Smith, R., Holmes-Elliott, S., Pettinato, M. & Knight, R.-A. (2014). Cross-accent intelligibility of speech in noise: Long-term familiarity and short-term familiarization. Quarterly Journal of Experimental Psychology, 67(3), pp. 590-608. doi: 10.1080/17470218.2013.822009

Cross-accent intelligibility of speech in noise: Long-term familiarity and short-term familiarization

Rachel Smith[a], Sophie Holmes-Elliott[a], Michèle Pettinato[b] and Rachael-Anne Knight[c],

[a] Glasgow University Laboratory of Phonetics, University of Glasgow, United Kingdom
[b] Division of Psychology and Language Sciences, University College London, United Kingdom
[c] Department of Language & Communication Science, City University London, United Kingdom

Corresponding author:
Dr. Rachel Smith
Glasgow University Laboratory of Phonetics
School of Critical Studies
University of Glasgow
12 University Gardens
Glasgow G12 8QQ
United Kingdom

Rachel.Smith@glasgow.ac.uk
Phone: +44 141 330 5533
http://www.gla.ac.uk/departments/englishlanguage/staff/rachelsmith/

Short title: Cross-accent intelligibility of speech.

Abstract

Listeners must cope with a great deal of variability in the speech signal, and thus theories of speech perception must also account for variability, which comes from a number of sources, including variation between accents. It is well-known that there is a processing cost when listening to speech in an accent other than one's own, but recent work has suggested that this cost is reduced when listening to a familiar accent widely represented in the media, and/or when short amounts of exposure to an accent are provided. Little is known, however, about how these factors (long-term familiarity and short-term familiarization with an accent) interact. The current study tested this interaction by playing listeners difficult-to-segment sentences in noise, before and after a familiarization period where the same sentences were heard in the clear, allowing us to manipulate short-term familiarization. Listeners were speakers of either Glasgow English or Standard Southern British English, and they listened to speech in either their own or the other accent, thereby allowing us to manipulate long-term familiarity. Results suggest that both long-term familiarity and short-term familiarization mitigate the perceptual processing costs of listening to an accent that is not one's own, but seem not to compensate for them entirely, even when the accent is widely heard in the media.

Introduction

This paper investigates how listeners respond to variations in speech related to regional accent. Specifically, we test whether or not exposure to another accent in the short and/or long term enables listeners to process it as fluently as their own, and discuss the implications for the perceptual representation of accent variation and the nature of perceptual learning about speech.

Variation in speech: idiolect and accent

It is well known that speech exhibits a great deal of variability. Differences in pronunciation may be caused by attributes of the speaker, including those which are fairly temporary (such as emotional state or attitude), and those which are relatively fixed (such as idiolectal and accentual variations). Listeners must deal with these sources of variation, but can usually understand the speech they hear, even in difficult listening conditions. Theories of speech perception must also deal with variation, and explain how variable word forms are mapped on to lexical representations. There are two main theoretical approaches to dealing with variation (see Tenpenny, 1995 for a review). Abstractionist theories suggest that variation is stripped from the signal, which is then matched to a single invariant lexical representation (e.g. Halle, 1985). Exemplar accounts, on the other hand, posit that the signal is matched to multiple stored representations of the 'same' word, with recognition being easier and faster the more similar the signal is to the stored representations (e.g. Goldinger, 1996, 1998; Johnson, 1997, 2006). Recent models (e.g., Pierrehumbert, 2002) take a hybrid approach, and consider that representations have both abstract and exemplar aspects, and that multiple levels of representation are necessary.

In recent years, adaptation to variation in the speech signal has been a focus of research (see e.g. Samuel and Kraljic, 2009, for a review). Two main areas, relevant to our purpose in this paper, are perception of idiolects and perception of accents.

Much work on perception of idiolects has indicated that listeners develop separate representations for words spoken by different speakers. Goldinger (1998) for example, reviews a number of studies which show that words are recognized more quickly, and more accurately when they are repeated in the same voice, than when they are repeated in a different voice. Furthermore, listeners who are successfully trained to recognize particular talkers are then better at recognizing words spoken by those talkers than words spoken in new voices, whereas no such advantage is found for listeners who are poorer at recognizing talkers. Taken together, these and similar results point towards the use of episodic information at some stage in the process of spoken word recognition (although support for abstract representations in some circumstances is also found in the literature e.g. McLennan et al., 2003).

The question of how words are represented when they are pronounced differently in different accents is also at issue. It is well known that listeners perform less well when listening to speech in an accent other than their own, as shown by experiments using a range of paradigms, regional accents, participant age groups and speech styles (e.g. Nathan, Wells & Donlan, 1998; Nathan & Wells, 2001; Fraser Gupta 2005; Floccia, Goslin, Girard, & Konopczynski, 2006; Floccia, Butler, Goslin, & Ellis, 2009; Adank, Evans, Stuart-Smith, & Scott, 2009; Sumner & Samuel, 2009). The typical pattern is that speech in one's own accent is easier to understand than speech in a different regional accent, which in turn is easier to understand than non-native accented speech (Adank et al., 2009; Floccia et al., 2006; but see also Floccia et al., 2009). The processing costs associated with listening to other accents may arise because the acoustic-phonetic input from another accent produces either weak activation of intended words, and/or inappropriate activation of competitor words (Clopper, Pierrehumbert & Tamati, 2010; Dahan, Drucker & Scarborough, 2008). Alternatively the costs may relate to the difficulty of making perceptual switches when exposed to a range of accents (Floccia et al., 2006; Floccia et al., 2009). However, as we discuss below, the processing costs of listening to another accent appear to be mitigated to some extent by exposure to that accent, whether in the long term (e.g. via the media; Floccia et al., 2006; Clopper & Bradlow, 2008; Adank et al., 2009), or in the short term (e.g. in experimental situations; Clarke & Garrett, 2004; Dahan et al., 2008; Maye, Aslin & Tanenhaus, 2008), again suggesting the use of episodic information in the recognition process.

The role of long-term familiarity with an accent

Long-term familiarity with an accent other than one's own appears to reduce, or possibly even remove, the difficulty of understanding that accent. For example, Adank et al. (2009) found that speakers of Glasgow English (GE), a non-standard regional variety of British English, showed no processing delay when listening to Standard Southern British English (SSBE), a standard variety. In contrast, SSBE speakers did show a processing delay when listening to GE. Floccia et al. (2006) report a similar pattern of findings for a variety of regional French accents and the standard Parisian variety. The authors of such studies suggest that the findings are due to familiarity with the standard accents via media exposure and/or dialect contact.

Likewise, in a study of U.S. varieties, Sumner & Samuel (2009) found that long-term familiarity with an accent conferred a perceptual advantage, especially if the familiar accent was the U.S. 'media-standard', General American (GA). They investigated how r-ful and r-less variants of words (e.g. [beɪkɚ] and [beɪkə], baker) were perceived by speakers from New York City (NYC), a mainly non-rhotic dialect area, and by speakers of GA, which is rhotic. In a range of priming tasks, they found that non-rhotic NYC speakers showed no impairment in processing standard, r-ful variants, presumably due to their familiarity with GA. GA speakers, in contrast, did show an impairment in processing non-standard, r-less variants, presumably due to unfamiliarity with them. Importantly, rhotic speakers from NYC were also tested and showed an impairment in processing r-less variants in certain tasks. This was a similar result to that of GA speakers, despite the fact that the NYC speakers presumably had long-term familiarity with the prevailing non-rhotic speech in New York.

