alarm rate. The false alarms that do occur are mainly from the Yorkshire and Lancashire speakers who use a monophthongal pronunciation of the GOAT vowel, similar to that of the familiar speaker. An accent disguise confuses the children to the extent that many are unable to recognise the familiar speaker. Although this is similar to findings from Sjöström et al. (2009), the present study has a narrower

---

10 Further work is needed to investigate the FOOT vowel and its salience in the recognition and discrimination process. Wells (1982) remarks that unlike the GOAT vowel, the FOOT vowel does not vary markedly in its use by English speakers. However this observation needs to be updated with more current production data.
focus; one-word accent-based phonetic realisations as opposed to whole phrases. Therefore this limits the range of features in the voice that the children could be using to recognise the speaker.

Wider implications of the findings from this experiment relate to how speakers’ voices are stored cognitively and the kind of social information that may be stored alongside the linguistic information. This is discussed in the next section.

5. General discussion

Results from these two experiments show that, overall, age and exposure play a key role in children’s ability to recognise and identify familiar speakers. Although this corresponds to intuitions about the nature of speaker recognition, the findings can be explained if we interpret them through cognitive models of recognition. Recognition involves long-term memory retrieval and this can manifest itself in two quite different ways. On the one hand, specific information or events may be recalled so that an instance of a particular familiar item is brought to the forefront of the mind. Alternatively, there may just be a strong sense of familiarity about something but nothing more specific than this is retrieved from memory (Reisberg, 2010). ‘Dual process’ theories of recognition specifically claim that these two types of memory actually involve separate memory processes. Diana et al. (2006) explain these as ‘recollection’ and ‘familiarity’ processes. The familiarity process is primarily based on the activation of conceptual information which has been stored alongside previous encounters with the familiar item. The recollection process goes further and accesses stored episodic information as well as conceptual information. This results in the retrieval of specific instances of the remembered item.

This framework can be applied to the present study, to explain the processes involved in identifying a familiar speaker. In being able to name a particular speaker, the children are recalling other stored instances of the speaker from their long-term memory and therefore accessing both conceptual and episodic information. Children’s conceptual abilities are qualitatively not all that different from those of adults. For example pre-school children have been shown to rely on basic-level categories and use hierarchical categorisation just as adults do (see Murphy, 2002 for an overview of such studies). Therefore any improvement of conceptual ability with age is due to general developmental advances that come with increased experience and exposure to members of categories, a better understanding and knowledge of the world, and increased processing speed that develops with maturation of the brain (Murphy, 2002). This helps to explain the results of the Identification experiment in which the older children who have had more exposure to, and therefore experience with, the nursery teachers perform better than the younger children. In the Recognition experiment, the girls who had parents from outside of Yorkshire performed better than those with at least one parent from Yorkshire. Therefore it appears that their increased exposure to members of accent categories outside of Yorkshire increases their conceptual ability to differentiate between individuals producing regionally-based phonetic accent features.

The improvement in the Identification experiment throughout the pre-school years can also be explained by the dramatic change in episodic memory during this time (cf. Newcombe et al., 2000). The remembering of particular episodes uses explicit memory which, unlike implicit memory, is strongly age dependent. For example, Drumme and Newcombe (1995) found that 5-year-olds outperformed 3-year-olds in remembering having seen particular animal pictures three months previously. However there was no age-related improvement in the implicit
memory task which involved identifying the animal from a blurry picture becoming increasingly focused.

Exemplar-theoretic accounts which rely upon the storing and accessing of episodic memories are now fairly commonplace in the speech perception and sociolinguistic literature (cf. Pisoni, 1997; Foulkes and Docherty, 2006). These accounts suppose that linguistic encounters, in the form of words, are stored in memory alongside other non-linguistic information about the encounters. This includes social information about the speaker, for example their voice quality or the particular segmental features used (Foulkes and Docherty, 2006). Therefore when particular linguistic exemplars are accessed through episodic memories, this social information is accessed as well. Which particular words are stored in this way depends on the listener’s attention and the importance of the encounter to the listener at the time (cf. Foulkes and Hay in press). This account would explain the inaccuracies of the children in the Recognition experiment. When Alice’s voice is disguised, the segmental phonetic information does not match closely that of the children’s stored exemplars of Alice. Therefore there is an impairment in the children’s ability to use the phonetic information to access the social information regarding who the speaker is. Furthermore, an unfamiliar speaker may be mis-identified as Alice if phonetic information is similar enough to activate stored exemplars of Alice, which in turn activates the social information regarding the speaker’s identity. This mistaking of an unfamiliar speaker for someone familiar has been found previously in studies with adults (cf. Ladefoged & Ladefoged, 1980).

Whether exemplars of speakers are stored and accessed in this way is a matter of considerable contention. Exemplar theory acknowledges that a listener might not always rely on detailed representations of instances stored in memory but some abstraction of them. However, as Foulkes and Hay (in press) highlight, the exact nature of the relationship between exemplar-specific instances and abstractions has yet to be fully worked out. Another possibility is that listeners primarily depend on prototypes. While still reliant on experience with particular speakers, prototypes form a more abstract cognitive representation of speakers. Öhman (2013) describes Papçun et al.’s (1989) memory model of voices which suggests that an average ‘prototype’ of voices is used as a means of comparison with any unfamiliar voices which are heard. For familiar voices, deviations from the average prototype can be stored through experience, and it is these idiosyncrasies that are searched for and accessed in the recognition process. Therefore, this theory makes a prediction that the more a particular voice deviates from the prototype, the easier it is to recognise. This could explain the results from the Identification experiment in which the speaker whose voice deviates most from the others in terms of pitch and voice quality (Jane) is the most readily recognised.11

Whether the memory retrieval process used in recognition is based on exemplars or prototypes is a matter for conjecture as this is very hard to test empirically. Perhaps both processes are used, as Reisberg (2010) suggests. Both models rely on the accessing of a memory and a similarity judgement. Also, both models advocate the storing of features, whether this is social information associated with particular instances or a series of deviations more generally. Therefore it is possible that prototypes are used in some cases of recognition and exemplars in others. This is likely to depend on the experience of the individual and the strength of familiarity they have for a particular memory they are accessing.

11 Something to further investigate would be the effect of including another deep female voice in the speaker line-up. Another speaker who similarly deviates from an average prototype would potentially be confused for Jane. Perhaps in this situation the listener would rely on other, less salient cues in order to identify the speaker as Jane over another, similar sounding speaker
Whichever model we use to interpret them, the findings from the present study indicate that linguistic information pertaining to a familiar speaker’s regional background is stored in memory and somehow accessed during the recognition process. Further work is needed to investigate, as suggested by Foulkes (2010), how children advance from the storing of accent information alongside familiar speakers’ exemplars to the forming of speaker groups based on social differentiations such as regional accent.

6. Conclusion

This paper has investigated pre-school children’s ability to identify familiar speakers and in particular, which aspects of a speaker’s voice seem to play an important role in the recognition process. The Identification experiment revealed that the familiar teacher with the most distinctive voice pitch and voice quality was most readily recognised by the children. This indicates the importance of these voice attributes in identifying an individual who has a particularly idiosyncratic voice pitch/quality. The Recognition experiment found that regionally-based phonetic accent features also appear to contribute to speaker recognition. The children were more likely to misidentify an unfamiliar speaker as their familiar teacher if they used the same phonetic accent features. Additionally, the children generally failed to recognise their teacher when she disguised her accent and used a phonetic realisation different to her own.

This study also looked at the effect of external influencing factors on the abilities of the children in both tasks. Generally, there was an improvement in both tasks throughout the pre-school years as the oldest children outperformed the youngest children. However, this was not found to be a straightforward relationship and must be considered along with other factors. In the Identification experiment, the amount of exposure the children had to the nursery teachers interacted with their years spent at nursery. This meant that overall, older children who spent more time at the nursery on a weekly basis were better able to identify their teachers. In the Recognition experiment, the children’s sex seemed to have more of an impact on their ability.

For the boys, there was a significant improvement in the task throughout the pre-school years, showing an age-related development. The girls, who generally outperformed the boys, did not show an improvement with age but with an exposure to different regional accents at home. This suggests a more advanced conceptual ability to differentiate speakers with different accents amongst those who have increased exposure to members of accent categories outside of their home region. Further work to investigate these findings and their implications is needed, with a larger data set across a greater range of ages.

Acknowledgements

This research is funded by the Arts and Humanities Research Council. I would like to thank Paul Foulkes, Carmen Llamas and Tamar Keren-Portnoy for feedback on earlier versions of this paper. Thanks also to Márton Sóskuthy for his help with statistics in R. I am very grateful to all those involved in the experiments; both the adults who were recorded and the children who participated. Finally, thanks to two anonymous reviewers for their comments and advice.
References


Ella Jeffries
Department of Language and Linguistic Science
University of York
Heslington
York
email: erwj500@york.ac.uk
THE ROLE OF GENDER-CORRELATED SOCIOLINGUISTIC VARIABLES IN IDENTIFYING SPEAKER-INDEXICAL INFORMATION

ANIA KUBISZ
University of York

Abstract

The present paper investigates perceptions of speaker social-indexical information, including gender, age and social-class, from smaller phonetic segments such as gender-correlated phonetic variants. Since fundamental frequency (F0) is not the only cue to speaker gender identification, the perceptions are examined using gender-ambiguous sounding speech. The results of the study show that while speaker social-indexical information is identifiable at the segmental level, some of the phonetic variants seem to encode social-indexical information to a more consistent degree than others.

1. Introduction

Previous socioperceptual studies focus on identifying speaker-indexical information such as ethnicity (Purnell et al., 1999; Wolfram, 2000), geographic origin (Bezooijen & Gooskens, 1999; Clopper et al., 2005) or personality traits (Ball & Giles, 1988; Bezooijen, 1988; Lambert et al., 1960). Researchers have also investigated female and male voice identification (Biemans, 2000; Munson & Babel, 2007). Even though it has been established that listeners are quite accurate at identifying adult female and male voices, it is still unclear how listeners identify gender in the speech signal (Munson & Babel, 2007). Literature provides evidence that fundamental frequency impacts femininity and masculinity judgments (Foulkes et al., 2010; Munson & Babel, 2007). However, fundamental frequency is not always a decisive factor. First of all, there is an overlap of female and male pitch ranges, such that a lower-pitched female voice might be erroneously taken for a higher-pitched male voice and vice versa (Biemans, 2000; Foulkes et al., 2010). Furthermore, Johnson et al. (1999) showed in their study that a voice judged as most stereotypically female had lower mean fundamental frequency than the non-stereotypical female voice. Also Klatt & Klatt (1990) demonstrated that voices judged as typically female were not always characterised by high pitch.

Finally, it has been reported that listeners are able to distinguish male and female speakers in the absence of acoustic information normally found in speaker fundamental frequency (Assmann & Nearey, 2007; Coleman, 1971; Hubbard & Assmann, 2013; Lass et al., 1975). These findings imply that parameters of the vocal tract are not the only factors deciding whether a speaker sounds feminine or masculine, which further implies that gender-specific acoustic information does not rely heavily on fundamental frequency.

Because fundamental frequency is not the only cue to speakers’ gender identification, it is hypothesised that when speaker-social information embedded in fundamental frequency is not accessible to the listener, this type of information can be identified from other cues such as gender-correlated phonetic variants.
Therefore, this paper examines whether speaker social-indexical information can be identified at the segmental level.1

This study builds on earlier research on perception of speaker-indexical information in child speech (Foulkes et al., 2010). Following the findings of Foulkes et al. (2010), it is hypothesised that listeners familiar with the dialect and particular variant realisations should be sensitive to speaker-indexical information carried by these variants. However, listeners with no previous exposure to the dialect are not expected to be able to access this information. The present study also goes further in investigating perceptions of speaker-indexical information than Foulkes et al. (2010). In addition to speaker gender, it investigates perceptions of speaker age and social class. The three types of speaker indexical-information will be discussed in detail in the Results section.

A set of gender-correlated phonetic variants identified in Tyneside English were selected for the purpose of this study. Variants are sociolinguistically marked in terms of speaker gender, age and social class. It was decided to use Tyneside English phonetic variants in the study for the following reasons. Firstly, in Tyneside English vowels and stops can be realised using a number of variants, as will be sown in the following section. Because Newcastle is considered to be the hub of the North East region, its dialect has been extensively researched and described (Beal et al., 2012; Docherty & Foulkes, 1999; Foulkes et al., 2005; Milroy, Milroy & Hartley, 1994; Milroy, Milroy, Hartley & Walshaw, 1994; Watt, 2000, 2002; Watt & Allen, 2003; Watt & Milroy, 1999). Furthermore, Tyneside English is stereotypically perceived as the variety spoken in all of the North East (Beal et al., 2012; Pearce, 2009).

Perceptions of Tyneside-localised variants were compared and contrasted with perceptions of other localised variants used in the wider North-East region, or non-marked supra-local variants. And so, localised vowel variants were contrasted with supra-local variants, while consonantal variants, with one exception, were contrasted with non-marked variants.

2. **Tyneside English**

Great Britain is characterised by an abundance of local dialects. The North East of England, with Tyneside English being one of many spoken in the region, is no different. However, outsiders tend to have a distorted view of the North East. They seem to neglect a number of distinct dialects, such as Sunderland or Middlesbrough dialects, present in the region and consider Tyneside English to be spoken anywhere up north (Beal et al., 2012; Pearce, 2009). However, each of the dialects in the region is characterised by distinctive phonetic features.

Variation in the use of some vowels and consonants is one of the main phonetic cues revealing social and regional characteristics of the speakers within the North East (Beal et al., 2012: 26). However, there is also considerable variation within Tyneside English in terms of the use of phonetic variants. In fact, Tyneside English is characterised by an array of localised phonetic variants, which are marked sociolinguistically, as they are not only gender- but also age- and class-specific (Watt & Allen, 2003: 269). It is these features that distinguish Tyneside speakers from speakers south of the River Wear or Teesside (Beal et al., 2012). It is also these features that distinguish speakers within Tyneside English.

---

1 While further research will investigate and compare perceptions of speaker-social information provided by Tyneside listeners and listeners from the whole of the North East region, the present paper focuses on listeners from Tyneside, who are familiar with the dialect.
Therefore, a set of Tyneside English gender marked phonetic variants have been selected for the purpose of this study. The vowel variables comprise realisations of the FACE, GOAT and NURSE vowels. The remaining variables constitute realisations of voiceless plosives /p, t, k/ in word-final and medial positions.

The section below provides an account of variation and possible realisations of the variables investigated in the study.