To explain these results, Sumner & Samuel (2009) suggest that non-rhotic NYC listeners' long-term familiarity with GA causes them to store two (r-less and r-ful) sets of perceptual representations for /r/-final words; one for their own accent, and one for the 'media-standard'. On the other hand, rhotic NYC speakers' familiarity with non-rhotic speech merely improves the efficiency with which they map r-less input to a single r-ful perceptual representation. Sumner & Samuel's view of a special representational status for the 'media-standard' GA is supported by Clopper & Bradlow's (2008) finding that GA was more intelligible in noise than speech from other U.S. accent regions, not only to GA speakers, but also to speakers of the other accents tested.

Thus, the general picture which emerges from the literature is that standard accents do not incur a processing cost, presumably due to them being familiar from the broadcast media, and therefore possibly having dedicated representations for use in word recognition. In some ways this is a surprising picture. It suggests powerful influences from exposure to standard accents via the media and/or dialect contact. Yet other research suggests that media influences are complex and subtle (Stuart-Smith et al., submitted; Evans & Iverson, 2007). And while data are lacking on the amount of exposure that speakers of regional accents have to standard accents, this is presumably small compared to exposure to their own accent. Therefore a question worth exploring is whether a processing cost might emerge for standard accents under a sufficiently challenging perceptual task.

The role of short-term familiarization to an accent

In addition to the effect of long-term familiarity with an accent, rapid perceptual shifts can occur after short-term exposure to real or simulated accent variation. Maye, Aslin & Tanenhaus (2008) exposed listeners to a 20-minute passage containing vowels that had been shifted in terms of phonemic category: each front vowel phoneme was replaced with the next lowest front vowel phoneme, e.g. witch /wɪʧ/ became wetch

/wɛʧ/. After exposure, listeners were more willing to accept shifted variants such as wetch as lexical items. Similarly, Dahan, Drucker & Scarborough (2008) showed that exposure to /æg/-final words in an accent where /æ/ is raised before /g/, but not /k/, improved listeners' speed of discrimination not only of /g/-final but also of /k/-final words that were never presented in the exposure period. They interpret this result as showing that exposure altered the dynamics of the lexical competition process. In both these studies, the shifts were specific to the parts of the phonetic/phonological space that had been manipulated, but there was generalization beyond the particular items tested to novel, phonologically-similar items. These findings parallel demonstrations of perceptual learning about voices (Nygaard & Pisoni, 1998; Norris, McQueen, & Cutler, 2003; Eisner & McQueen, 2005; Kraljic & Samuel, 2006).

The above familiarization studies manipulated very specific, limited properties of the accent in which stimuli were spoken. However, when more naturalistic speech materials are used, the evidence that short-term familiarization improves speech perception is more mixed. Such improvements have been documented for non-native accented speech, both for intelligibility (where the extent of improvement seems to depend on initial intelligibility; (Bradlow & Bent, 2008) and for response times (Clarke & Garrett, 2004). However, Floccia et al. (2006) found that improvement in response times for non-native speech was less rapid and less extensive than that observed by Clarke & Garrett (2004), and also failed to find evidence for improvement using regional accents, as did Floccia et al. (2009). Thus it remains unclear to what extent listeners can adapt in the short term to accents other than their own.

Experimental purpose

The current paper addresses two issues in adaptation to accents which remain poorly understood. The first issue concerns the relationship between perception of one's native regional accent vs. a long-term familiar accent such as the media standard: We probe whether processing costs can be observed for standard accents under challenging perceptual conditions. The second issue concerns the relationship between long-term familiarity and short-term familiarization with an accent: What is not clear from the contradictory findings described above is whether short-term familiarization facilitates processing of speech in a regional accent that is different to the listener's own. If it is facilitatory, it is unclear whether it has different effects depending on whether the accent is unfamiliar to the listener or already has a degree of long-term familiarity via, for example, the broadcast media. Also unknown is the extent to which the benefit of familiarization with an accent is contingent upon the particular voice with which participants are familiarized.

Thus, in the experiments described below we aim to establish how listeners perform when they differ in terms of their long-term familiarity and short-term familiarization with a regional accent. We compare the performance of Glasgow English and Standard Southern British English listeners when they listen to either their own or the other accent in noise, before and after familiarization with that accent in good listening conditions.

Method

Overview and design of experiment

The pair of accents we investigated were Glasgow English (GE) and Standard Southern British English (SSBE). By Glasgow English, we refer to a range of varieties belonging to the Scottish English continuum typical of the West Coast of Scotland, whose Scots end derives from West Central Scots, and whose Standard English end is Glaswegian Standard English (cf. Stuart-Smith 1999; 2003). By Standard Southern British English, we refer to varieties spoken in England, ranging from local varieties characteristic of the south-east of England, to the less regionally marked Received Pronunciation (Wells, 1982). GE and SSBE are highly divergent from one another in a range of phonetic and phonological properties: their vowel systems and vowel realizations, rhoticity, various aspects of consonant system and realisation, and suprasegmentals (Wells, 1982; Stuart-Smith, 1999, 2003, 2004).

The choice of this accent pair allowed us to explore the following degrees of long-term familiarity with an accent:

1) An accent which is the same as the listener's own, and thus entirely familiar
    a. GE for GE listeners
    b. SSBE for SSBE listeners
2) An accent which is a 'media-standard' variety and thus likely to be familiar to listeners who have the other accent (SSBE for GE listeners)

3) An accent which is non-standard and receives relatively little media representation, and is thus likely to be unfamiliar to listeners with the other accent (GE for SSBE listeners).

Short-term familiarization to an accent was assessed by testing perception before and after familiarization to speech materials presented in the listeners' own accent or the other accent. Finally, by manipulating the individual speaker heard in the familiarization period, we were able to address to what extent learning about an accent is specific to, or generalizes beyond, the particular voice tested, and to consider the relationship between learning about voices (or idiolects) and learning about accents.

Thus, the experiment consisted of a pre-test, familiarization phase and post-test. In the pre- and post-test, participants heard short, decontextualized low-predictability sentences in background noise, and typed (in standard orthography) what they heard. In the familiarization phase, listeners heard the same sentences, but each was preceded by a disambiguating context and presented without background noise, and listeners answered questions about their meaning. Performance between pre-test and post-test was compared, as was performance in the familiarization phase for different conditions.

To summarize, Table 1 shows the three crossed factors that were manipulated: Listener Accent (GE or SSBE), Test Accent (the accent heard in the pre- and post-tests (GE or SSBE), and Familiarization Voice (Identical to the voice in the pre- and post-tests), or Novel). Every participant heard the same single voice in the pre- and post-tests (the Test voice), and half the participants heard this voice in the familiarization period too, with the other half hearing a novel voice with the same accent as the Test voice.

Thus there were eight conditions, labelled according to Listener Accent, Test Accent and Familiarization Voice: for example, GGident refers to the group with GE Listener Accent who heard the GE Test Accent and were familiarized with the Identical voice that they heard in the tests, while GSdiff refers to the GE Listener Accent group who heard the SSBE Test Accent and were familiarized with a Different voice from that heard in the tests. Note that our design was between-subjects not only with respect to Listener Accent, but also Test Accent and Familiarization Voice (i.e., each listener heard only one accent in the experiments, and was familiarized with only one voice). This allowed us to offer an appropriately long period of short-term familiarization while keeping the experiment manageable in duration.

(Table 1 about here)

We predicted 1) better performance in the pre-test and familiarization period when the Listener Accent and Test Accent were the same (i.e. when listeners heard their own accent, so an advantage for GG and SS groups over SG and GS groups). In terms of the expected degree of improvement after familiarization, we predicted 2) stronger improvement for conditions where initial performance was better: if perceptual learning is meaning-driven (Norris, McQueen & Cutler, 2003; Davis et al., 2005), better initial access to meaning should lead to more effective learning (see Bradlow & Bent, 2008, for similar argumentation). Finally, we expected to see 3) a greater improvement at the post-test when familiarization was in the identical voice as in the pre- and post-test, but predicted that this advantage would be reduced when the accent has less long-term familiarity, since in this case there is a potential advantage to learning about properties that characterize the accent as well as those that are specific to the speaker.

Participants

Participants in the GE Listener Accent conditions were 77 speakers (41 female) of Glasgow English, who were randomly assigned to conditions 1 to 4 of Table 1 (condition 1: 19 participants; condition 2: 18 participants; conditions 3 and 4: 20 participants each). All were aged between 17 and 34 (mean 20.1) and were students or staff of the University of Glasgow. They were tested at the University of Glasgow in a sound-treated room.

Participants in the SSBE Listener Accent conditions were 80 speakers (57 female) of Standard Southern British English (SSBE), who were randomly assigned to conditions 5 to 8 of Table 1 (20 participants in each). All were aged between 18 and 35 (mean 21.7). Participants in conditions 5 and 6 were students or staff of City University London and were tested at City University London in a sound-treated room. Participants in conditions 7 and 8 were students or staff of the University of Cambridge and were tested at the University of Cambridge in a sound-treated room, except for 8 participants who were students or staff of the University of Glasgow, and were tested at the University of Glasgow in a sound-treated room. These 8 participants were all SSBE speakers who had moved to Glasgow recently prior to testing, and, as they were being tested on SSBE speech, their exposure to GE was not considered a cause for concern. Data from participants in conditions 7 and 8 were previously reported in Smith & Hawkins (2012).

Participants filled in a questionnaire identifying where they had grown up and the areas where they had lived. GE participants had grown up in Glasgow or the surrounding urban conurbations. SSBE participants had grown up in regions of England where SSBE is spoken. No participant had spent any time living in the other dialect region, with the exception of 3 GE participants who had briefly lived in Southern England during infancy. In line with similar studies, we did not control for exposure/engagement with broadcast media in either accent, or for possibilities for dialect contact with GE or SSBE, a point we will return to in the discussion. Participants were paid a small fee.

Materials

Stimuli were 24 pairs of sentences that were phonemically identical (in SSBE) but differed (usually grammatically) in a critical portion, underlined, e.g. a) So he diced them vs. b) So he'd iced them; a) Those are cat size vs. b) Those are *cats' eyes* (full list in Appendix A; average frequencies of the a and b members of the pairs in Table 2). Thus, for SSBE the two phrases in every pair corresponded to two different parses of the same phoneme string, and were therefore difficult to segment. Allophonic cues can be expected to provide some limited evidence as to which parse is intended, but these cues may differ across accents (Abercrombie, 1979). The materials were originally designed for use in a study of SSBE (Smith & Hawkins, 2012). Thus for GE, six pairs of sentences were not phonemically identical. For example, in the pair That surprise for the child vs. *That's a* prize for the child, the first member of the pair contains, in GE, a coda /r/ in the first syllable of surprise. We address this point in the Results section below.

(Table 2 about here)

For each experimental sentence a meaningful context was prepared, which disambiguated members of the pair (Appendix A), e.g.

A: He wanted the carrots to cook fast. So he diced them.
B: The top of the cakes had come out looking uneven. So he'd iced them.

A list was prepared consisting, in random order, of 8 tokens of each of the (24 x 2 =) 48 experimental sentences, each with its context, plus 216 filler sentences. The complete sentence list thus had (8(48) + 216 =) 600 sentences. The number of 8 tokens per sentence was sufficient to ensure that no tokens needed to be either repeated during the 40-minute exposure period, or included in both the exposure and the tests. As a result, effects of familiarization could be attributed to learning about the speakers' pronunciation patterns for these sentences in general, rather than about individual tokens. Individual tokens of a sentence differed only in minor aspects of their pronunciation; there were no major differences across a given speaker's tokens of a given sentence in prosody or segmental realisations.

Two male speakers of SSBE (aged 27 and 53), and two male speakers of GE (aged 19 and 54), recorded the list of experimental and filler sentences. For each accent, the younger speaker was assigned to be the Test voice, and the older speaker was assigned to be the Familiarization voice for the novel-voice conditions. Informal piloting suggested that the degree of intelligibility of the SSBE Test talker to SSBE listeners was similar to that of the GE Test talker to GE listeners.

Recordings of the SSBE speakers were made in a double-walled IAC booth at the University of Cambridge using a Sennheiser MKH40 P48 condenser microphone and a

Sony 55ES High Density Linear A/D D/A recorder, and were digitised at 16 kHz to a Silicon Graphics machine running xwaves. Recordings of the GE speakers were made in a double-walled IAC booth at the University of Glasgow using a Sennheiser MKH40 P48 condenser microphone, a Symetrix 302 pre-amplifier and an Edirol UA1-EX A/D D/A converter; they were recorded direct to disc at a sampling rate of 44.1 kHz, and downsampled to 16 kHz. The speakers read as naturally and informally as possible. Disfluent tokens were omitted and additional tokens recorded, to achieve a total of 8 tokens per sentence per speaker.

For the pre- and post-test stimuli, one token of each sentence (e.g. So he diced them) was selected at random from those recorded. This token was mixed with randomly-varying cafeteria noise at an average signal-to-noise ratio (SNR) of +2 dB (average amplitude of sentence:average amplitude of noise). This SNR was selected based on informal piloting by the four authors. The noise was ramped up to its maximum amplitude over 5 s before the sentence began, continued at this average amplitude for 15 s after sentence offset and was then ramped down to zero over a further 5 s. The 15 s period was established in pilot tests as necessary for the response to be typed.

Stimuli for the familiarization phase were different tokens of the same 48 sentences, each in its disambiguating context (e.g. He wanted the carrots to cook fast. So he diced them.). For each speaker, six tokens of each context+sentence were selected at random from the seven remaining after removal of the token used for the pre- and post-test.

Stimuli in each phase of the experiment were presented in a fixed pseudo-random order, with the constraint that members of a sentence pair never occurred adjacent to one another.