Two perceptually prominent vowels in Tyneside English are the FACE and GOAT vowels (Beal et al., 2012; Watt, 2000). Not only can they be realised using a number of variants but also different variants are used by older and younger speakers (Watt, 2000).

Watt (2000, 2002) lists three types of realisations of the FACE and GOAT vowels and groups them into monophthongs, centring diphthongs and closing diphthongs. The most commonly occurring and unmarked variants of FACE and GOAT in Tyneside English are the monophthongal realisations, [eː] and [oː]. These realisations are also found in other varieties of North East English, and as such, are supra-local (Beal et al., 2012: 31).

Monophthongal [eː] and [oː] are found across male and female speakers of different ages and social backgrounds in Tyneside English. The only exceptions are older working-class male speakers who, instead, use the centring diphthong [ɪə] as a realisation of the FACE vowel. The GOAT vowel is realised as monophthongal [ʊə], the centring diphthong [ʊə] or the fronted monophthongal [ɛə]2 in this group of speakers (Beal et al., 2012; Watt, 2000, 2002; Watt & Milroy, 1999).

While the diphthongal FACE and GOAT variants [ɪə] and [ʊə] are found in all of the North East, they are, in fact, associated with Tyneside English and considered to be traditional and old-fashioned, and are most often used by older working-class males (Beal et al., 2012; Watt, 2000, 2002). [ɪə] can also be found in the speech of younger working-class males, although much less frequently than in older working-class males (Watt & Milroy, 1999). [ʊə] is less frequently used by other groups of male speakers than older working-class. For example, older middle-class or younger working-class speakers use it less frequently, and younger middle-class speakers use it very rarely (Watt & Milroy, 1999).

The closing diphthongs are [eɪ], which is a realisation of the FACE vowel, and [ʊʊ], which is a realisation of the GOAT vowel. Overall, [eɪ] is not a common variant in Tyneside English, yet it is becoming more popular among younger middle-class speakers. It is used most often by young female middle-class speakers, followed by young middle-class male speakers (Watt, 2000). Both diphthongs are also Standard English forms.

The closing diphthong [ʊʊ] is also widely used in other parts of the country. In Tyneside English, this realisation is used by young middle-class speakers (Watt, 2000; Beal et al., 2012). The fronted monophthongal [ɛə], on the other hand, is largely found in male speakers and is used most frequently by younger middle-class males but also older and younger working class males. However, the variant is becoming less common in general and female speakers refrain from using it (Watt, 2000; Watt & Milroy, 1999).

Finally, Watt & Allen (2003: 269) and Viereck (1968: 69,70) provide more examples of the realisation of the GOAT vowel which make the vowel contrast in Tyneside English even more varied. For example, [ɪə] can be found in words like [stɪən] stone, [hɪəm] home and [bɪən] bone, and [ɛɹ] in words like snow [snaɹ]. These pronunciations occur in older working-class male speakers and are considered to be old-fashioned even by Viereck (1968).

---

2 Although the fronted monophthongal [ɛə] is also used in other varieties of English, for example in Yorkshire English (Haddican et al., 2013), for the purpose of this study it was treated as a localised variant.
Another vowel associated with significant variability in the region is the NURSE vowel, which can be realised as the localised retracted [ɔː], fronted [ɔː] and centralised [ɜː] (Beal et al., 2012; Watt, 1998; Watt & Milroy, 1999).

While the first variant is now rare and used mostly by older working-class male speakers, the two other variants are more commonly used in Tyneside English than [ɔː]. The centralised [ɜː] is most common and also supra-local. Watt (1998) and Watt & Milroy (1999) point out that the fronted variant [ɔː] is marked for age and gender, as it is found in female speakers, and especially younger middle- and working-class females who use it more frequently than [ɜː].

In general, localised vowel variants seem to be used by older and usually male speakers. Younger speakers, especially females, tend to prefer supra-local variants, widely found across the region and the country (Beal et al., 2012).

Overall, a decrease in the use of localised, traditional forms can be observed in Tyneside English (Watt, 2000). In their place, new, non-regional forms are adopted. The process leads to a reduction of the number of vowel variants in use and implies dialect levelling, which results in formation of a more uniform repertoire of phonetic variation, one that is closer to other varieties of British English (Watt, 2000, 2002). At the same time, the supra-local forms new to the region seem to be less socially and geographically marked.

The situation is quite the opposite in the case of consonant productions. Some localised consonantal realisations seem to be becoming more widely used by speakers of Tyneside English (Beal et al., 2012: 45).

One of the characteristic features of the North East dialects is a wide variety of realisations of voiceless stops. For example, /t/ in Tyneside English can be realised as T-to-R variant, a glottalised stop, a pure glottal stop, a fully released /t/ or pre-aspirated /t/ (Beal et al., 2012; Watt & Milroy, 1999). Also /p/ and /k/ are glottalised, pre-aspirated or realised as glottal stops in Tyneside English (Beal et al., 2012; Watt & Milroy, 1999).

As has been already mentioned, word-final /t/ can be realised as [ɫ]. This feature is commonly referred to as T-to-R and it is found in some restricted contexts, in pre-vowel position in a limited number of common verbs and non-lexical words, for example put on or shut up (Beal et al., 2012; Milroy, Milroy, Hartley & Walshaw, 1994; Watt & Milroy, 1999: 30).

Given the fact, that T-to-R is found most frequently in older working class female speakers and, to a lesser extent, in younger working class females, the process is marked sociolinguistically (Watt & Milroy, 1999).

Another marked feature of Tyneside English is glottalisation or glottal reinforcement of voiceless plosives (Beal et al., 2012; Docherty et al., 1997; Docherty & Foulkes, 1999; Milroy, Milroy & Hartley, 1994; Milroy, Milroy, Hartley & Walshaw, 1994; Watt & Allen, 2003; Watt & Milroy, 1999). Overall, glottalisation is a feature preferred by male speakers and it is found in intervocalic position when a stop gap is followed or preceded by a laryngealised voice, in words like for example: city, copy, happy, bottle, wanted, lucky or local (Watt & Allen, 2003: 268). These variants are associated with a double articulation and as such they are usually transcribed as [ɬp], [ɬt] and [ɬk] (Beal et al., 2012). At the same time, glottalised variants, and particularly glottalised /p/ and /k/ are claimed to be recessive (Docherty et al., 1997: 306).
As far as glottalisation of /p, t, k/ in intervocalic position is concerned, of the three stops, /p/ is most often realised as a glottalised stop. Table 1 presents rates of use of glottalised /p/ by male and female speakers in Tyneside English.

<table>
<thead>
<tr>
<th>Rates of use of glottalised /p/ by gender in Tyneside English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
</tr>
<tr>
<td>Women</td>
</tr>
</tbody>
</table>

Table 1: Rate of use of glottalised /p/ by male and female speakers in Tyneside English (Beal et al., 2012; Docherty et al., 1997; Milroy, Milroy & Hartley, 1994).

Glottalised /t/ is also a significantly frequent realisation of /t/ in intervocalic contexts. This feature is characteristic of older working class male speakers. Table 2 presents rates of use of glottalised /t/ by male and female speakers in Tyneside English.

<table>
<thead>
<tr>
<th>Rates of use of glottalised /t/ by gender in Tyneside English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
</tr>
<tr>
<td>Women</td>
</tr>
</tbody>
</table>

Table 2: Rate of use of glottalised /t/ by male and female speakers in Tyneside English (Beal et al., 2012: 39, Watt & Milroy, 1999).

Table 3 presents rates of use of glottalised /k/ by male and female speakers in Tyneside English.

<table>
<thead>
<tr>
<th>Rates of use of glottalised /k/ by gender in Tyneside English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
</tr>
<tr>
<td>Women</td>
</tr>
</tbody>
</table>

Table 3: Rates of use of glottalised /k/ by male and female speakers in Tyneside English (Beal et al., 2012: 38).

Voiceless plosives in Tyneside English can be also realised as pure glottal stops []. In Tyneside, /t/ is realised as []/ more often than /p/ or /k/ (Beal et al., 2012; Docherty & Foulkes, 1999; Milroy, Milroy & Hartley, 1994; Milroy, Milroy, Hartley & Walshaw, 1994). While realisations of /t/ as glottal stops are quite commonly used in southern parts of the North East region, in Tyneside English this type of realisation is found word medially before syllabic L (Beal et al., 2012). Apart from this context, intervocalic word-medial glottal stops are not very common in Tyneside English, where only young speakers use it 11 per cent of the time (Beal et al., 2012: 46). According to Watt & Milroy (1999), Milroy, Milroy, Hartley & Walshaw (1994) and Milroy, Milroy & Hartley (1994), this realisation is most commonly occurring in the speech of young middle class females, who are also responsible for spreading the variant.
Glottalised /p/, on the other hand, is not very common in Tyneside English and amounts to only 1 per cent of use for male and female speakers. It is more widely used for example, in Sunderland and especially Middlesbrough, where it has been a growingly popular variant among young female speakers (Beal et al., 2012).

As far as realisations of /k/ as [k] are concerned, they have not been found in Tyneside English (Beal et al., 2012: 39).

Finally, realisations of word final pre-pausal voiceless stops as pre-aspirated variants, with /t/ being most frequently pre-aspirated of the three variables, have been reported in the North East (Jones & Llamas, 2003; Beal et al., 2012). These variants have been also found in Tyneside English, where they are most frequent in young female speech (Docherty & Foulkes, 1999; Foulkes et al., 2005; Watt & Allen, 2003).

3. Method

For the purpose of this study, speaker pitch was shifted to obtain the effect of a gender-ambiguous-sounding voice.

This study uses single-word stimuli. The advantage of using single words over connected speech is that listeners can focus with greater ease on the specific type of information present in the acoustic signal (Munson, 2007). At the same time, this approach allows the researcher to control for more parameters and therefore draw more reliable conclusions from the data when analysing which phonetic cues listeners rely on.

In constructing each stimulus, it was important to ensure that listeners responded only to the variant under investigation as opposed to any other segment of a word. This was achieved by realising the variant of interest as localised, with all other segments being standard pronunciation.

3.1 Stimuli

A total of four voices were used in this study. Two phoneticians recorded target stimuli using Tyneside English variants and two other speakers recorded fillers used in the study. Speakers were in their forties and mid-twenties.

Stimuli selected for this study account for specific phonological contexts. Vowels occur in three phonological contexts: word-finally in open syllables, preceding a nasal, and preceding a fricative in one instance. For example, the words in the GOAT group included: grow, stone, dough, go, home and toe. Consonantal stimuli were also one-syllable words, where the variables under investigation occurred in intervocalic word-medial and final position. All consonants were adjacent to vowels, and so for example in the T-to-R group get off, put on and shut up were used, whereas in the glottalised /t/ group, bottle, wanted and city were found.

Preliminary tests with Adobe Audition 3.0 (Adobe, 2007) revealed that regarding the range of possible pitch manipulation and the final outcome in terms of voice naturalness, male voices gave better results than female voices. In other words, when working with male voices, it was possible to apply a wider range of pitch manipulations before the voice started to sound
unnatural. The results were less optimistic for female voices which would lose their naturalness before they started to sound gender-ambiguous. Therefore, only male voices were used in this study.

The tokens were recorded in a recording studio to .wav sound files at a sampling rate of 44.1 kHz and 16 bit mono resolution. All tokens were manipulated in Adobe Audition 3.0 (Adobe, 2007) using the Pitch Shifter function to raise pitch and obtain the effect of gender-ambiguous-sounding voice. In addition to preserving the tempo of the samples, high precision and default appropriate settings were selected. Pitch Shifter allows changes in fundamental frequency by semitones and cents, where 1 semitone is equal to 100 cents. Each token was manipulated individually between 1.0 and 4.0 semitones. Average F0 of target stimuli was 135 Hz.

The algorithm implemented by the Pitch Shifter allows the speech tempo to be preserved and the formant values to be adjusted to changes in pitch (Adobe, 2007). Because this study investigates perception of gendered phonetic variables in the absence of gender-specific fundamental frequency, the aim was to manipulate only one of the phonetic cues, that is, fundamental frequency. Preserving tempo and adjusting formant values to changes in pitch sustained other acoustic features of the recordings. Furthermore, this approach allowed to control for pitch and draw more specific conclusions about the acoustic cues responsible for perceptions of speaker-indexical information.

Figure 1: The [ɔː] variant in stir. Before manipulation of F0.
All tokens were normalised for volume in Adobe Audition CS5.5 (Adobe, 2012) using the Match Volume function. A single token was pre-selected and the remaining tokens were matched in volume to the pre-selected token using the file total root mean square power (RMS) function and limiting settings to ensure the output files were not clipped or overly loud.

In one instance, token duration required elongation using the Stretch and Pitch function in CS5.5. Previous shifting of F0 resulted in a speeded and unnatural output of the token. While duration was stretched by 50 per cent using iZotope Radius algorithm, pitch and speech characteristics were preserved.

At the end of the process, the naturalness and gender-ambiguity of the stimuli and fillers were judged by a male and a female sociophonetician familiar with the dialects of North East England.

3.2 Procedure

The experiment was conducted in laboratory conditions and administered in SurveyGizmo (SurveyGizmo, 2014). At the beginning of the experiment there was a short training session during which participants familiarised themselves with the types of scales used in the experiment. Participants heard four words, each with a different scale, and attempted to evaluate the speakers. Data from the training session were excluded from analysis. After the training participants were given time to ask questions. A total of 531 single-word stimuli and fillers were presented via headphones at a comfortable hearing level, one at a time. Each stimulus was played once only. The entire session took about 90 minutes and there were three breaks during which participants were asked to complete sudoku.

During the experiment, a visual representation of a stimulus was displayed on screen. Sound was played after an image and scale were loaded. The onset delay for audio was about a second.
The role of pictures was to help listeners not familiar with Tyneside English understand the recordings. The images also served as an additional element in the study, which alleviated a possible feeling of boredom. In order to avoid visual priming, except for two instances referring to filler words, pictures excluded images of men or women. Care was also taken to ensure that pictures used in the study were not associated with males or females. Images included photographs, drawings, cartoons and computer icons. They presented objects, concepts or activities illustrating words played to the listeners. Nevertheless, it could be argued that for some listeners some of the pictures could be associated with men or women. For example, with the word bat, a picture of a bat or a picture of a baseball bat was shown. With blur, a picture of blurred birds in motion or a picture of a blurred view through a windshield was shown. It could be argued that participants may have stereotypically associated driving fast with men. The same might be true for a baseball bat. A bowl filled with dough however, might be stereotypically associated by some with women. However, associations of this type may depend on an individual and their experience; as a result they may vary from person to person.