Procedure

Participants were tested individually using high-quality Sennheiser headphones and a PC laptop running DMDX. Listeners took part in the pre-test (48 items, 25 minutes), then the familiarization phase (288 items, 40 minutes), and finally the post-test (48 items, 25 minutes). The pre- and post-tests were each preceded by one practice item and the familiarization phase by two practice items.

In the pre- and post-tests, listeners heard decontextualised sentences in noise, and their task was to type what they heard into a computer. They were instructed to type as many words as they had understood, and the words they entered appeared on the screen as they typed.

In the familiarization phase, listeners heard each sentence in its disambiguating context without background noise. They were told that they would hear descriptions of events, and should judge whether it was LIKELY or UNLIKELY that an event involved the object, person, emotion or idea specified by a question displayed (for 3 s) on the computer screen immediately after they had heard the sentence. For example, for the sentence He wanted the carrots to cook fast. So he diced them, questions (with their correct responses in parentheses) included “Does the event involve something orange (LIKELY) / using a knife (LIKELY) / a telephone call (UNLIKELY) / strong emotion

(UNLIKELY)?", etc. Listeners responded by pressing one of two labelled keys on a keyboard as fast as possible, and their reaction time and accuracy were measured. Assignment of responses to hands was counterbalanced over participants. A short break occurred after every 20 items, and a longer self-paced break half-way through the familiarization phase. Listeners were offered self-paced breaks between pre-test, familiarization phase and post-test.

Results

For the orthographic data from the pre- and post-test, the words correctly reported were scored for each listener separately. Words had to be in the same order as in the test sentence in order to be scored as correct. Obvious mis-spellings and homophones were scored correct, morphological variants incorrect.

For the reaction time data from the familiarization phase, incorrect responses (6%) were removed, as were those that were longer than 5000 ms, or shorter than 200 ms. Next, data points for each subject were examined separately. Data points were removed from the two extremes of the distribution, until successive points were within 2 ms. For most subjects this involved the removal of only the single highest and lowest response time. Distributions were then visually examined for normality and RTs were log-transformed.

As explained above, the sentences had initially been designed for SSBE, and thus 6 pairs did not form minimal pairs in GE speech. Data from the pre- and post-tests, and for the familiarization phase, were initially analysed with and without these pairs, but, as no differences were found, results are reported below for the entire data set.

Task performance in same-accent vs. other-accent

We began by testing our first prediction: that processing would be more efficient when the Test accent was the same as the listener's accent, even when the Test accent was a media standard accent with which all listeners are likely to have long-term familiarity. Note that this prediction can be assessed in two ways, by comparing a) whether participants who have a given Listener accent respond differently to the two Test accents, and b) whether a given Test accent is responded to differently by the two Listener accent groups. Previous studies have taken approach a) (cf. Adank et al., 2009). Since we were sceptical of the idea that standard, long-term familiar accents incur no processing cost relative to native accents, we chose instead to take approach b), and our planned comparisons focus on whether a given Test accent is responded to differently by the two Listener accent groups.

We analysed the orthographic accuracy data from the pre-test, and, separately, the RT data from the familiarization phase. These analyses allowed us to assess how listeners process accented speech-in-noise before and during exposure to the materials and test accent. Performance after exposure is dealt with separately below.

Orthographic accuracy at pre-test

The percentage of correctly-reported words in the pre-test is shown in Figure 1a. Results were analysed using generalized linear mixed-effects modelling (Baayen, 2008). The model fitted had as predictors Listener Accent (GE vs. SSBE) and Test Accent (GE vs. SSBE), and their interaction, plus random effects for Subject and Sentence. Full statistical results are in Table 3. Effect sizes associated with significant pairwise differences were calculated using Cohen's d, i.e. $\bar{x}_1 - \bar{x}_2 / s$, where s is defined as the total between-subjects standard deviation.

(Figure 1 about here)

Accuracy was low overall, at 13-34% according to condition: these accuracy rates parallel those for unpredictable keywords in the SPIN test (Kalikow, Stevens, & Elliott, 1977) at S/N ratios between -5 dB and 0 dB, i.e. substantially lower than used here. The low rates likely reflect the casual speech style adopted by the speakers, and the susceptibility of their individual voices to the babble masker.

We found a significant interaction between Listener Accent and Test Accent ($p < .0001$). For both Test accents, words were more accurately identified by listeners who had the same accent than by listeners who had the other accent (GG vs SG: 12 percentage points, $z = -8.4$, $p < .0001$, Cohen's $d = .93$; SS vs GS: also 12 percentage points, $z = 6.7$, $p < .0001$, Cohen's $d = .72$). We note, however, the salient pattern in Figure 1 whereby SSBE listeners who heard the GE voice were much less accurate than SSBE listeners who heard the SSBE voice (SG vs SS: 21 percentage points) while GE listeners who heard the SSBE voice were no less accurate than GE listeners who heard the GE voice (GS vs GG: 2 percentage points). We note also that there appears to be a difference between the two same-accent groups, with the GG group performing more poorly, by 9%, than the SS group. In summary, the planned comparisons support prediction 1, and suggest that processing is more efficient when the Test accent is the same as the listener's accent, even if the Test accent has long-term familiarity. Nonetheless, in line with previous research, SSBE listeners appear to be much more affected by hearing a different accent than GE listeners are.

Response times in familiarization phase

Figure 1c shows the results for the familiarization phase, i.e. the time taken to correctly respond LIKELY or UNLIKELY to comprehension questions of the form "Does the event involve X?". RTs were analysed using linear mixed-effects modelling. The model fitted (Tables 3 and 4) had as predictors Listener Accent (GE vs. SSBE), Test Accent (GE vs. SSBE), Trial, and the interaction of all three factors. Random effects for Subject and Sentence were also included. As accuracy was 94% overall, error rates were not analysed.

We again found a significant interaction between Listener Accent and Test Accent ($p = .001$). In planned comparisons, when listening to the SSBE test accent, SSBE listeners responded significantly faster than GE listeners (SS vs GS: 1440 ms vs. 1723 ms, $t = -4.67$, $p = .0002$, Cohen's $d = .46$). In contrast, when listening to the GE test accent, GE listeners did not respond significantly faster than SSBE participants (GG vs SG: 1710 ms vs. 1815 ms, $t = 0.01$, $p = .999$). Again, as in the pre-test, we note the pattern whereby SSBE listeners who heard the GE test accent responded much more slowly than SSBE listeners who heard the SSBE test accent (SG vs SS: 1815 ms vs. 1440 ms), whereas GE listeners who heard the SSBE test accent responded no slower than GE listeners who heard the GE test accent (GS vs GG: 1723 ms vs. 1710 ms). Of the two same-accent groups, the GG group responded more slowly than the SS group (1710 ms vs. 1440 ms, respectively). Finally, there was a strongly significant effect of Trial, and Trial interacted significantly with Listener Accent and Test Accent. Figure 2 shows that RT decreased as the experiment went on for all listeners, but the most pronounced decrease was in the SS condition.

(Figure 2 about here)

In summary, the results indicate that listening to an accent other than one's own causes some impairment, even if the accent is familiar (supporting prediction 1). At the same time, there is a limitation on the generality of this effect: the impairment when listening to another accent is clearly greater when the accent is unfamiliar (i.e., for SSBE listeners compared to GE listeners).