As far as the words are concerned, these were not tested for gender bias. Listeners were instructed to listen to each stimulus and evaluate it using a Visual Analogue Scale (VAS) slider with a 0 to 100 point scale, incrementing by 1 point and logging participant choices on the x axis. Listeners were also asked to go with their first impressions and to not over think their choices.

![Image](image_url)

Figure 3: Visual Analogue Slider (VAS) scale for evaluating perceived speaker-femaleness.
The role of gender-correlated sociolinguistic variables

Figure 4: Visual Analogue Slider (VAS) scale for evaluating perceived speaker-maleness.

Figure 5: Visual Analogue Slider (VAS) scale for evaluating perceived speaker-age.
The scales of choice were Visual Analogue Scales (VAS), which are continuous, fine-grained scales. One of their major advantages over other types of scales is that they give the subjects more flexibility when providing subjective ratings. This, in turn, makes the analysis more precise (Llamas & Watt, 2014).

However, to be effective, VAS needs to be clearly defined and polar rather than bipolar (Torrance, Feeny & Furlong, 2001). Thus, the left and right anchor points of the scale are required to not introduce two contrastive concepts, for example Definitely male – Definitely female. Instead, the concepts should form end-points of a continuum, for example Definitely female – Definitely not female or Definitely male – Definitely not male. This was the reason for developing two separate scales to evaluate speaker-gender in the present study. Applying bipolar scales would make it impossible to conduct analysis and interpret data.

Wording in each of the scales was colour-coded for the benefit of the participant. Distinctive colours aimed to associate a particular colour with a particular scale. Colour-coding was consistent, and so the femaleness scale used red wording, the maleness scale -- navy blue, the age scale -- orange and the social-class scale -- green.

Stimuli were presented in a fixed order and the slider was reset to a midpoint position on the scale after each evaluation. Additionally, the slider did not allow for stimuli to be left unrated and so, in order to proceed, participants had to move it.

Each stimulus was evaluated four times along four dimensions: perceived speaker gender – maleness and femaleness and perceived speaker age and social class. These alternatives were
presented in a mixed order, in such a way, that every stimulus was rated along only one dimension per block and on all four of them in total.

Data were saved on an external server owned by the software provider.

3.3 Participants

Listeners who participated in the study were from Tyneside, and so they fulfilled the criterion of being familiar with the dialect under investigation. They were volunteers recruited from the undergraduate and graduate student bodies at the University of York and Newcastle University. The majority of participants recruited at the University of York were in their fresher year, thus had lived away from Tyneside for about 6 months. Four of the York students were more advanced in their coursework and had lived in York longer, between 1 to 4 years. However, all of the students maintained to have family and friends in Tyneside and to visit home often. As far as participants from Newcastle University are concerned, they claimed to have lived in Tyneside all their lives.

Listeners were also close in age. With the exception of three persons, whose age ranges were 25-34, the rest of the participants were ages 18-24. Even though participants were asked to provide age ranges rather than specify their age, from additional questions in the survey it was clear that the oldest participant was 28 years of age.

Twenty-four female and seven male listeners participated in the study. Although the aim was to obtain a balanced sample of male and female participants, this proved difficult in practice.

In terms of social background, 13 participants considered themselves to be of working-class background and 11 described themselves as middle-class. The rest declared to be of lower-middle class or did not have a clear idea.

Only five participants claimed to speak a foreign language, out of whom two spoke two non-native languages. Participants also came from a wide variety of fields, the most popular being: speech and language therapy, history, philosophy and psychology.

None of the listeners reported a hearing impairment and only two suffered from a mild cold.

Each participant was paid £12 upon completing the study.

3.4 Data analysis

As has been already mentioned, 31 listeners participated in the study. Each participant evaluated between two to three words in each of the vowel and consonantal groups when evaluating localised variants and, in some cases, as many as six non-localised words per group. This resulted from the study design. Furthermore, this meant that data were clustered and they could be correlated rather than independent (Galbraith, Daniel & Vissel, 2010; Zyzanski, Flocke & Dickinson, 2004). In order to analyse data, first, average values for each of the participants were determined. It was decided to use averages rather than medians because of the small number of measurements per participant. Applying this type of measure ensured that data were independent which, in turn, allowed statistical analysis. Because in a number of cases data distributions were skewed, a non-parametric Wilcoxon test was applied. This type of test compares differences between two medians. Additionally, the exact probability option of the test was selected.
4. Results

This section presents and discusses experimental results of the perceptual study. First, the FACE, GOAT and NURSE vowel variants are discussed in terms of perception of speaker gender, age and social class. Next, results for T-to-R, glottalised voiceless stops and pre-aspirated /t/ are presented.

The data analysis presented below is based on all 31 Tyneside listeners who participated in the study.

4.1 Analysis of listener perception of vowels

The following section focuses on evaluation of speaker-indexical-social information of localised and supra-local vowel variants occurring in Tyneside English. Table 4 presents patterns of use of the FACE vowel variants in Tyneside English.

<table>
<thead>
<tr>
<th>Variants of the FACE vowel</th>
<th>Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ɪː] – a localised variant</td>
<td>most often used by older working-class (WC) male speakers</td>
</tr>
<tr>
<td>[ɛː] – a supra-local variant</td>
<td>used across all groups of speakers</td>
</tr>
</tbody>
</table>

Table 4: Usage patterns of the FACE vowel variants in Tyneside English (Beal et al., 2012; Watt, 2000; Watt & Milroy, 1999).

A closer look at the results presented below provides some interesting observations. Figure 7 and Figure 8 show evaluations of perceived speaker femaleness and maleness, respectively. The localised variant is used most often by older working-class males. As can be seen, both variants were evaluated as overall male sounding. Also a Wilcoxon signed-rank test did not show a statistically significant difference between evaluations of the two variants in terms of gender (maleness and femaleness). It can be noticed that Figures 7 and 8 are mirror images of one another. This would imply listeners were consistent with their evaluations.

Furthermore, from the results presented in the plots it can be concluded that evaluations provided by listeners seem to reflect results of the production studies (Beal et al., 2012; Watt, 2000; Watt & Milroy, 1999).
Figure 7: FACE localised [Iː] (1) and supra-local [ɛː] (2) variants -- evaluation of speaker gender.

Figure 8: FACE localised [Iː] (1) and supra-local [ɛː] (2) variants -- evaluation of speaker gender.
In terms of the perceived age of speakers (Fig. 9), the results show that the localised and supra-local variants were judged equally in the middle range of the scale which might imply that most participants found the voices to be mature but young-sounding. Furthermore, the spread of evaluations, particularly in the case of the localised variant [l→l] would suggest that listeners varied considerably in their perceptions.

A Wilcoxon signed-rank test showed a statistically significant difference (p < 0.001) between evaluations of the localised variant [l→l] and pan-northern variant [ɛ:] in terms of speaker social-class (Fig. 10). As can be noticed, the localised [l→l] was found to sound more working-class which corroborates findings of the production studies. By contrast, the pan-northern variant [ɛ:] was found to be less working-class sounding when compared with the localised variant. Furthermore, when taking a closer look at the box-plot it can be noticed that the [ɛ:] variant is characterised by a considerably wide spread of evaluations including working-class judgements as well as non-working class judgements. What is interesting, especially in comparison with the FACE results obtained for gender and age, is that participants seemed to be quite sensitive to the social-class information of the speakers.
The role of gender-correlated sociolinguistic variables

Table 5 presents patterns of use of the GOAT vowel variants in Tyneside English.

<table>
<thead>
<tr>
<th>Variants of the GOAT vowel</th>
<th>Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>[øː] – a fronted monophthongal localised variant</td>
<td>most often by younger middle-class (MC) males but also older and younger WC males</td>
</tr>
<tr>
<td>[iː] – an archaic localised variant</td>
<td>found in the speech of older WC males</td>
</tr>
<tr>
<td>[ʊə] – a centring diphthongal localised variant</td>
<td>most often used by older WC males</td>
</tr>
<tr>
<td>[oː] – a supra-local monophthongal variant</td>
<td>used across all groups of speakers</td>
</tr>
</tbody>
</table>

Figure 11 presents evaluations of speaker gender of the three localised GOAT variants. It seems that listeners found the fronted monophthongal and archaic variants to be very similar sounding in terms of the perceived speaker femaleness – that is, as not female sounding. The centring diphthong received the widest spread of evaluations in the second quartile, which might suggest that listeners did not have any strong associations of the variant sounding typically male (Figure 11). Finally, the supra-local variant was found to be overall not female sounding. However, less so than the fronted monophthongal and archaic variants.

In terms of evaluation of perceived femaleness of the GOAT variants, a Wilcoxon signed-rank test showed a statistically significant difference ($p < 0.01$) between evaluations of the localised [ɔː] and [ʊə] and between the archaic variant [iːə] and localised [ʊə] in terms of the perceived femaleness of voices. However, no difference in perception has been found for the localised [ɔː] and the archaic [iːə].

Listeners found the localised and archaic male variants [ɔː] and [iːə] to be overall male sounding, which is consistent with the results presented for evaluations of perceived femaleness. However, as can be seen in Figure 11, the spread of evaluations shows that some listeners found the variants to be altogether female sounding. As far as evaluations of the [ʊə] variant are concerned, a spread of evaluations noticed especially in the second and third quartile (Figure 11) suggests that listeners had quite varied perceptions of this variant and some of the listeners did not have a strong association of the variant as male sounding.
In terms of statistically significant differences in perception of maleness of the four GOAT variants, a Wilcoxon signed-rank test did not show differences between any of the localised variants. However, such differences were reported for the localised [ʊə], archaic [iːə] and the supra-local variant [oː], where p < 0.05 in both cases.

The results for the perceived speaker-age seem to be reflecting the findings of the production studies only to a certain extent. While the supra-local variant [oː] and the localised [ʊə] male variant were found to be in general mature-sounding, which correlates with the age ranges of speakers using these variants, the evaluations of the localised male variant [ʊə] and the archaic male variant [iːə] were not as clear-cut. Even though the evaluations of [ʊə] show a significant dispersion of ratings in the first and second quartile, the median itself is located towards the end of the scale, which means that the variant was perceived as older sounding. This, in turn, corroborates with findings of the production studies. This finding is particularly interesting when compared with the results of gender evaluation of this variant. Of the four GOAT variants, [ʊə] was judged as somewhat female sounding by the highest number of respondents. This links with the fact that the same variant was perceived as old sounding. This is the first instance of a possible correlation between perceived speaker-age and gender identified in this study. It will be investigated in the following sections whether more examples of such correlations can be reported.

As far as evaluations of the archaic [iːə] are concerned, a spread of evaluations can be noticed especially in the second and third quartile (Figure 12). This might suggest that listeners did not necessarily have any strong associations of this variant with older working-class males simply because of the young age of the listeners and the fact that the variant was already

Figure 12: GOAT localised [ʊə] (1), [iːə] (2) and [ʊə] (3) and supra-local [oː] (4) variants – evaluation of speaker age.
recessive in the mid- to late-1990s. This spread and the fact that the median evaluation is at mid-point on the scale, might actually reflect listeners’ uncertainty as to how to evaluate it.

Statistical differences were recorded for the following localised variants: [ɔ:] and [ʊɨ] (p < 0.01) and [iː] and [ʊə] (p < 0.05) as well as for the localised [ʊə] and supra-local [oː] (p < 0.01).

A Wilcoxon signed-rank test showed a statistically significant difference (p < 0.01) between evaluations of almost all variants under investigation in terms of perceived speaker social-class. The only non-significant comparison was found between the localised archaic [iː] and localised [ʊə]. A closer look at Figure 13 reveals that the archaic [iː] and local [ʊə] variants were, in fact, evaluated as working-class sounding. While the localised [ɔː] was found to be much less working-class sounding than the other two localised variants it was also perceived to be more working-class sounding in comparison with the supra-local variant [oː]. Once again, it seems that the listeners were quite sensitive to social-class information, even in the case of the archaic variant.

The results so far show that listeners are sensitive to the speaker social-indexical information embedded in the phonetic variants under investigation to a varied degree.

Table 6 presents patterns of use of the NURSE vowel variants in Tyneside English.
The role of gender-correlated sociolinguistic variables

<table>
<thead>
<tr>
<th>Variants of the NURSE vowel</th>
<th>Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>[øː] – a fronted localised variant</td>
<td>most often used by young MC and WC females but also older WC females</td>
</tr>
<tr>
<td>[ɔː] – a retracted localised variant</td>
<td>most often used by older WC males</td>
</tr>
<tr>
<td>[ɜː] – a supra-local centralised variant</td>
<td>used across all speaker groups</td>
</tr>
</tbody>
</table>


The final vowel investigated in the study is the NURSE vowel (Figure 14). When looking at Figure 14 it can be noticed that the female variant [øː] was evaluated as more female sounding than the male variant [ɔː]. At the same time a Wilcoxon signed-rank test showed a statistically significant difference (p < 0.05) between evaluations of these two localised variants in terms of their perceived femaleness. A difference between the male and non-marked variant was also reported, where p < 0.01 when evaluating the femaleness and maleness of the speaker.
As far as evaluations of speaker age are concerned (Figure 15), only the medians for the male [ɔː] and female [øː] illustrate a difference in perception. While the female variant [øː] was evaluated as slightly less old sounding, the male variant [ɔː] was found to be slightly older sounding in comparison. The supra-local variant, on the other hand, was evaluated similarly to the female variant. As could be expected from the spread of evaluations, no statistically significant differences were reported.

The results in Figure 16 show that, as for the previous variables, listeners were again quite sensitive to social-class information carried by the three variants. The results corroborate with findings of the production studies. Furthermore, perceptions of these variants in terms of social class were statistically significant (p < 0.001).
The role of gender-correlated sociolinguistic variables

4.2 Analysis of listener perceptions of plosives

The following section focuses on evaluation of speaker gender, age and social class of localised and supra-local variants of voiceless plosives occurring in Tyneside English.

Table 7 presents patterns of use of the GOAT vowel variants in Tyneside English.

<table>
<thead>
<tr>
<th>Variants of word-final pre-vocalic /t/</th>
<th>Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ᵰ] – T-to-R</td>
<td>most often used by older WC female speakers</td>
</tr>
<tr>
<td>[ʔt] – glottalised /t/</td>
<td>most often found in older MC and WC male speech</td>
</tr>
</tbody>
</table>

Listener perceptions of speaker femaleness and maleness were mirror images of one another. This would imply that listeners have a clear and consistent perception of the two variants in terms of their perceived gender. It is also worth drawing attention to the fact that the results of the perceptual test correlate with the results of production studies (Docherty & Foulkes, 1999; Watt & Milroy, 1999; Milroy, Milroy, Hartley & Walshaw, 1994; Milroy, Milroy & Hartley, 1994). While glottalised /t/ was strongly perceived as a male variant (see Figure 17), T-to-R received a wider spread of evaluations. This might imply that listeners were in fact sensitive to the gender information carried by the variant under investigation.

A Wilcoxon signed-rank test showed a statistically significant difference for age and social class for the two variants (p < 0.001). Thus, the results indicate a clear difference in perception of the two variants. In other words, social-indexical information encoded in these variants seems to be quite salient to Tyneside listeners.

Figure 17: T-to-R [ɹ] (1) and glottalised /t/ [ɾ] (2) variants -- evaluation of speaker gender.
Although a statistically significant difference between perceived speaker age of T-to-R and glottalised /t/ has been found, it seems that evaluations of speaker age do not reflect findings of the production studies (Figure 18), which report both variants to be used mostly by older speakers (Watt & Milroy, 1999; Beal et al., 2012; Docherty et al., 1997; Milroy, Milroy & Hartley, 1994). Nevertheless, a relation between perceived speaker age and gender can be noticed in the case of glottalised /t/. This relation results most probably from the design of the present study. Upon raising F0, male voices may seem to sound younger than before manipulation. A similar age-gender correlation has been already reported for one of the GOAT variants, which was found to be somewhat female-sounding and, at the same time, older-sounding (see Figures 11 & 12).

In terms of evaluation of speaker social class, a clear difference in perception of the two variants can be noticed (Figure 19). The female variant was found to be definitely working-class sounding. The male variant, by contrast, was perceived as less definitely working-class sounding. These results corroborate findings of the production studies. Furthermore, there was a wider spread of evaluations of the male variant, which suggests that listeners varied more in their judgments.
Figure 19: T-to-R [ɹ] (1) and glottalised /t/ [ɹ] (2) variants -- evaluation of speaker class.

Table 8 presents patterns of use of /t/ in intervocalic context in Tyneside English.

<table>
<thead>
<tr>
<th>Variants of intervocalic /t/</th>
<th>Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ɹ] – glottalised /t/ - a localised variant</td>
<td>most often found in older MC and WC male speech</td>
</tr>
<tr>
<td>[/] – glottalled /t/ - a localised variant</td>
<td>most often used by younger MC female speakers</td>
</tr>
<tr>
<td>released /t/ – a non-local variant</td>
<td>used across all speaker groups</td>
</tr>
</tbody>
</table>

Table 8: Realisations of intervocalic /t/ in Tyneside English (Beal et al., 2012; Docherty & Foulkes, 1999; Milroy, Milroy & Hartley, 1994; Milroy, Milroy, Hartley & Walshaw, 1994).
The role of gender-correlated sociolinguistic variables

Figure 20: Glottalised /t/ [ɾ] (1), glottal stop [ʔ] (2) and released /t/ (3) variants -- evaluation of speaker gender.

Figure 20 presents evaluations of the perceived speaker gender of three realisations of /t/ found in the Tyneside dialect. As can be observed, the average and spread of ratings for [ɾ] stimuli differs in this and the previous group. This might be because [ɾ] stimuli in different phonological contexts were used in opposition to [ʔ] and [ʔ]. It can be noticed that listeners did not differ in their perceptions of speaker-gender of the three variants (Figure 20). Glottalised, glottal and released /t/ were quite consistently evaluated as male sounding. Furthermore, when taking a closer look at the plots, it can be seen that even though listeners found these variants to be definitely male sounding, /t/ was found to be the most male sounding and the most definitely not female sounding in evaluations of perceived maleness and femaleness respectively. This is interesting since the variant is supra-local and as such is used by males and females alike.

No statistically significant results were recorded.
In terms of perception of speaker age (Figure 21), a difference can be noticed between evaluations of glottalised and glottal /t/. This difference was also statistically significant ($p = 0.001$). As far as the released /t/ variant is concerned, it was found to be the oldest sounding of the three variants under investigation. A Wilcoxon signed-rank test showed a statistically significant difference between evaluations of $[ʔ]$ and released /t/ ($p = 0.005$). This would mean that listeners were sensitive to information about speaker age embedded in the phonetic variants to a certain degree. We can definitely say that perceptual differences were found for the variants under investigation. These differences reflect the pattern of results of production studies, which have established that glottalised /t/ is most often used by older speakers, while glottal /t/ is used by younger speakers. Interestingly enough however, from the results presented in the plots, it can be seen that overall, all the variants were found to be young sounding. This might result from the fact that all three variants were evaluated by listeners to be male sounding.
Even though a Wilcoxon signed-rank test showed statistically significant differences between evaluations of the two localised variants (p < 0.01) and each of the localised variants and the supra-local variant (p < 0.001) in terms of speaker social-class (Figure 22), the results presented in the plot do not reflect results of the production studies, especially when comparing evaluations of the two local variants with the evaluations of the non-local variant. While glottalised /t/ is used by both working- and middle-class speakers, the listeners evaluated it to be less definitely working-class sounding than [ʔ] which is used by middle-class speakers. The main-stream variant, released /t/, was found to be perhaps middle-class sounding with the median evaluation around the mid-point of the scale. This, as well as the fact that the variant received the widest spread of evaluations might suggest that listeners associated it with speakers of all social strata.

Table 9 presents patterns of use of /p/ in intervocalic context in Tyneside English.

<table>
<thead>
<tr>
<th>Variants of word-medial /p/</th>
<th>Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ʔp] – glottalised /p/ - a localised variant</td>
<td>most often used by male speakers</td>
</tr>
<tr>
<td>released /p/ – a non-localised variant</td>
<td>used across all speaker groups</td>
</tr>
</tbody>
</table>

Figure 23: Glottalised /p/ [ʔp] (1) and non-marked released /p/ [p] (2) variants -- evaluation of speaker gender.

As in the case of glottalised and released /t/, glottalised and released /p/ were both evaluated as overall male sounding (Figure 23). Listeners seemed to be quite certain when judging perceived femaleness of the speakers (Figure 23), which is reflected by narrow spread of evaluations. No statistically significant differences were reported for evaluations of perceived speaker gender.
The role of gender-correlated sociolinguistic variables

The evaluations of speaker perceived age (Figure 24) of the localised variant do not corroborate the results of the production studies. The fact that glottalised /p/ was found to be young sounding might mean that, again, voice characteristics influenced listeners’ judgements of the perceived speaker age. It has been noticed already that variants perceived as male sounding were also found to be young sounding (see for example Figures 20 & 21). This correlation however, cannot be noticed for the standard released /p/ variant, which was found to be male- and older than teenage-sounding. Since a wider spread of evaluations close to the mid-point on the scale characterises this variant, it could be argued that it was found to be used most often by middle-aged speakers. This, on the other hand, reflects findings of the production studies.

A Wilcoxon signed-rank test showed a statistically significant difference between evaluations of the two variants in terms of perceived speaker age (p < 0.001).

The results for evaluation of speaker social-class draw attention to a difference in perception of the two variants (p < 0.001). While the localised variant was found to be rather working-class sounding, the standard variant was perceived as much less working-class sounding, perhaps even middle-class sounding. The results show that social-class is quite a salient feature to listeners.
Figure 25: Glottalised /p/ [p] (1) and non-marked released /p/ [p] (2) variants -- evaluation of speaker class.

Table 10 presents patterns of use of /k/ in intervocalic context in Tyneside English.

<table>
<thead>
<tr>
<th>Variants of word-medial /k/</th>
<th>Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ʔk] – glottalised /k/, a localised variant</td>
<td>most often used by male speakers</td>
</tr>
<tr>
<td>released /k/ – a non-localised variant</td>
<td>used across all speaker groups</td>
</tr>
</tbody>
</table>

Table 10: Realisations of intervocalic /k/ in Tyneside English (Docherty et al., 1997: 306; Milroy, Milroy & Hartley, 1994; Milroy, Milroy, Hartley & Walshaw, 1994).
The role of gender-correlated sociolinguistic variables

Figure 26: Glottalised /k/ [ʔk] (1) and non-marked released /k/ [k] (2) variants -- evaluation of speaker gender.

Listeners perceived the glottalised and released /k/ as different in terms of gender of the speakers. The difference can be noticed especially when examining perceptions of femaleness (Figure 26), where $p < 0.001$. While [ʔk] was found to sound definitely male, released /k/ was found to sound overall male, however, listeners perceived it as a less definitely male feature.

However, with regards to perceived speaker maleness, the perceptual difference between the two variants was less significant ($p < 0.01$).
In terms of speaker-age, glottalised and released /k/ were found to be perceived as different sounding (p < 0.01) (Figure 27). Glottalised /k/ was evaluated as younger sounding than released /k/. Both variants are characterised by wide spreads of evaluations, especially in the first and fourth quartiles. The medians, however, show that the variants were found to be overall mature-sounding. Thus, the results corroborate with findings of production studies only to some extent.

Also, differences between evaluations of speaker social-class were statistically different (p < 0.001). While glottalised /k/ was found to be working-class sounding, released /k/, by comparison, was evaluated as much less working-class sounding (Figure 28). Once again, it seems that listeners were quite sensitive to social-class information embedded in the phonetic variants.
The role of gender-correlated sociolinguistic variables

Figure 28: Glottalised /k/ [ʔ] (1) and non-marked released /k/ [x] (2) variants -- evaluation of speaker class.

Table 11 presents patterns of use of word-final post-vocalic /t/ in Tyneside English.

<table>
<thead>
<tr>
<th>Variants of word-final post-vocalic /t/</th>
<th>Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ʰt] – pre-aspiration /t/, a localised variant</td>
<td>characteristic of younger female speakers</td>
</tr>
<tr>
<td>released /t/ – a non-localised variant</td>
<td>used across all speaker groups</td>
</tr>
</tbody>
</table>

Table 11: Realisations of word-final post-vocalic /t/ in Tyneside English (Beal et al., 2012; Docherty & Foulkes, 1999; Foulkes et al., 2005).
Figure 29: Variants of pre-aspirated /t/ [ʰːτ] (1) of longer duration and more breathy pre-aspiration, shorter duration and less breathy pre-aspiration [ʰτ] (2) and non-marked released /t/ [τ] (3) -- evaluation of speaker gender.

Two variants of pre-aspirated /t/ were contrasted with released /t/, which is a realisation found in Tyneside and other parts of the country. A closer examination of pre-aspirated and released realisations of /t/ reveals that there were no differences between perceptions of the two variants of pre-aspirated /t/ in terms of femaleness (Figure 29). However, listeners did perceive localised and supra-local variants as different (p < 0.01 & p < 0.01). The supra-local variant was evaluated as the least definitely not female sounding.

Interestingly, however, listeners did not seem to be sensitive to speaker-gender information encoded in the pre-aspirated variants which were reported by production studies to be typically female. This might imply that the linguistic change has progressed and that word-final pre-aspirated /t/ is used by men and women alike.

Evaluations of perceived maleness of the stimuli under investigation revealed that the three variants were found to be overall male sounding. However, it should be pointed out that the pre-aspirated variant of longer duration had the highest spread of evaluations, including evaluations of definite femaleness, in the fourth quartile.
As far as age ratings are concerned, the non-marked fully released /t/ was evaluated as mature adult-sounding (Figure 30). The localised variant of shorter duration of pre-aspiration is characterised by a spread of evaluations in the second quartile. Overall, it was found to sound similar to the non-local variant.

Nevertheless, the localised variant of longer duration of pre-aspiration is characterised by even larger spread of evaluations, especially in the third quartile. This would imply that for some listeners this variant sounded older than the other two. This could be correlated with the fact that this variant was found to be the most female sounding of the three variants under investigation. These findings can be tied to earlier findings of age-gender correlation resulting from the design of the study. However, if any age-gender correlation exists in this case, it does not seem to be very strong.

![Figure 30](image-url)
As can be noticed, variants of pre-aspirated and fully released /t/ were evaluated as rather middle-class sounding (Figure 31). None of the variants were perceived as definitely working-class sounding.

A Wilcoxon signed-rank test did not show a statistically significant difference between evaluations of the three variants in terms of social-class.

5. Discussion and Conclusions

The results presented in the previous sections indicate that listeners familiar with the dialect under investigation were sensitive to speaker-indexical information on the segmental level. This information was identified from gender-ambiguous sounding speech. Furthermore, certain localised variants seemed to encode social-indexical information to a more consistent degree than others. While some of the presented results constitute very clear-cut cases, for example T-to-R, others reflect findings of the production studies only to some extent, for example the FACE vowel or pre-aspirated /t/. One of the possible explanations is that some
other cues came into play and overshadowed indexical information encoded in phonetic variants themselves.

As far as evaluations of gender of the speaker are concerned, it should be reminded that even though the voices sounded gender-ambiguous, it is often the case that localised variants are attributed to male speakers. This could explain why listeners not always evaluated variants most frequently used by female speakers as female. Nevertheless, even if this was the case, often statistically significant differences between variants were reported. This would indicate that, in fact, listeners were sensitive to gender differences between the variants at least to some extent.

Evaluating age of the speaker seemed to pose a similar problem to the listeners. Overall, variants were evaluated as middle aged sounding. The only exceptions were the majority of the glottalised variants which were found to be young sounding. Another finding was the age-gender correlation which could be noticed, for example, for a variant of the GOAT vowel and T-to-R contrasted with glottalised /t/. Variants perceived as more female sounding were also evaluated as older sounding, whereas variants found to be definitely male sounding were also rated as younger sounding.

It seems that after removing the cue of gender specific fundamental frequency social class of the speaker became the most salient social-indexical feature. From the evaluations of speaker social-class presented in the Results section, is it clear that listeners seemed to be quite sensitive to this type of indexical information and correctly identified patterns conditioned by social class of the speaker. It may be worth mentioning at this point that a number of participants reported feeling uncomfortable having to evaluate social class of the speaker.

It should be pointed out that perhaps different results would be obtained from listeners older than the current group. Especially perception of speaker age could differ. Even though the scales used in the study had clearly defined end points, it is common knowledge that the general perception of age and what or who is thought to be young or old varies between age-groups.

From the results presented above it is clear that listeners were able to extract speaker-indexical information from phonetic segments alone. Even though the success rate varied between variants and depended on the social-indexical feature being evaluated, the present study provides evidence that speaker social-indexical information is identifiable at the segmental level.