(Table 3 about here)

Task improvement in same-accent vs. other-accent listeners

We went on to explore listeners' improvement at the task, and thus the role of short-term familiarization, by including in our statistical modeling the predictor Test (Pre-test vs Post-test), and its interactions with other factors. We began by fitting a full model with the predictors Test (Pre-test vs. Post-test), Listener Accent (GE vs. SSBE), Test Accent (GE vs. SSBE), and Familiarization Voice (Identical to test voice or Novel), and all their interactions, plus random effects for Subject and Sentence. Non-significant predictors were sequentially removed until the most parsimonious model had been found. Statistical results for the final model are in Table 4. Note that predictors close to significance ($.05 < p < .15$) were retained in the models so that they could be explored.

Figure 3 shows that all listeners' performance improved dramatically between pre- and post-test (average improvement: 25%; Test: $p < .0001$). However GE listeners improved more than SSBE listeners did (26% vs. 23%; Test x Listener Accent: $p < .05$). Whilst, from Figure 3, this result is driven by the 'GE test accent' condition, there is no significant interaction to support this interpretation (Test x Listener Accent x Test Accent $p = .13$). Further, listeners improved less with the GE test accent than with the SSBE test accent (22% vs. 27%; Test x Test Accent: $p < .025$). Improvement with the GE speech was, numerically, much weaker for the SSBE listeners (19%) than the GE listeners (26%), but again the three-way interaction was non-significant (Test x Test Accent x

Listener Accent, as above). Thus, in summary, SSBE listeners improved less over the course of the task than GE listeners, and GE test speech also evoked less improvement than SSBE speech, with the lowest improvement score occurring in the SG condition.

(Figure 3 about here)

Figure 4 shows that listeners improved more at the task when the voice they heard in the familiarization phase was identical to the voice they were tested on (Test x Familiarization Voice: $p < .0001$). Neither the Listener Accent nor the Test Accent significantly affected the size of this identical-voice advantage, although two trends were observed that mirror the general effects of Listener Accent and Test Accent upon improvement at the task. That is, the SSBE test accent evoked marginally stronger identical-voice effects than the GE test accent (Test x Test Accent x Familiarization Voice: $p = .079$), and the GE listeners also showed slightly stronger identical-voice effects than the SSBE listeners (Test x Listener Accent x Familiarization Voice: $p = .105$). Thus, numerically, the smallest identical-voice advantage was again for the SG group.

(Table 4 about here)

(Figure 4 about here)

To summarize the results as a whole, in all phases of the experiment, both GE and SSBE listeners were affected to some extent by the accent in which speech was presented. Planned comparisons revealed that, as predicted, listeners who spoke with the same accent as the test accent (i.e. who were listening to their own accent) always identified words significantly more successfully than listeners who had the other accent. We also note that the interactive pattern from the perspective of listener group differences indicates an asymmetry related to long-term familiarity, congruent with previous literature. That is, those GE listeners who heard SSBE speech performed as well as those GE listeners who heard GE speech, whereas those SSBE listeners who heard GE speech performed much worse than those who heard SSBE speech. In the Discussion we attempt to reconcile this pattern with the main finding.

We also found that the extent of improvement conferred by short-term familiarization depended in complex ways on the test accent and listener's accent, with GE participants improving more than SSBE participants, and SSBE speech evoking more improvement than GE speech. One way of thinking about these results is that the speech comprehension situation which was most difficult initially—that of the SG group—improved least over the course of the experiment. Finally, results for all groups of participants showed, as expected, a benefit if short-term familiarization was in the identical voice as that used in the test phases, with a trend for the identical-voice advantage to be weaker for the most difficult condition, SG.

Discussion

This paper investigated the dependence of speech intelligibility on short- and long-term familiarity with an accent. We compared intelligibility in noise of the UK broadcast media standard accent (Standard Southern British English) and a non-standard regional accent (Glasgow English), to speakers of each accent, before and after 40 minutes' familiarization with the task and materials. The main contribution of the results is to show that the dependencies between a listener's accent, their presumed degree of long-term familiarity with a test accent, and the short-term familiarization they receive, are more complex than has been assumed in previous research: Perception of a media-standard accent can be challenging for listeners who speak with a non-standard regional accent, and short-term familiarization cannot fully compensate for lack of long-term familiarity.

Role of long-term familiarity

Our first prediction (see page 12) concerned the influence of long-term familiarity on intelligibility. We predicted that performance on our challenging task and materials would be better when listeners heard their own accent, even if the other accent was familiar to them.

GE listeners did respond less accurately and less quickly to SSBE speech than SSBE listeners did, and SSBE listeners responded less accurately and less quickly to GE speech than GE listeners did, demonstrating that listening to an accent other than one's own caused some impairment, even if the accent was familiar. However, an important qualification is that GE listeners who heard SSBE did not actually appear to perform worse than GE listeners who heard GE. This represents a substantial limitation on the generality of our results.

Though this conflicting pattern is surprising, it proves to be paralleled elsewhere in the literature. Although Adank et al. (2009) conclude that GE listeners process SSBE as efficiently as they process GE, closer inspection of their numerical results reveals a similar discrepancy to ours. That is, although their GE listeners performed equivalently on SSBE and on GE, on average they appear to have been slower and less accurate in comprehending SSBE than SSBE listeners were. And, again as in our results, GE listeners also appear to have performed less well on their own accent, GE, than SSBE listeners did on their own accent, SSBE. Adank et al. (2009) report a different set of planned comparisons to us, so results cannot be directly compared, but the similarity in the numerical patterns is striking.

We suspect that a key factor underlying the conflicting pattern in the present study is the poor performance of the GG group relative to the SS group. That is, even when GE listeners were presented with their own accent they performed relatively poorly. As a similar pattern is present in the results of Adank et al. (2009), it is unlikely to be due to sampling issues. If the GE speech was, for some reason, inherently more difficult to understand even by native GE listeners, then this difficulty would have masked the advantage conferred by its nativeness, leading, as a net result, to the absence of a difference between results for GG and GS groups.

This suggestion leads us to ask what could have been inherently difficult about the GE speech in the present study: the specific speakers chosen, the testing situation, or general properties of the accent? We consider these possibilities in turn. First, the GE Test speaker may have been more effectively masked by the background noise than the SSBE speaker, due to (e.g.) his pitch range or phonation quality, and we know that intelligibility of speech in babble noise can be affected by the native accent (Fox et al., 2010; Heinrich, Bruhn, & Hawkins, 2010) as well as the language (García Lecumberri & Cooke, 2006; van Engen & Bradlow, 2007) of the babble speakers. Our noise was that used by Tunley (1999), but whilst it seems likely that the speaker accents were SSBE, we have not been able to confirm this. Regardless, differential effects of masking could not account for Adank et al.'s similar finding, since they used multiple speakers and a different type of noise. Further testing of our materials produced by different and multiple speakers, and using different accents in the babble noise would elucidate this issue.