References


Ania Kubisz
Department of Language and Linguistic Science
University of York
Heslington
York
email: ania.kubisz@york.ac.uk
Implicit Associations with Welsh in Two Educational Contexts

RACHELLE LEE
University of York

Abstract
This study explores whether adolescents in Welsh- and English-medium schools in Cardiff, Wales, differ in the strength of their positive and negative associations with the Welsh language. An Implicit Association Test (IAT), which uses latency times to determine associations with dichotomous target concepts, is employed. The results suggest that there are indeed differences in both the strength and direction of positivity toward the Welsh language between pupils at different types of schools. Pupils from the Welsh-medium school displayed significantly more positive associations with the Welsh language than by pupils from the English-medium school, even when controlling for first language and parental language effects. While self-rated proficiency in Welsh is positively correlated with implicit associations with the language at the Welsh-medium school, no such correlation was found among pupils at the English-medium school. Finally, pupils from the Welsh-medium school who opted to complete the IAT in Welsh had significantly more positive associations with Welsh than those who completed the task in English, which is evidence of an implicit association-behaviour link. The implicit association data suggests that there may be a substantial ideological divide between Welsh- and English-medium schools in Cardiff.

1. Introduction

One of four nations that comprise the United Kingdom, alongside Scotland, Northern Ireland, and England, Wales is a country of around 3 million people. Although the majority language of Wales is English, nearly one-fifth of its inhabitants—around 562,000 speakers—also speak Welsh, a Celtic language that is thought to predate English on the British Isles by nearly 1,000 years (Lewis, 2008). While Welsh-English bilingual speakers can be found throughout Wales, they have traditionally been concentrated in the ‘heartland’ areas of northwest and west Wales, as shown in Figure 1.
Welsh is often cited as one of the ‘success stories’ of the endangered language movement (Berdichevski, 2004). Language revitalisation efforts over the past fifty years have succeeded in slowing, and even beginning to reverse, language decline in Wales. Between 1991 and 2001, the number of Welsh speakers increased from 18% to 20.5%, the first such increase recorded since 1901 (Price, 2010). This has been hailed as a huge victory for the language and has been attributed in large part to increased Welsh instruction in English-medium schools and the expansion of Welsh-medium schools. Indeed, young people between the ages of 3 and 15 represent the largest proportion of Welsh speakers in the country, at around 30% (Office for National Statistics, 2011). Since the 2001 census, however, there has been mounting evidence that many pupils, even fluent Welsh speakers, infrequently use the Welsh language outside of school (Jenkins, 2001; Jones & Martin-Jones, 2004; Thomas & Roberts, 2011). Additionally, there is a marked decrease in Welsh language use as children get older (Baker, 2003; Jones & Martin-Jones, 2004; Edwards & Newcombe, 2005; Hodges, 2009). Indeed, despite a 31% increase in the number of Welsh-medium schools between 1989 and 2009 (Musk, 2010), the latest Census language statistics revealed an overall decrease in speakership from 20.5% in 2001 to 19% in 2011 (Office for National Statistics, 2011). A speaker for the Office of National Statistics posited that this was likely to be the case because pupils were not using Welsh outside of the classroom (BBC News, 11 December 2012).

Young people have long been identified as a critical demographic for language revitalisation (Fishman, 1991), and in the Welsh context researchers have begun to ponder the reasons for their diminished use of Welsh over time. The problem appears to lie not in the delivery of Welsh-medium instruction, which is flourishing, but in the use of Welsh as an everyday language. Several studies have shown that adolescents do not feel that the language is useful for employment (Davies & Trystan, 2011), and that it is not practical for everyday communication with those who also speak English, or necessarily linked to their Welsh identities (Coupland et al., 2005; Price, 2010). Others have expressed trepidation about speaking imperfect or code-mixed Welsh, as purist attitudes prevail in certain sectors of the population (Robert, 2009).
The question of why English seems to be the preferred medium of communication among Welsh adolescents, particularly in regions with low densities of Welsh speakers, is critical to language planning. As of this moment, there is no clear resolution to this question and as such there is no clear path for moving forward. Interest in this topic has arisen in the media, in the Welsh government, and, to a lesser extent, among researchers in sociolinguistics, bilingual education, and language shift. Much of the research on this topic has taken the form of self-reported, structured questionnaires. In 2010, BBC One produced a documentary entitled *The Welsh Knot*, which was dedicated to addressing the question of why children opted to speak English outside of school, and was based on a language attitudes survey (Williams, 2010). In 2012, BBC Radio Cymru commissioned a large-scale survey of all age groups on attitudes toward the Welsh language (BBC News Wales, 2012). In 2013, BBC Radio Cymru, the Welsh-language television station S4C, and the Welsh Assembly government jointly commissioned a research study on language attitudes and use that incorporated some qualitative data and a literature review on language alongside use survey data (Beaufort Research, 2013). While these studies point to some important patterns in motivations for language choice and indicate broadly positive attitudes toward Welsh, they tend to be methodologically uniform and, in general, lacking in-depth sociolinguistic and/or social psychological analysis.

Fortunately, a mounting body of research examining young people’s attitudes toward the Welsh language has emerged over the past decade, through questionnaire data (Coupland et al., 2005; Price, 2010), focus groups (Davies & Trystan, 2011; Musk, 2006; 2010), matched-guise tests (Robert, 2009), interviews (Hodges, 2009), and observation of code-switching behaviour (Musk, 2006). These studies and others have made a tremendous contribution to our understanding of attitudes toward the Welsh language and the motivations behind language choice. Nevertheless, a multitude of questions remain, including the role of the school in fostering particular attitudes, the role of the Welsh language in Welsh identity, and a realistic assessment of whether language planners are capable of improving attitudes.

What all of these studies do reveal is that language attitudes are highly complex, multidimensional, contextually situated, and dynamic. Such a complex issue requires a methodology that incorporates qualitative data alongside quantitative figures, implicit measures alongside explicit ones, and a cross-disciplinary approach. The present study focuses on the implicit dimension, using a methodology that has been not yet been applied to the context of the Welsh language in Wales. Thus, it fills a gap by contributing to the understanding of ethnolinguistic social contexts in Wales through exploring social cognitive processes.

1.1 Language attitudes and revitalisation

The link between attitude and language vitality has long been acknowledged by researchers of language shift (Fishman, 1964). Although linguistic attitudes and behaviours do not always correlate neatly (Ladegaard, 2000), attitude has nonetheless been identified as both a measure of current vitality, and a predictor of its vitality in the future (Baker, 1992). This is especially true for minority languages, as “attitudes play a key role in their successful transmission, revitalisation and survival” (Loureiro-Rodriguez, 2012:1) and “appear to be important in language preservation, restoration, decay or death” (Baker, 1992:9), which can be seen in other Celtic languages such as Irish, Gaelic, and Scottish Gaelic (Edwards, 2010). Some have suggested that fostering positive attitudes is particularly critical during the teenage years, when young people are given the freedom to choose whether to continue using the language or to let their language skills deteriorate (Bo-Yuen Ngai, 2007; Jones & Uribe-Jongbloed, 2013; Ó Riagáin et al., 2008). Indeed, in some areas of Wales, particularly the southern and...
eastern regions, there is a noted incongruence between the successful implementation of Welsh-medium education and the lack of use in the family and the community (Hodges, 2012). Consequently, even if today’s adolescents desire to pass the language on to their children later in life, they will lack the necessary language skills to do so. This may explain why, while the Welsh language appears to have stabilised somewhat over the past several decades in demographic terms, its health in terms of community and family use is unsteady.

1.2 Role of the educational institution in the formation of language attitudes

The relative stability of the Welsh language over the last three decades can be attributed in large part to the proliferation of Welsh language education in Wales. As of 2012, there were 517 Welsh-medium primary and secondary schools in Wales—one school for every 903 pupils—serving roughly 25% of the pupil population. In regions with particularly high concentrations of Welsh speakers, nearly all pupils attend Welsh-medium schools; in the northwest local authority of Gwynedd, for example, 98.5% of pupils attend Welsh-medium schools. In regions with high concentrations of English speakers, however, there may be few or no Welsh-medium alternatives to English-medium schools. For instance, in the southeastern local authority of Monmouthshire, only 2.5% of pupils attended Welsh-medium schools in 2012. In other regions, the proportion varies. In Cardiff, where the present study takes place, 13.3% of pupils attended Welsh-medium schools.

The term ‘Welsh language education’ has two primary manifestations in Wales. Around 80% of secondary school pupils attend English-medium schools, where Welsh is offered as a second language. About 20% attend Welsh-medium or bilingual schools, where the majority of the curriculum is taught through the medium of Welsh, and where fluency is the target outcome (Welsh Assembly Government, 2013). Both forms of Welsh language education have grown since the passing of the 1988 Welsh Language Act, which granted Welsh the status in the National Curriculum of a ‘core subject’ for Welsh-medium schools and a ‘foundation subject’ in English-medium schools. Wales has seen a dramatic expansion in Welsh-medium education at both the primary and secondary levels, and community demand continues to increase. In fact, as of January 2012, there were 56 Welsh-medium secondary schools, comprising about a quarter of secondary schools in the country. The increase in the availability and uptake of Welsh-medium education was clearly reflected in the 2011 Census, which showed that children aged 5–14 comprise the largest percentage of Welsh speakers in Wales (ONS, 2011).

The two forms of Welsh language education differ dramatically in their target outcomes. The goal of Welsh-medium education is to produce fluent Welsh-English bilinguals who will, ideally, pass the language on to their children and thus perpetuate its vitality. This is achieved by providing “as much provision through the medium of Welsh as is necessary for learners to achieve fluency in two languages” (WAG, 2010:9). The goal in English-medium schools, however, is for learners to “[benefit] from opportunities to develop language skills which enrich their experience of living in a bilingual country” (WAG, 2010:9). Because Welsh learners receive only a maximum of a few hours per week of Welsh language instruction

1 In four local authorities (Merthyr Tydfil, Blaenau Gwent, Monmouthshire and Newport), there were no available Welsh-medium secondary schools as of 2012 (Office for National Statistics, 2012).
2 In practice, this means pupils in Welsh-medium schools are assessed in Welsh, while pupils in English-medium schools are required to take courses in Welsh from ages 5-16, but are not required to complete GCSE exams in Welsh.
3 24% of Welsh speakers in Wales are in this age bracket (Office for National Statistics, 2011).
throughout their statutory education\textsuperscript{4}, they are unlikely to become fluent speakers, but rather gain basic language skills such as those one might acquire in a secondary-level foreign language course. In addition, English-medium schools must comply with certain requirements designed to raise the profile of the Welsh language. These include displaying bilingual signage around the school, making communications with parents available in both languages, and using a few token Welsh phrases in the classroom on a daily basis.

The discrepancy in language proficiency of pupils in these two types of schools is an important consideration with respect to attitudes toward Welsh. There is a well-documented link between language proficiency and attitude (Baker, 1992; Garrett, 2010). Pupils who feel their language skills are lacking may have negative attitudes about the language, and likewise those who harbour negative feelings about a language are less likely to succeed in their learning of that particular language. Conversely, positive attitudes can be both the reason for and the result of high levels of language achievement. Baker asserts that ‘there is considerable evidence that ability in a language and attitude to that language are linked . . . attitudes and achievement may be both the cause and effect of each other. In a cyclical, spiral relationship, one builds on the other—in an upward or downward relationship’ (1992:44). This assertion is in line with those propounded by Coupland et al. (2005), whose large-scale study of perceived ethnolinguistic vitality demonstrated that engagement with Welsh, feelings of Welshness, and perception of demographic vitality were all correlated with language competence. This has critical implications for assessing differences in attitude between Welsh- and English-medium schools, where the gap in proficiency levels might well be reflected in attitudes.

Proficiency, of course, is not the sole indicator of attitude toward the Welsh language. Family language, media, and other social experiences are substantial influences on attitudes toward language (Garrett, 2010). Likewise, the level of proficiency facilitated by the school is not its only influence on the attitudes and perceptions of its pupils. Schools explicitly and implicitly communicate ideological positions on language use through the written materials they display, the way Welsh is discussed in the classroom, the linguistic behaviours of teachers and staff, and official policies. Fishman asserts that “schools are often important in connection with enriching their pupils’ attitudinal and overt-implementational commitments to language by providing and stressing the historical, cultural and moral rationales for such commitments” (1991:372). Different types of schools encourage such commitments in different ways and to different degrees. Baker points out that “a school where covert anti-Welsh attitudes are conveyed is likely to have a different effect from schools where the survival and nurturance of Welsh language cultural forms is one prime raison d’être” (1992:44). While Baker is not referring explicitly to differences between Welsh- and English-medium schools here, the principle could be aptly applied to that context. Welsh-medium schools are staffed by Welsh speakers, are engaged with other Welsh-speaking communities, and are often highly motivated by prospects for language revitalisation. English-medium schools, on the other hand, have the flexibility to be far more diverse in their ideological positions, yielding varying results.

Previous work on language ideologies in English-medium schools in Wales reveals the range of orientations toward the Welsh language that results from this flexibility. The Welsh-Medium Education Strategy developed by the Welsh Assembly Government (2010) is purportedly based on the foundations of Iaith Pawb (‘Everyone’s Language’), a seminal document put forth by the WAG that calls for “a truly bilingual Wales, by which we mean a country where people

\textsuperscript{4} Since Welsh has been designated a ‘core subject’, it is up to the individual school to decide how many hours per week are dedicated to Welsh instruction, which in practise ranges from 30 minutes to three hours. Additionally, the school determines whether pupils are required to sit GCSE exams in Welsh.
can choose to live their lives through the medium of either or both Welsh or English” (WAG, 2003:1) and which endeavours to “create the right conditions in which the Welsh language can grow and flourish in all aspects of Welsh life” (2003:2). Though the Welsh-Medium Education Strategy is intended for all schools in Wales, this ideology is manifested differently in different settings. Coupland et al.’s comprehensive examination of language attitudes in secondary schools throughout Wales led them to conclude that “the policy of compulsory teaching of Welsh to age 16 in Welsh secondary schools can […] be said to be delivering some level of Welsh language competence to the great majority of students in Wales […] but not to be delivering the ideological principle of Iaith Pawb” (2005:16). This was evidenced by a relatively low commitment to interactional uses of Welsh in the English-medium schools studied, despite a high level of appreciation of its symbolic and ceremonial values.