Alternatively our GE listeners may have been primed by the SSBE accent of the experimenter to expect SSBE, leading to their difficulties with their own accent; there may have been an effect of the fact that the materials were originally designed for SSBE listeners; or the formality of the laboratory testing situation might have led them to expect a formal register of speech, which, in our experience, they associate with SSBE. Floccia et al. (2009) showed that participants' expectations about the accents to be heard (based on experimental instructions) affected their response times, while Hay & Drager (2010) demonstrated that even when an experimenter simply attracts participants' attention in an incidental manner to a toy kangaroo, koala or kiwi this can markedly affect listeners' interpretation of incoming speech as spoken with an Australian or New Zealand accent. One might expect that switching accents or registers would have caused difficulties only at the start of our experiment rather than (as we observed) throughout, as some studies (e.g. Clarke and Garrett, 2004) find adaptation to an accent occurs within a few seconds. However, Floccia et al. (2009) found that listeners' reaction times never recovered from the initial disruption caused by an accent switch. It is difficult to tease this issue apart by conducting further tests of this type away from a formal laboratory setting, but future research could manipulate the accent of the experimenter to see, for example, if a GE speaking experimenter would elicit different results.

Finally, a less well understood factor—linguistic-internal difference—might contribute to the difficulty in intelligibility for the non-standard accent. Clopper et al.

(2008) found General American to be more intelligible than other American English accents, irrespective of the listener's own accent. Our results, taken together with Clopper et al.'s, suggest that possible linguistic bases for asymmetric mutual intelligibility between pairs of accents should be explored in future work. For example, more 'prototypical' accents (ones that share phonetic or phonological properties with many other accents) might be easier to understand than less prototypical accents, regardless of their familiarity. Alternatively, the phonetic realization of linguistic contrasts (such as Scottish varieties' more frequent resyllabification of syllable-final consonants; Abercrombie, 1979) could affect intelligibility. Further work is needed to establish whether phonological or phonetic differences between accents result in differential intelligibility. One way to approach this is with acoustic measures of relevant properties, such as differences in durational juncture cues (Smith and Knight, in prep).

In summary, we return to our original point concerning whether listeners who have a regional accent show any processing impairment when listening to the familiar 'media standard'. Previous studies (Clopper & Bradlow, 2008; Floccia et al., 2006; Adank et al., 2009) have concluded that the answer is no, but our results compel us to draw a different conclusion, namely that perception of a media-standard accent can be more challenging for listeners who speak with a non-standard regional accent than for those who speak with the standard accent—at least under the difficult conditions that our listeners experienced, namely hearing casually-spoken, difficult-to-segment phrases in background noise. However, there remain some intriguing asymmetries, which call for further research into whether all accents are equally intelligible to their own listeners.

Exposure, familiarity and perceptual representations of speech

We follow previous studies (Adank et al., 2009; Floccia et al., 2006, 2009) in taking the view that long-term familiarity is a plausible part of the explanation for our results, and the finding that speakers perform better with their own accent. However, in line with these and other similar studies we did not control for the actual amount of exposure that speakers had to the 'other' accent, which is almost impossible to measure with any accuracy, since people might well be exposed to accents in real life, or through the media, without realizing it. Nevertheless we did control for residence and parental birthplace, and thus most of our SSBE listeners are likely to have had little or no experience with GE, whereas our GE listeners are likely to have at least some experience listening to SSBE from the media or from encounters with SSBE speakers (e.g. at university).

Accordingly, the GE listeners presumably have some type of stored cognitive representation of SSBE, in a similar way to Sumner and Samuel's (2009) interpretation that non-rhotic New Yorkers have stored representations of General American. GE listeners can access these representations as SSBE speech is heard, and subsequently refine them. Even if accessing these stored representations is a resource-consuming process and not error-free, it will still result in reasonable task performance. In contrast, SSBE listeners, with less supposed direct or indirect experience of GE, do not have stored representations of GE, and must therefore process GE speech through the 'filter' of the cognitive representation of their own accent, until new accent-appropriate representations can be formed. Differences in stored representations could also affect listeners' processing of their own accent. That is, the more diverse nature of the GE listeners' exposure to accents and cognitive representations of those accents, might adversely affect their processing of their own accent relative to mono-dialectal SSBE listeners' processing of SSBE: episodic models predict that the aggregate response to a stimulus is stronger, the more tokens it matches. This suggestion could be tested by investigating whether SSBE listeners who have long-term familiarity with a different accent of English process SSBE less effectively than mono-dialectal SSBE listeners.

Interaction of long-term familiarity and short-term familiarization

Leaving aside questions of long-term familiarity, we also investigated short-term familiarization by looking at improvement between the pre- and post-test. We predicted firstly that we would find more improvement for conditions where initial performance was better (prediction 2 on page 12). We also expected to see a greater improvement at the post-test when familiarization was in the identical voice as in the pre- and post-test, but speculated that this advantage might be reduced when the accent has less long-term familiarity (prediction 3).

Supporting prediction 3, the experiment replicated previous demonstrations of perceptual learning for individual voices (Nygaard & Pisoni, 1998): familiarization with the same individual voice heard in the pre- and post-tests led to slightly greater improvement at the task than familiarization with a different voice, consistent with the predictions of exemplar or hybrid accounts of perception (e.g. Goldinger, 1998; Johnson, 2006; Lachs, McMichael, & Pisoni, 2003; Hawkins, 2003; Pierrehumbert, 2006).

SSBE speech elicited greater task improvement after familiarization than GE speech, while GE participants also showed more improvement. The participant group who performed most poorly overall (i.e. SSBE listeners hearing GE speech) also showed least improvement, and there was a trend for the facilitative effect of hearing the identical voice in all phases of the experiment to be smallest for this group. We predicted this pattern based on the availability of meaning (prediction 2). That is, if understanding meaning is critical to effective perceptual learning (Norris et al. 2003; Davis et al., 2005), speech that is harder to understand (such as that with which listeners have little long-term familiarity), and consequently poorly encoded initially, will trigger less effective perceptual learning about both the accent and the speaker (cf. Bradlow & Bent, 2008, for similar argumentation). A further possible explanation for the GE listeners' greater improvement after the familiarization phase is that, despite initial perceptual disruption caused by listening to an accent not their own, over the course of the experiment they were able to access underlying familiarity with, and stored representations of, SSBE speech. To summarize, then, it appears that short-term familiarization with a voice, accent and set of speech materials helps all listeners, but conveys an increased benefit for speech that is more intelligible at the outset: that is, contrary to Clark & Garrett (2004), short-term familiarization cannot fully compensate for lack of long-term familiarity.

Conclusion

It is clear from this and other studies on perception of regional accents that listeners process speech in their own accent more efficiently than in another accent. Although our results on long-term familiarity leave some intriguing questions unanswered, the clear conclusion from them must be that long-term familiarity with a standard accent does not equate to native-like perception as straightforwardly as has generally been assumed. What is more, short-term familiarization with an accent—perhaps unsurprisingly—does not compensate for the lack of long-term familiarity, and seems to benefit the listener only in proportion to how good performance was to begin with. On one level, these results qualify the idea that listeners are highly flexible and adapt rapidly to any aspect of variation that they encounter. On another, they are consistent with the fundamental predictions of episodic models of speech perception, namely that perceptual representations and consequent responses depend rather closely on a listener's specific experience, as shown by our demonstrations of perceptual learning for individual voices. More work is needed to explore the interplay of familiarity, social priming and phonetic/phonological factors in determining how speech in another accent is understood. The results we have observed suggest that exploring these factors will contribute to understanding the structure of the perceptual neighbourhood space and the way it alters in response to exposure to an accent, and will allow us to discover more about the bases of communication, miscommunication and possibly even change across accents.

Acknowledgements

This work was supported by an award to the first and last authors from the Nuffield Foundation (SGS/35300: Perceptual learning about word boundaries in familiar and unfamiliar accents).

Appendix: Experimental sentences and precursors

Critical portions of experimental sentences are underlined. * marks pairs of sentences that are not phonemically identical in GE; in these cases phonemic transcriptions for GE are given.

1a:We lined up all the wines from the cheapest to the most expensive. Then <u>we drank</u> them in order.
1b:We decided that we'd mark all the tests first. Then <u>we'd rank</u> them in order.

2a: I bought a new car stereo in time for the trip. <u>I drove</u> all over Europe playing music.
2b: I thought I would become a travelling bard. <u>I'd rove</u> all over Europe playing music.

3a: He wanted the carrots to cook fast. So <u>he diced</u> them.
3b: The top of the cakes had come out looking uneven. So <u>he'd iced</u> them.

4a: Her teenage son insisted all his T-shirts had to be black. <u>She dyed</u> them resentfully.
4b: She hadn't been able to afford the jeans in the shop window. <u>She'd eyed</u> them resentfully.

5a: All writers are terrified when their book first comes out. <u>We dread</u> the reviews.
5b: We expected the play to be unusual. <u>We'd read</u> the reviews.

6a: John thought they needed a challenge. He was pleased that <u>we dared</u> them.
6b: All Mark's clothes had been damp. He was pleased that <u>we'd aired</u> them.

7a: There are some people out there with a morbid streak. Apparently the gentlemen <u>collect skulls</u>.
7b: We have a dedicated bird collector living next door. Apparently the gentleman <u>collects gulls</u>.

8a: Most people would have borrowed money to pay for it. <u>Pete stole</u> money.
8b: She eventually realised what the cheque on the table was. <u>Pete's dole</u> money.

9a: Chocolate and sugar are comforting when you're in danger. That's why the airmen <u>eat sweet</u> products.
9b: Carbohydrates are good for endurance. That's why the airman <u>eats wheat</u> products.

*10a: "The Lord is my Shepherd" is popular at St. Mary's. The congregation certainly <u>like psalms</u>. GE /lʌɪk samz/.
*10b: The pastor and his flock are gun-toting NRA members. The congregation certainly <u>likes arms</u>. GE /lʌɪks armz/

11a: For a while the fallen trees blocked access. But <u>Pat sawed</u> them.
11b: To begin with they were unimpressed. But <u>Pat's awed</u> them.

12a: Among the dog baskets, Sarah was surprised to see some much smaller ones. She said, "Those are <u>cat size</u>".
12b: I wanted to know what the little lights in the road were. She said, "Those are <u>cat's eyes</u>".

*13a: "The film was sentimental, but one scene was genuinely moving," she said. "<u>That surprise</u> for the child." GE / ðat sərpraez /.

*13b: “See the mountain bike over there?” he said. “That’s a prize for the child.” GE /ðats ə praez/.

14a: I agree that Simon excels. But Ralph surpasses.
14b:Geoff’s grades are all distinctions. But Ralph’s are passes.

15a: Don’t say venereal. Say veneer for me.
15b: Don’t eat the whole of the chocolate rabbit. Save an ear for me.

*16a: John gave up pretty early. Ruth sustained. GE /ruθ səstend/.
*16b: I’m not sure whose trousers to borrow. Ruth’s are stained. GE /ruθs ər stend/.

17a: He hasn’t worked with the Gurkhas before. It’s no wonder he didn’t recognise that salute.
17b: He doesn’t know anything about music. It’s no wonder he didn’t recognise that’s a lute.

18a: There was still one thing she didn’t like about the fireplace. That surround.
18b: That song’s not a fugue. That’s a round.

*19a: They’d told him not to leave the house. But he went for a sly stroll. GE /slae strol/.
*19b: There were lots of fancy sandwiches on the table. But he went for a sliced roll. /slʌɪst rol/.

20a: After each show I arrange the props backstage, ready for the next day. And I lay Steve’s costume out in the wings.
20b: Just before the curtain went up, I did the actors’ make-up in the green room. And I laced Eve’s costume out in the wings.

21a: His parents take the idea of corporal punishment too far. They whack Stan’s legs quite painfully.
21b: I don’t recommend the King’s Road salon. They waxed Ann’s legs quite painfully.

*22a: I don’t know what caused my migraines. It may have been eye strain. GE /ae stren/.
*22b: You must have misread the sentence: hail isn’t acid rain. It may have been iced rain. GE /ʌɪst ren/.

23a: It’s not just that his foster parents are well off. They also offer Mick stability.
23b:The yoga centre has a few beginners’ classes this year. They also offer mixed ability.

24a: They’ll have a game of football with anyone that’s willing. They even play strangers in the park.
24b: The local authority took all sorts of measures to combat crime. They even placed rangers in the park.

References

Abercrombie, D. (1979). The accents of standard English in Scotland. In A. J. Aitken &. T. McArthur. (Eds.), Languages of Scotland (pp. 68-84). Edinburgh: Chambers.

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. Journal of Experimental Psychology: Human Perception and Performance, 35(2), 520-529. doi: 10.1037/a0013552

Baayen, R. H. (2008). Analyzing Linguistic Data. A Practical Introduction to Statistics Using R. Cambridge: Cambridge University Press.

Baayen, R., H., Piepenbrock, R. & Gulikers, L. (1985) The CELEX Lexical Database (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. Cognition, 106 (2), 707-729. doi: 10.1016/j.cognition.2007.04.005

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. The Journal of the Acoustical Society of America, 116 (6), 3647-3658. http://dx.doi.org/10.1121/1.1815131

Clopper, C. G., & Bradlow, A. R. (2008). Perception of dialect variation in noise: Intelligibility and classification. Language & Speech, 51, 175-198. doi: 10.1177/0023830908098539

Clopper, C. G., Pierrehumbert, J. B., & Tamati, T. N. (2010). Lexical neighborhoods and phonological confusability in cross-dialect word recognition in noise. *Laboratory Phonology, 1* (1), 65-92. doi: 10.1515/labphon.2010.005

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition, 108* (3), 710-718. doi:10.1016/j.cognition.2008.06.003

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, 13 4(2), 222-241. doi: 10.1037/0096-3445.134.2.222

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67* (2), 224-238. doi: 10.3758/BF03206487

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America, 121* (6), 3814-3826. doi: 10.1121/1.2722209

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in English: Can listeners adapt? *Journal of Psycholinguistic Research*, 38(4), 379-412. doi: 10.1007/s10936-008-9097-8

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? Journal of Experimental Psychology: Human Perception and Performance, 32 (5), 1276-1293. doi: 10.1037/0096-1523.32.5.1276

Fox, R. A., Jacewicz, E., & Chang, C.-Y. (2010). Speech intelligibility in cross-dialectal multi-talker babble. Journal of the Acoustical Society of America, 127, 1903.