In her study of language ideologies in two secondary schools in southwest Wales, however, Selleck (2013) found evidence that some ideological principles of Iaith Pawb were indeed present in one English-medium school setting. Selleck’s focus group data in an English-medium school indicated an openness to the idea of using both Welsh and English in daily life, albeit a code-mixed version of Welsh frequently referred to as ‘Wenglish’. Her data showed that the English-medium school was more closely aligned with the ideology of ‘choice’ laid out by Iaith Pawb than the Welsh-medium school, precisely because at the English-medium school, “a particular institutional ideology of bilingualism (flexible bilingualism) gives “choice” to students, offers discretion to students, and increases the potential pool of new learners and users of Welsh” (2013:18). In contrast, the Welsh-medium school in her study promoted an ideology of monolingualism that restricted the language choices of its pupils within the school setting.

The present study aims to consider these ideological discrepancies through the lens of implicit social cognition, defined as “the introspectively unidentified (or inaccurately identified) trace of past experience that mediates R’ where R refers to the category of responses that are assumed to be influenced by that construct (Greenwald & Banaji, 1995:5; cited in Nosek et al., 2007:266). While the study of language ideologies requires a more multidimensional analysis than is possible with a single measure, the investigation of implicit associations can shed light on ideological processes at work in different environments. Van Dijk (1995) conceptualises ideologies as the interactions between cognitive processes and social interactions, manifested in interactional behaviours. That is, attitudes do not exist in a cognitive vacuum, but rather function within social structures. In this view, attitudes are “influenced by powerful ideological positions” (Garrett, 2010:34), and can therefore point to the nature of the social contexts in which they operate. In this study, the examination of implicit attitudes is expected to prove a useful tool for uncovering broader ideological structures in Welsh schools.

2. Methodology

2.1. About the IAT

The Implicit Association Test has its roots in social psychology, and was first developed by Greenwald et al. (1998) to examine social stereotypes. Its purpose is to indirectly measure the strength of associations between target concepts and diametrically-opposed positive or negative attributes. This information is obtained via a computer-based categorisation task in which participants match dichotomous attributes (e.g., ‘friendly’ and ‘mean’) to a target concept representing ‘positive’ or ‘negative’, and textual or pictorial exemplars (e.g., a picture
of a man and a picture of a woman) to the target concepts being assessed (e.g., ‘male’ and ‘female’). In three of the trial blocks, only one target concept is shown in each corner of the screen, which serves to familiarise the participant with the task. In two pairs of trial blocks, however, attribute and exemplar stimuli are mixed, and all four target concepts appear at once, where one target concept is matched with ‘positive’ and the other is matched with ‘negative’ (e.g., ‘MALE OR POSITIVE’ on the left side, ‘FEMALE OR NEGATIVE’ on the right side). The target concepts are then reversed (i.e., ‘FEMALE OR POSITIVE’ on the left side and ‘MALE OR NEGATIVE’ on the right). The difference in reaction times between these two block pairs indicates the strength of the association, where a faster mean reaction time and fewer errors indicate an easier categorisation and thus a stronger association.

The primary benefit of the Implicit Association Test is its capacity to indirectly access information about attitude, associations, and stereotypes that may be difficult to elicit through more direct methods. For example, social desirability bias, which is error arising from the desire to avoid embarrassment or disapproval, may impede a participant from answering a questionnaire item honestly (Garrett, 2010). The IAT has also been tested extensively for ‘fakeability’ and has been found to be far more resistant to dishonest manipulation than self-reporting (Nosek et al., 2007). Additionally, the IAT may reveal associations of which the participant is not consciously aware (Nosek et al., 2007). Like other indirect measures that have been used in sociolinguistics, such as the matched-guise test (Lambert, 1960) and priming (Bourhis and Giles, 1976; Hay and Drager, 2010), the IAT is a valuable complement to more direct elicitation techniques. The present study is the first to use the IAT to examine attitudes toward the Welsh language, and indeed the first to compare implicit associations toward two different languages on a large scale.5

2.1.1. The IAT and attitude

There has been some question as to whether the IAT is indeed a measure of implicit attitude. To begin with, there is debate about how attitude should be defined both within and across the fields of sociolinguistics, social psychology, sociology and others. The concept of attitude originates from social psychology but has been applied across numerous fields, each of which has had to grapple with its meaning in particular contexts and in frameworks built upon certain theoretical assumptions. Social psychologists and social constructionists, in particular, disagree about whether attitude is a stable entity of the mind that can be discovered through behaviour, or whether it is purely a socially-constructed phenomenon that emerges through interaction (Soukup, 2012). Discursive psychologists attempt to bridge this gap by defining attitude as a cognitive component of a larger ideological framework wherein mental representations and the social actors who possess them interact (Van Dijk, 1995). In any case, it is generally agreed that attitudes a) are complex in both their multifaceted nature and manifestations (Garrett, 2010); b) incorporate cognitive, affective, and behavioural components (Baker, 1992); and c) are at least somewhat socially embedded. Given these caveats, Garrett et al.’s definition of an attitude as ‘an evaluative orientation to a social object’ (2003:3) provides some common ground for analysis.

A further question is whether the IAT measures ‘attitude’ in the same way that an explicit instrument, such as a questionnaire, might. Karpinski and Hilton’s (2001) study showed that data from their IAT and from an explicit task patterned independently of one another. Because of this fact, they assert that the IAT detects “the associations the person has been exposed to in

5 In his doctoral dissertation, Redinger (2010) carried out a pilot IAT study using two languages, Luxembourgish and French. His study, however, involved only five participants.
his or her environment, not that individual’s level of endorsement regarding the attitude object” (2001:786). Uhlmann et al. (2012) argue that while there may be discrepancies between explicitly endorsed views and automatic associations, they are not totally independent entities. Cultural and environmental factors have an effect on people’s orientations toward evaluative objects whether those orientations are explicitly endorsed or not, which may in turn affect behaviour. Uhlmann et al. maintain that “implicit attitudes reveal the power of cultures to reproduce themselves in individual minds and the futility of conscious protests to the contrary” (2012:250). Though this assumes a rather deterministic view of the relationship between culture, attitude, and behaviour, it does acknowledge the permeability of those aspects.

Although it is necessary to be cautious in making claims about what the IAT actually does measure, the data it provides are valuable in any case, if judicious choices are made concerning how to make use of them. For this study, participants’ environmental exposure is particularly relevant, since it is concerned with examining the relationship between attitudes and ideologies in different educational contexts. Therefore, regardless of whether the associations measured by the IAT reflect the environment or personal subjectivities, they provide information about ideological frameworks. For analytical purposes in this study, ‘implicit attitude’ will be defined by the strength of association between positive or negative attributes and the target concepts.

2.1.2. The IAT and behaviour

Evidence regarding whether implicit attitudes correlate with specific behaviours is fragmentary. The inconsistency in the relationship between behaviour and attitude has long been a problem in all types of attitude research (Garrett, 2010; Ladegaard, 2000). Incongruent attitude-behaviour relationships may be due to conflicts between intention and actual commitment, competition between multiple attitudes, and an inability to control one’s behaviour. Likewise, the IAT has performed inconsistently in correlations with behaviours, particularly for stigmatised topics (Swanson et al., 2001). However, there is evidence that the IAT does outperform explicit self-report in predicting behaviour, especially for socially sensitive subjects where social desirability bias is a factor. In a meta-analysis of 122 IAT reports, Greenwald et al. (2009) found that while both implicit and explicit measures were able to predict behaviour to some degree, there was far less variability in effect size for implicit measures. Furthermore, the capacity of IAT measures to validly predict behaviour outperformed that of explicit measures in studies involving black-white interracial behaviour. For this reason, the IAT is a useful instrument for studies involving attitudes toward controversial topics, including aspects of language and culture.

2.1.3. The IAT in sociolinguistic research

Despite its popularity in social psychology, the IAT has been adopted only recently by sociolinguists. The majority of these studies have examined attitudes toward specific linguistic groups. In her study of New Zealanders’ degree of accommodation to Australian English, Babel (2010) found that attitude toward Australia in general was a significant predictor of phonetic convergence, or accommodation. Redinger (2010) piloted a small-scale IAT as part of a comprehensive study on language attitudes in Luxembourg and found that his participants associated Luxembourgish with more positive attributes than French. Pantos (2012) conducted an IAT in which linguistics pupils were asked to align American-accented and Korean-accented English speech with ‘GOOD’ and ‘BAD’, and found a pro-American-accented English bias, which contradicted participants’ questionnaire responses. Recently, Campbell-
Kibler (2012) used the IAT to investigate sociolinguistic and sociophonetic variability. Her data revealed associations between the American English variable (ING) and Northern/Southern states, blue-collar/white-collar professions, and country singers/news anchors. Using audio files, she also tested associations between (ING) and region, [ay] monophthongisation, and [t]-release. The data showed that there was indeed an association between region and the socially enregistered [ay] monophthongisation, but not the less enregistered [t]-release. Finally, she correlated her IAT data with two more explicit measures, and found little correlation for individual speakers. Nonetheless, she regarded the IAT as a promising tool for the study of sociolinguistic cognition.

2.2. Context of the study

The Implicit Association Test discussed herein is one component of a larger doctoral dissertation study on the relationship between the ideologies espoused in Welsh- and English-medium schools, adolescents’ attitudes toward the Welsh language, and the impact of language revitalisation efforts in Cardiff’s educational institutions. Its role in this broader context is to gain insight into the implicit positive and negative associations pupils have with the Welsh language, with the English language as a contrastive backdrop. Critically, this study aims to obtain information about attitudes that may be obscured by more explicit measures. The IAT used for the present study was modelled after the one used in Redinger’s (2010) study of attitudes toward Luxembourgish and French, and the parameters laid out by Nosek et al. (2007) were used to guide both the design of the instrument and the analysis of the data.

2.3. Community context

The study took place in Cardiff, the capital of Wales and its largest city at 346,100 residents. 10.6% of Cardiff’s population is Welsh-speaking, compared to the national average of 19% (ONS, 2011). 11.6% of pupils attend one of Cardiff’s three Welsh-medium secondary schools (Welsh Assembly Government, 2013). Parental demand for Welsh-medium education has been particularly high in this region, owing in part to the city’s increased need for Welsh-speaking employees. Perhaps surprisingly, Cardiff has received relatively little attention in sociolinguistic research on Wales in the past decade. Recent language attitude studies have been generally been conducted in areas of Wales with either very high or very low concentrations of Welsh speakers (Coupland et al., 2005; Price, 2010; Williams, 2009). While Cardiff is not generally regarded as the cultural capital of Wales (Carter, 2010), it is a ripe domain for sociolinguistic study in several respects. First, it has a unique status as the largest urban centre in Wales, drawing a diverse population from all over the world. Approximately one-fifth of the population of Cardiff belongs to an ethnic minority group, compared to 4.4% in Wales as a whole. This ethnic and, in some cases, ethnonlinguistic diversity introduces a level of complexity in orientation toward Welsh and Welshness. Second, the number of Welsh speakers in Cardiff has increased dramatically over the past twenty years, mainly due to young Welsh speakers seeking employment in the public sector, while former rural strongholds of the Welsh language such as Carmarthenshire and Ceredigion are rapidly losing speakers. In fact, Cardiff was one of only two local authorities in Wales to experience an increase in Welsh speakership between 2001 and 2011 (ONS, 2011). This demographic shift has prompted the Welsh Assembly Government to rethink its language planning strategies and to assess the

---

6 Aberystwyth, a city in a dominantly Welsh-speaking region of west Wales, is widely regarded as the unofficial ‘cultural capital’ of Wales (Carter, 2010).
potential for language revitalisation in urban areas (WAG, 2010:37). Finally, Cardiff houses the central offices of some of the country’s most prominent Welsh-language institutions, including BBC Radio Cymru, S4C, the Welsh Assembly Government, and Cardiff University, making it a crux of influence for language policy and planning.

2.4. Research sites

The study was conducted at one Welsh-medium and one English-medium secondary school in Cardiff. The Welsh-medium school, which I will refer to as Ysgol Gyfun Gymraeg (‘Welsh Comprehensive School’), or YGG, is one of three designated Welsh-medium secondary schools in the city. It is a comprehensive secondary school, serving pupils from years 7 to 13, inclusive (ages 11 to 18). In 2013, the school’s enrolment was around 850 pupils, a number which has increased steadily since the school’s opening. Due to its large catchment area, YGG has a diverse pupil population from a wide range of ethnic, linguistic, and socioeconomic backgrounds. Approximately 20% of the pupils come from homes where the predominant language spoken is Welsh; other home languages include Urdu, Punjabi, Korean, Spanish, French, and German, although nearly all pupils’ primary home language is English. At YGG, all courses except English are taught through the medium of Welsh, and pupils sit all exams, including GCSEs, in Welsh. For this reason, pupils are expected to speak Welsh at all times throughout the school day and face sanctions if they do not. In practise, however, it is well acknowledged by staff that most pupils primarily speak English outside the classroom, including on school grounds.

The English-medium school, which I will call English-Medium Secondary, or EMS, also caters to pupils from Years 7 to 13, and is larger than YGG with approximately 1,500 pupils. EMS is situated in a relatively affluent area of the city, and is ethnically and linguistically diverse. 37% of the pupils come from minority ethnic backgrounds, and many speak languages other than English at home, including Czech, Polish, Cantonese, Bengali, Cebuano (also known as Visayan), Hindi, and Spanish, among others. Only 1% of pupils speak Welsh as a first language. At EMS, as in all English-medium secondary schools in Wales, pupils are required to take Welsh as a second language from ages 5 to 16. The aim of this language policy is not to produce fluent speakers, but to teach pupils basic communicative skills in Welsh as well as to educate them about Welsh culture. Therefore, although pupils are encouraged to speak as much Welsh as they can inside the Welsh classroom, they are allowed to speak English as needed, and are permitted to speak whatever language they like outside the classroom. In practise, as in YGG, the preferred language is almost always English.

2.5. Participants

Sixty pupils from each school completed the Implicit Association Test. Thirty from each school were in Year 9 (ages 13–14), and thirty were in Year 10 (ages 14–15). Participants were selected by a staff member at each school, each of whom was told to choose participants from a wide range of ethnic and linguistic backgrounds and to balance for gender. At the time of the IAT, personal information such as home language and ethnicity had not yet been collected. This information was gathered during the questionnaire phase of the larger study. In order to reduce the possibility of biasing IAT results with participants’ explicit consideration of language attitudes, the IAT was conducted prior to the collection of this information.

---

7 The third Welsh-medium school opened in 2012 to keep up with parental demand for Welsh-medium education in Cardiff.

8 Foreign language courses are taught primarily through the medium of the target language; however, certain concepts are explained in Welsh.

9 Data from one
pupil at each school were discarded due to user error that interfered with the recording of reaction times, leaving a total of 118 participants. Of the 59 pupils who participated in the IAT from YGG, 31% (n=18) spoke Welsh as a first language, 15% (n=9) learned English and Welsh at the same time, 47% (n=28) spoke English as a first language, 3% (n=2) had a first language other than English or Welsh, and for two, no data were available. 61% (n=36) of participants had at least one caregiver or parent who was proficient in Welsh. At EMS, 88% (n=52) spoke English as a first language, 10% (n=6) had a first language other than English or Welsh, and for one, no data were available. No participant at EMS had Welsh as a first language, but 7% (n=4) had one caregiver or parent who was proficient in Welsh.

<table>
<thead>
<tr>
<th>Language</th>
<th>EMS</th>
<th>N</th>
<th>YGG</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welsh</td>
<td>0%</td>
<td>0</td>
<td>31%</td>
<td>18</td>
</tr>
<tr>
<td>English</td>
<td>88%</td>
<td>52</td>
<td>47%</td>
<td>28</td>
</tr>
<tr>
<td>Learned Welsh and English</td>
<td>0%</td>
<td>0</td>
<td>15%</td>
<td>9</td>
</tr>
<tr>
<td>concurrently</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>10%</td>
<td>6</td>
<td>3%</td>
<td>2</td>
</tr>
<tr>
<td>No Data</td>
<td>1%</td>
<td>1</td>
<td>3%</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 1. First language by school

2.6. Instrument design

The IAT was designed and run using the Python-based experimental software PsychoPy (Peirce, 2007). The task was completed on iMac 21 Core II Duo computers at YGG and on Dell PCs at EMS. Due to teachers’ scheduling preferences, the IAT was administered over four sessions of thirty pupils each. Four pupils at EMS were missing from their session and therefore completed their test several days later. Pupils at YGG were given the choice to perform the IAT in English or in Welsh. Since it was determined that the level of Welsh language in the Welsh IAT was too advanced for the pupils at EMS, all pupils from that school performed the IAT in English.

Five attribute pairs were selected for the IAT. Four of the pairs were selected based on salience and relevance to Welsh- and English-medium secondary school pupils, based on pilot interview data. These included INTELLIGENT / UNINTELLIGENT, USEFUL / IMPRACTICAL, COOL / UNCOOL, and HARDWORKING / LAZY. The fifth set of attributes, WEAK / STRONG, was included to incorporate an element of ethnolinguistic vitality perception, which has appeared frequently in explicit Welsh-language attitude studies (Coupland et al., 2005; Giles & Johnson, 1987). The IAT was piloted with five Welsh-English bilingual participants (who did not participate in the main study) before the main study and the attribute stimuli were judged to be appropriate.
The order in which attribute stimuli appeared was randomised within PsychoPy. Additionally, the order in which the judgment attributes were presented was counterbalanced, such that half the participants saw ‘WELSH LANGUAGE OR POSITIVE’ in Blocks 3 and 4 and ‘WELSH LANGUAGE OR NEGATIVE’ in Blocks 6 and 7, while the other half saw ‘WELSH LANGUAGE OR NEGATIVE’ in Blocks 3 and 4 and ‘WELSH LANGUAGE OR POSITIVE’ in Blocks 6 and 7.

2.7. Procedure

The IAT consisted of seven trial blocks, including five practise blocks and two test blocks. Blocks 1, 2, 3, 5, and 6 consisted of 20 trials each, while Blocks 4 and 7 each had 40 trials. Participants were instructed to match the terms that appeared in the centre of the screen with the terms in the right- and left-hand corners of the screen by pressing the ‘D’ or ‘K’ keys. If the answer was incorrect, a red ‘X’ would appear and participants would have four seconds to correct their responses before the next stimulus item appeared.

In Block 1, the target concepts ‘WELSH LANGUAGE’ and ‘ENGLISH LANGUAGE’ appeared in the upper right- and left-hand corners of the screen, respectively, as shown in Figure 1. In the centre of the screen, these same phrases appeared as exemplars.

---

**Figure 1: IAT Block 1**

---

11 In Greenwald et al.’s (2003) original methodology, ‘practice’ blocks 3 and 6 were not calculated in the D score. This was revised, however, in Nosek et al. (2007). It is now standard practice to include these blocks in the calculation.

12 Traditionally, the terms appearing in the centre during this block would be images or words that are symbolically associated with the target concepts. However, because there was no means of representing the two languages with words or images without conflating them with their associated countries (i.e., Wales and England), using identical terms was determined to be the most appropriate alternative to the conventional design.
In Block 2, the words ‘POSITIVE’ and ‘NEGATIVE’ appeared on the upper left- and right-hand corners of the screen, while adjectives representing positive and negative attributes appeared one a time in the centre, as shown in Figure 2.

![Figure 2: IAT Block 2](image)

In Block 3, all four target concepts appeared at one time, so that ‘WELSH LANGUAGE OR POSITIVE’ appeared in the upper left-hand corner and ‘ENGLISH LANGUAGE OR NEGATIVE’ appeared in the upper right-hand corner, as shown in Figure 3. The exemplar terms ‘WELSH LANGUAGE’ and ‘ENGLISH LANGUAGE’ randomly alternated with positive and negative attributes in the centre of the screen. Block 4 was identical to Block 3, but with 40 trials rather than 20.
In Block 5, the side of the screen on which the target concepts ‘WELSH LANGUAGE’ and ‘ENGLISH LANGUAGE’ appeared was reversed, as shown in Figure 4. Participants practised the new configuration over the course of 20 trials.

In Block 6, the upper left corner now read ‘ENGLISH LANGUAGE OR POSITIVE’ and the upper right ‘WELSH LANGUAGE OR NEGATIVE’, as shown in Figure 5. Again, the exemplar terms ‘WELSH LANGUAGE’ and ‘ENGLISH LANGUAGE’ alternated randomly with positive and
negative attributes in the centre of the screen. Block 7 was identical to Block 6, only with 40 trials instead of 20. The entire test took about seven minutes to complete.

![IAT Blocks 6 and 7](image)

Figure 5: IAT Blocks 6 and 7

Figure 6 shows the full content and structure of the IAT, organised by trial block.

<table>
<thead>
<tr>
<th>BLOCK</th>
<th>TARGET CONCEPTS (L-R)</th>
<th>STIMULI (RANDOMISED)</th>
<th>NUMBER OF TRIALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WELSH LANGUAGE/ENGLISH LANGUAGE</td>
<td>WELSH LANGUAGE/ENGLISH LANGUAGE</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>POSITIVE/NEGATIVE</td>
<td>COOL/UNCOOL</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>INTELLIGENT/UNINTELLIGENT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>USEFUL/IMPRactical</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HARDWORKING/LAZY</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WEAK/STRONG</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>WELSH LANGUAGE OR ENGLISH LANGUAGE / OR POSITIVE NEGATIVE</td>
<td>WELSH LANGUAGE/ENGLISH LANGUAGE</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COOL/UNCOOL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>INTELLIGENT/UNINTELLIGENT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>USEFUL/IMPRactical</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HARDWORKING/LAZY</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WEAK/STRONG</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>WELSH LANGUAGE OR ENGLISH LANGUAGE / OR POSITIVE NEGATIVE</td>
<td>WELSH LANGUAGE/ENGLISH LANGUAGE</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COOL/UNCOOL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>INTELLIGENT/UNINTELLIGENT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>USEFUL/IMPRactical</td>
<td></td>
</tr>
</tbody>
</table>
2.8. *D* Score calculation

The strength of associations was quantified using *D* scores, which are calculated using an algorithm (for a full discussion, see Nosek et. al, 2007). Data from Blocks 3, 4, 6, and 7 were included in the calculation. Trial latencies of less than 300 milliseconds were eliminated. For trials where the first answer was incorrect, the time of the corrected keypress was used. For trials where no answer was provided within 4000ms, the average block mean was used with an added 600ms penalty. The combined means from Blocks 3 and 6 and from Blocks 6 and 7 were divided by their respective standard deviations, and the two resulting quotients were averaged. The end result is a *D* score between -2 and +2. A positive *D* measure indicates a positive implicit attitude toward a target concept, while a negative *D* score indicates a negative implicit attitude. Since *D* scores are designed to show relative association strengths the units on the association strength scale are not meaningful in themselves. However, as a frame of reference for this study I will refer to the break points of .15 (slight), .35 (moderate), and .65 (strong) delineated by Harvard’s Project Implicit (Nosek et al., 2006; “Race Attitude”, 2013).

3. Hypotheses

The hypotheses that the experiment described in this paper is testing are as follows:

---

13 Nosek et. al (2007) provide two nearly identical algorithms: one for IATs where respondents can correct their responses and one where it is not possible to make corrections. In the present IAT, respondents were instructed to correct their responses, but because of technical issues, the stimulus that followed would appear after 4000ms of nonresponse. To compensate for this, procedures from both algorithms were used for calculating means for trials with errors.
(a) Participants who attend the Welsh-medium school will demonstrate more positive implicit associations toward the Welsh language than participants who attend the English-medium school, as measured by IAT $D$ scores.

(b) Participants at the Welsh-medium school whose first language is a language other than Welsh will have more positive attitudes toward the Welsh language than will participants at the English-medium school whose first language is a language other than Welsh.

(c) Participants at the Welsh-medium school who have no Welsh-speaking parents will demonstrate more positive implicit associations toward the Welsh language than will participants at the English-medium school who have no Welsh-speaking parents.

(d) Self-rated proficiency will be positively correlated with implicit attitude toward Welsh, as measured by IAT $D$ score.

(e) Those at the Welsh-medium school who choose to complete the questionnaire in Welsh will show more positive associations with the Welsh language than those who choose to complete the questionnaire in English.

4. Results

An independent-samples t-test showed that on average, participants at YGG had stronger positive associations with the Welsh language ($D = 0.302$, SE = 0.05) than did participants at EMS ($D = -0.195$, SE = 0.07). This difference was highly significant ($t(105.891) = 5.872$, $p < 0.0001$, $r = 0.50$). The mean $D$ scores in Table 3 show that YGG pupils demonstrated a weakly positive association while EMS pupils showed a weakly negative association with the Welsh language.

<table>
<thead>
<tr>
<th>School</th>
<th>Mean $D$ score</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>YGG</td>
<td>.302</td>
<td>59</td>
</tr>
<tr>
<td>EMS</td>
<td>-.195</td>
<td>59</td>
</tr>
</tbody>
</table>

Table 3. $D$ scores, all participants

Excluding participants whose first language was Welsh revealed a similar pattern. Participants at the YGG demonstrated more positive associations with the Welsh language ($D = 0.125$, SE = 0.08) than participants at the EMS ($D = -0.195$, SE = 0.07). The difference was highly significant ($t(82) = 2.836$, $p < 0.01$, $r = 0.30$). The strength of positive associations at YGG was attenuated by this exclusion.

<table>
<thead>
<tr>
<th>School</th>
<th>Mean $D$ score</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>YGG</td>
<td>.125</td>
<td>26</td>
</tr>
<tr>
<td>EMS</td>
<td>-.204</td>
<td>58</td>
</tr>
</tbody>
</table>

Table 4. $D$ scores, participants with first languages other than Welsh
The same discrepancy emerged when participants with one or more parents proficient in Welsh were excluded from the analysis. Participants at the Welsh-medium school demonstrated more strongly positive associations with the Welsh language ($D = 0.144$, SE = 0.09) than those at the English-medium school ($D = -0.214$, SE = 0.07). The difference was highly significant ($t(76) = 2.904$, $p < 0.005$, $r = 0.32$). Once again, mean $D$ scores for both schools went in opposite directions, and again the exclusion of children with Welsh-speaking parents in the home attenuated the strength of the positive association.

<table>
<thead>
<tr>
<th>School</th>
<th>Mean $D$ score</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>YGG</td>
<td>0.144</td>
<td>23</td>
</tr>
<tr>
<td>EMS</td>
<td>-0.214</td>
<td>55</td>
</tr>
</tbody>
</table>

Table 5. $D$ scores, participants with no Welsh-speaking parents

A Pearson correlation showed that among participants in the Welsh-medium school, self-rated proficiency was positively correlated with $D$ score ($r = 0.364$, $p$ (one-tailed) < 0.01), which supported hypothesis (d). For the English-medium school, however, no such correlation was found ($r = 0.007$, $p$ (one-tailed) = 0.48). This result did not support hypothesis (d).

The choice of IAT language at YGG was a significant predictor of $D$ score; those who completed the IAT in Welsh had stronger positive associations with the Welsh language ($D = 0.452$, SE = 0.07) than those who completed it in English ($D = 0.206$, SE = 0.08). The difference was significant ($t(57) = 2.523$, $p < 0.05$, $r = 0.32$). Moreover, those who chose to complete the task in Welsh showed a moderately positive association with the Welsh language, while those who chose to complete the task in English showed a weakly positive association with Welsh.

<table>
<thead>
<tr>
<th>Language</th>
<th>Mean $D$ score</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welsh</td>
<td>0.452</td>
<td>23</td>
</tr>
<tr>
<td>English</td>
<td>0.206</td>
<td>36</td>
</tr>
</tbody>
</table>

Table 6. $D$ scores by IAT language choice

5. Discussion

The difference between implicit associations at the two schools was substantial. The large discrepancy in $D$ scores in Table 3 might be attributed to participants’ home language, since there were no participants who spoke Welsh as a home language in the English-medium school.\(^\text{14}\) While bilinguals do not necessarily regard both of their languages with equal esteem (Baker, 2011), Welsh-English bilinguals are very likely to have come from home environments where use of the Welsh language is overtly and/or covertly encouraged. In Wales, it is quite common to have families with one Welsh-speaking and one English-speaking parent. Often in

\(^\text{14}\) Notably, only one pupil in all of the pupil population of EMS in Years 9 and 10 spoke Welsh as a first language. This pupil did not take part in the study.
these cases, the default language of the home is English, since that is regarded as the “language of inclusion” (Thomas & Gathercole, 2011:92). Thus, children who grow up speaking Welsh in the home are likely to have grown up either with two Welsh-speaking parents or caregivers, or with one Welsh-speaking parent or caregiver who is highly committed to passing the language on. In this environment, it could reasonably be assumed that children grow up with more positive associations toward the Welsh language than do those whose come from homes where Welsh is not spoken.

It might also be assumed that the difference in $D$ scores is due to other influences from the home environment, as YGG had more participants with one or more Welsh-speaking parents than EMS. It might be expected that having a Welsh-speaking parent would foster more positive associations with the Welsh language, even if Welsh was not acquired as a first language in the home. Some parents are first-language Welsh speakers themselves; others have learned Welsh as a second language to improve their employability, to communicate with their children, for personal enrichment, or to reconnect with their own Welsh identities. The linguistic identities and behaviours of parents could serve to promote positive associations with the Welsh language among their children. On the other hand, parents who choose to speak Welsh only outside the home may have little influence on their children’s attitudes, and some of those who speak Welsh as a first language but choose not to speak it to their children may do so because they hold negative attitudes of their own.

Despite the differences in their linguistic home environments, the pupils from YGG and EMS have multiple broader social contexts in common. Both groups reside in Cardiff, a metropolitan, multicultural, and multilingual city. Both come from schools with culturally and socioeconomically diverse pupil populations. English is the dominant language in which pupils at both schools communicate with their peers and interact with the media. And importantly, both groups fall within an age demographic that is characterised by self-exploration and identity formation, which is often manifested in rebellion against parental attitudes and expectations (Erikson, 1968). For example, in their large-scale questionnaire study of 12,000 10–15-year-olds in Wales, Sharp et al. (1973) found the oldest group of pupils they surveyed (14–15 years of age) to have the least positive attitudes toward the Welsh language, regardless of the school attended. During adolescence, linguistic values may be tested and called into question, and parental authority might lose its potency relative to that of peer influence. Given these commonalities, it is clear that the primary dividing factors of the participants in this study are family language behaviour and the type of school attended. If the former is controlled for, it can be determined whether the latter is a plausible influence on language attitudes.

Due to the differences in the linguistic backgrounds of the participants in the two schools, it is not possible to entirely isolate the school as a factor in comparing language attitudes. However, if first language and having a Welsh-speaking parent are eliminated as variables, the school emerges as a likely influence. Table 3 shows the comparative $D$ scores of pupils at YGG and EMS, eliminating pupils who speak Welsh as a first language. As predicted, participants at YGG had more strongly positive implicit associations with Welsh than those at EMS. Moreover, $D$ scores at YGG were positive, while those at EMS were negative. Given that no

---

15 On the questionnaire, information was gathered on whether pupils had a Welsh-speaking parent. However, no information was elicited on whether these parents communicated with their children through the medium of Welsh.

16 It should be noted that pupils from bilingual schools did have more positive attitudes than their counterparts who attended English-medium schools in all age groups. However, the apparent decline was evident among Sharp et al.’s pupils in both school types.

17 Individual social and linguistic experiences are also influential (Garrett, 2010); however, these cannot be accounted for on a group level.
participant in this analysis was a first-language Welsh speaker, it seems likely that the school environment was influential in fostering these associations.

Table 5 shows the comparison in $D$ scores between the two schools when participants with one or more Welsh-speaking parents were factored out of the analysis (logically, this also eliminated all first-language Welsh speakers). Once again, there is a significant difference in $D$ scores between the two schools, where YGG shows a positive implicit association with the Welsh language, and EMS a negative association. Again, this is notable, as all of the pupils in this sample came from homes where neither parent spoke Welsh. It is reasonable to assume from the data that the school plays a role in nurturing these associations.

Surprisingly, the two schools differ with respect to the correlation between self-rated proficiency and implicit attitude. While there is a significant moderate correlation between self-rated proficiency and $D$ score among participants at YGG, no such pattern appears among participants at EMS. The correlation at YGG is in accordance with previous research on proficiency and language attitudes (Baker, 1992). This connection, however, has been found in both monolingual and bilingual schools, and thus a similar pattern would have been expected at EMS. The discrepancy might be attributed to the fact that in the Welsh-medium school, the Welsh language is a central focus of the school curriculum, and emphasis on performing well in Welsh is far greater than it is in the English-medium school. This increased pressure may have a strong attitudinal effect on learners at YGG. It could also be the case that YGG and EMS pupils based their proficiency evaluations on different criteria. The emphasis on Welsh language competence at YGG may have led YGG pupils to overestimate their proficiency or align it with their attitudes on the questionnaire, in which case a correlation with implicit attitude would be expected. Finally, as has been suggested by Karpinski and Hilton (2001), it is possible that the IAT is indeed measuring factors other than attitude—such as environmental and cultural influences—that are not necessarily correlated with proficiency for all pupils. Further research is needed to determine whether such a correlation exists in other populations of Welsh language learners.

Remarkably, more than a third of participants at YGG chose to complete the IAT in Welsh. This was surprising for a number of reasons. First, in interviews with staff, it was frequently attested that the pupils at YGG spoke English the majority of the time in recreational spaces at school, even though they are strongly discouraged from doing so. Several teachers at the school attributed this pattern to ease of communication, especially for non-first-language Welsh speakers, to peer pressure, and in some cases to rebellion against school norms. Although the IAT was conducted in an IT lab at the school site during school hours, it was clearly stated that participants were welcome to complete the test in either language and that their data would not be shared with anyone. Additionally, these instructions were delivered in English by myself, an American English speaker. Second, about one-quarter of those who chose to complete the IAT in Welsh were first-language English speakers (and non-first-language Welsh speakers), who would presumably have been more comfortable completing the task in English. It is quite possible that the academic setting and the presence of a school administrator were major influences on language choice. Since this was a confidential individual task, however, it is also possible that given the freedom to choose a language in the absence of peer pressure, pupils freely chose to use Welsh.

Table 6 shows that those at YGG who chose to complete the IAT in Welsh had significantly more positive $D$ scores than those who chose to complete it in English. The data establish a tenable link between implicit associations and behaviour. This is similar to Redinger’s (2010) finding that attitudes toward Luxembourgish were most positive on the affective, instrumental, and integrative dimensions among participants who elected to complete a self-report
questionnaire in Luxembourgish. The majority of participants who completed the IAT in Welsh were first-language Welsh speakers. However, it is notable that there were several non-first-language Welsh speakers who chose to complete the IAT in Welsh as well. While the data here cannot establish a causal relationship between language choice and implicit attitude, they do suggest that those with more positive implicit associations with Welsh are more likely to use the Welsh language, at least in certain contexts. Further investigation is needed to determine to what extent this pattern extends to other types of language behaviour.

The IAT on its own is limited in what it can indicate about ideologies, and indeed about attitudes. It is a rather blunt instrument that is designed to take a snapshot of implicit associations in a specific context at a specific point in time and space. Given a different physical setting (e.g., in a recreational space outside of the school grounds or in the home), the data might look different. Language attitudes are subject to contextual influences, both in their realisations and their impact, and “could have varying shapes and effects across […] contexts” (Garrett, 2010: 225). In addition, the IAT does not capture the multidimensionality of attitudes. Unlike the matched-guise test or a comprehensive questionnaire (both of which were used in the larger study), the IAT cannot detect which specific attributes are most strongly associated with particular languages. As mentioned previously, there is also uncertainty about whether the IAT measures personal attitudes or reflects exposure to cultural environments, if indeed the two can be separated (Uhlmann et al., 2012). Finally, it relies on dichotomies—positive versus negative, Welsh versus English—that must be prescribed by the researcher in the design of the instrument. Thus, fine-grained nuances in implicit attitudes are undetectable.

That said, the IAT is valuable for language attitude studies for a number of reasons. Its most valuable asset is its capacity to reveal psychosocial associations without the need for introspection (Nosek et al., 2007) and without running the risk of skewing data with social desirability bias (Campbell-Kibler, 2012). Balancing implicit and explicit measures is critical for examining such a complex social psychological construct as attitude. Apart from the small number of studies mentioned previously, sociolinguists who have used these types of measures in tandem have been limited to matched-guise tests and priming. The IAT provides an additional tool in the sociolinguist’s repertoire, and one that is relatively easy to implement. Since the test only takes five to seven minutes to complete, it is user-friendly and minimises experimental fatigue. Furthermore, the IAT is particularly suitable for examining ideologies in different social environments, since it assumes that social psychological constructs are ideologically formed. In this way, the IAT functions as an index to broader ideological frameworks and a jumping-off point for further attitudinal study.

6. Conclusion

This article has examined Welsh- and English-medium secondary school pupils’ implicit associations with the Welsh and English languages. The present study employs the Implicit Association Test, a social psychological instrument designed to assess indirect attitudes, for the first time in the Welsh context. By exploring how associations with the Welsh language diverge within different educational contexts, this study contributes to our understanding of the role of the school in fostering positive or negative attitudes.

The IAT data show that implicit associations with the Welsh language among pupils in the Welsh- and English-medium schools differ quite dramatically, even when accounting for the effects of first language, home language, and parental proficiency. The data indicate that non-first-language Welsh speakers at a Welsh-medium school have significantly more positive
Implicit associations with Welsh in two educational contexts

associations with the Welsh language than their English-medium-school counterparts. While other influences may be at play, there are strong indications that the school environments factor heavily into implicit attitudes. As pupils in Wales spend roughly 40 hours per week in school, educational institutions make up a substantial portion of their social and cultural environments. Through official policies, language display, and discourse about language inside and outside of the classroom, schools communicate both overt and covert ideologies to pupils about the value and usefulness of the languages they speak. It appears that the Welsh-medium school in this study is fostering an environment that is conducive to forming positive associations with Welsh, which may go a long way in promoting Welsh language revitalisation.

While the Welsh language has stabilised demographically over the past two decades in relative terms, it is again in decline, and there are numerous indications that the language attitudes of adolescents are at the heart of the issue. In recent years, language planners in Wales have sought to address this issue by providing young people with both the language skills they need and sufficient opportunities to use those skills. Relatively little attention, however, has been paid to the broader influence of the school environment. While it has been clear for some time that Welsh-medium schools produce Welsh speakers with much higher competency levels in the language than English-medium schools, these data suggest that there may be a rather substantial ideological divide as well. With respect to preparing young Welsh language learners to carry the language forward, attitude is just as critical as proficiency, and language planners need to focus a proportional amount of effort in that direction. A better understanding of the school’s role in developing positive attitudes could open up opportunities for increased Welsh language support in English-medium schools as well.

References


Welsh Government. (2012). Proportion of people (aged 3 and over) able to speak Welsh, by local authority, digital image, viewed 9 April 2015, Available at:


Rachelle Lee
Department of Language and Linguistic Science
University of York
Heslington
York
email: rachelle.lee@york.ac.uk
LISTENERS, WHICH TYPE OF “NATIVE EAR” WORKS BETTER?

CHEN SHEN, DOMINIC WATT

University of Sheffield, University of York

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
dialec 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 20\(^{th}\) April 2015).
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.

<table>
<thead>
<tr>
<th>Western</th>
<th>Sichuan Province</th>
<th>Gansu Province</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xinjiang 1</td>
<td>Sichuan 2</td>
<td>Gansu 3</td>
</tr>
<tr>
<td>Xinjiang 5</td>
<td>Sichuan 3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Northern</th>
<th>Heilongjiang Province</th>
<th>Hebei Province</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liaoning 1</td>
<td>Heilongjiang 2</td>
<td>Hebei 3</td>
</tr>
<tr>
<td>Liaoning 3</td>
<td>Heilongjiang 4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Southern</th>
<th>Guangdong Province</th>
<th>Jiangsu Province</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhejiang 1</td>
<td>Guangdong 2</td>
<td>Jiangsu 41</td>
</tr>
<tr>
<td>Zhejiang 3</td>
<td>Guangdong 3</td>
<td></td>
</tr>
</tbody>
</table>

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 a piece 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).](image)

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.

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean %</th>
<th>SE % &amp; SD %</th>
<th>Min %</th>
<th>Max %</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ET</td>
<td>68.3</td>
<td>3.81</td>
<td>12.6</td>
<td>44</td>
<td>18</td>
</tr>
<tr>
<td>EU</td>
<td>68.1</td>
<td>2.08</td>
<td>6.91</td>
<td>60</td>
<td>11.5</td>
</tr>
<tr>
<td>CT</td>
<td>94</td>
<td>1.52</td>
<td>4.81</td>
<td>84</td>
<td>6.5</td>
</tr>
<tr>
<td>CU</td>
<td>91.8</td>
<td>0.96</td>
<td>3.05</td>
<td>88</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean %</th>
<th>SE % &amp; SD %</th>
<th>Min %</th>
<th>Max %</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ET</td>
<td>30.9</td>
<td>2.93</td>
<td>9.73</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>EU</td>
<td>38.8</td>
<td>2.35</td>
<td>7.8</td>
<td>24</td>
<td>7</td>
</tr>
<tr>
<td>CT</td>
<td>38.4</td>
<td>3.13</td>
<td>9.9</td>
<td>24</td>
<td>18</td>
</tr>
<tr>
<td>CU</td>
<td>40.4</td>
<td>4.65</td>
<td>14.7</td>
<td>17</td>
<td>26</td>
</tr>
</tbody>
</table>

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.
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, 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 Chinese listeners’ familiarity and performance scores were tested. A correlation, albeit not a strong one, between the two factors was found ($r(18) = 0.38, p = .119$). 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).

---

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 ($\chi^2(1) = .040, p = .841$).

<table>
<thead>
<tr>
<th>Listeners</th>
<th>West (n = 4)</th>
<th>North (n = 7)</th>
<th>South (n = 9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>West</td>
<td>42.5</td>
<td>34.3</td>
<td>38.5</td>
</tr>
<tr>
<td>North</td>
<td>35.0</td>
<td>35.7</td>
<td>43.3</td>
</tr>
<tr>
<td>South</td>
<td>37.5</td>
<td>38.5</td>
<td>43.3</td>
</tr>
</tbody>
</table>

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.
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

<table>
<thead>
<tr>
<th></th>
<th>Most correctly categorised</th>
<th>Least correctly categorised</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 English listeners</td>
<td>Liaoning 3 (86%)</td>
<td>Jiangsu 41 (25%)</td>
</tr>
<tr>
<td>L1 Chinese listeners</td>
<td>Liaoning 1 (100%)</td>
<td>Jiangsu 41 (60%)</td>
</tr>
<tr>
<td></td>
<td>Zhejiang 3 (100%)</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Most correctly categorised</th>
<th>Least correctly categorised</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 English listeners</td>
<td>Korea-R2 (86%)</td>
<td>Korea-R3 (57%)</td>
</tr>
<tr>
<td>L1 Chinese listeners</td>
<td>Japan-R2 (100%)</td>
<td>Japan-R3 (100%)</td>
</tr>
<tr>
<td></td>
<td>Korea-R3 (100%)</td>
<td>Korea-R3 (80%)</td>
</tr>
</tbody>
</table>

Table 6: The most correctly and incorrectly categorised non-CE speakers (speakers labelled by country of origin plus identifying code number)
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 /l/, 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 /l/ as [ɻ] and [p], respectively. /l/’s voiced counterpart /l/ 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 [b] or [v]. 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 /a/ 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 [ɑ] 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


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