Fraser Gupta, A. (2005). Inter-accent and inter-cultural intelligibility: A study of listeners in Singapore and Britain. In D. Deterding, A. Brown & E. L. Low (Eds.), English in Singapore: Phonetic research on a corpus (pp. 138-152). Singapore: McGraw-Hill (Education) Asia

Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. Journal of the Acoustical Society of America, 119, 2445-2454.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 1166–1183. doi: 10.1037/0278-7393.22.5.1166

Goldinger, S. D. (1998). Echoes of echoes? an episodic theory of lexical access. Psychological Review, 105, 251-269. doi: 10.1037/0033-295X.105.2.251

Halle, M. (1985). Speculations about the representation of words in memory. In V.A. Fromkins (Ed.), Phonetic Linguistics (pp. 101-104). New York: Academic Press.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. Journal of Phonetics, 31, 373–405. http://dx.doi.org/10.1016/j.wocn.2003.09.006

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. Linguistics, 48(4), 865-892. doi: 10.1515/LING.2010.027

Heinrich, A., Bruhn, K., & Hawkins, S. (2010). Masker efficiency of accented English background babble for British English sentences heard by young English adults. Poster presented at Psycholinguistic Approaches to Speech. Recognition in Adverse Conditions Workshop, Bristol, 8-10 March 2010.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In: K. Johnson, & J. W. Mullennix (Eds.), Talker variability in speech processing (pp. 145–165). San Diego: Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. Journal of Phonetics, 34(4), 485-499. doi:10.1016/j.wocn.2005.08.004

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. Psychonomic Bulletin & Review, 13(2), 262-268. doi: 10.3758/BF03193841

Lachs, L., McMichael, K., & Pisoni, D. B. (2003). Speech perception and implicit memory: Evidence for detailed episodic encoding. In J. S. Bowers, & C. J. Marsolek (Eds.), Rethinking implicit memory (pp. 215-235). Oxford: Oxford University Press.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. Cognitive Science, 32(3), 543-562. doi: 10.1080/03640210802336920

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29(4), 539-553. http://dx.doi.org/10.1037/0278-7393.29.4.539

Nathan, L., & Wells, B. (2001). Can children with speech difficulties process an unfamiliar accent? Applied Psycholinguistics, 22(3), 343-361. http://dx.doi.org/10.1017/S0142716401003046

Nathan, L., Wells, B., & Donlan, C. (1998). Children's comprehension of unfamiliar regional accents: A preliminary investigation. Journal of Child Language, 25(2), 343-365. http://dx.doi.org /10.1017/S0305000998003444

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. Cognitive Psychology, 47(2), 204-238. doi:10.1016/S0010-0285(03)00006-9

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. Perception & Psychophysics, 60(3), 355-376. http://dx.doi.org/10.3758/BF03206860

Pierrehumbert, J. (2002). Word-specific phonetics. In: C. Gussenhoven, & N. Warner (Eds.), Laboratory phonology (Vol. 7, pp. 101–139). Berlin: Mouton de Gruyter.

Pierrehumbert, J. B. (2006). The next toolkit. Journal of Phonetics, 34(4), 516-530. doi:10.1016/j.wocn.2006.06.003

Samuel, A.G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception & Psychophysics, 71*, 1207-1218. doi:10.3758/APP.71.6.1207

Smith, R., & Hawkins, S. (2012). Production and perception of speaker-specific phonetic detail at word boundaries. *Journal of Phonetics, 40*, 213-233 doi:10.1016/j.wocn.2011.11.003.

Smith, R., & Knight, R.-A. (in preparation). Phonetic cues to word boundaries in two accents of English.

Stuart-Smith, J. (1999). Glasgow: Accent and voice quality. In P. Foulkes, & G. Docherty (Eds.), *Urban voices: Accent studies in the British Isles* (pp. 203-222). London: Arnold.

Stuart-Smith, J. (2003). The phonology of modern urban Scots. In J. Corbett, D. McClure & J. Stuart-Smith (Eds.), *The Edinburgh companion to Scots* (pp. 110-37). Edinburgh: Edinburgh University Press.

Stuart-Smith, J. (2004). The phonology of Scottish English. In: B. Kortmann, , K. Burridge, E. Schneider, R. Mesthrie, and C. Upton, (Eds.), *A Handbook of Varieties of English: 1: Phonology* (pp. 47-67). Mouton de Gruyter, Berlin, Germany.

Stuart-Smith, J., Timmins, C., Pryce, G. and Gunter, B. (submitted). 'Television is also a factor in language change: Evidence from an urban dialect'.

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. Journal of Memory and Language, 60 (4), 487-501. http://dx.doi.org/10.1016/j.jml.2009.01.001

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. Psychonomic Bulletin & Review, 2(3), 339-363. doi:10.3758/BF03210972

Tunley, A. (1999). Coarticulatory influences of liquids on vowels in English (Doctoral dissertation, University of Cambridge).Van Engen, K. J., & Bradlow, A. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. Journal of the Acoustical Society of America, 121, 519-526.

Wells, J. C. (1982). Accents of English. Cambridge: Cambridge University Press.

Table 1.

Experimental conditions.

Group	Listener Accent	Test Accent	Familiarization phase voice
1 GGident	GE	GE	Identical GE speaker as Test
2 GGnovel	GE	GE	Novel GE speaker
3 GSident	GE	SSBE	Identical SSBE speaker as Test
4 GSnovel	GE	SSBE	Novel SSBE speaker
5 SSident	SSBE	SSBE	Identical SSBE speaker as Test
6 SSnovel	SSBE	SSBE	Novel SSBE speaker
7 SGident	SSBE	GE	Identical GE speaker as Test
8 SGnovel	SSBE	GE	Novel GE speaker

Table 2.

Mean log frequencies of critical phrases and of critical words in a and b members of sentence pairs (standard deviations in brackets). Frequencies of phrases are derived from Google Books (averaged over all time periods from 1810 to 2009). Word frequencies are from the combined spoken and written COBUILD databases in CELEX (Baayen et al., 1995).

	Mean log frequency of critical phrase (s.d.)	Mean log frequency of word 1 (s.d.)	Mean log frequency of word 2 (s.d.)
a sentences	2.13 (2.01)	2.04 (1.48)	0.60 (0.68)
b sentences	1.00 (1.41)	1.23 (0.96)	1.43 (1.45)

Table 3.

Results of mixed-effects modelling. *: χ^2, df and significance levels are for the factor plus interaction terms including that factor.

	Pre-test			Familiarization session		
Effect	χ^2	**df**	**p**	χ^2	**df**	**p**
Listener Accent	87.20	2 *	< .0001	77.50	4 *	< .0001
Test Accent	123.72	2 *	< .0001	53.36	4 *	< .0001
Listener Accent x Test Accent	86.53	1	< .0001	37.82	2 *	< .0001
Trial		n/a		352.84	4 *	< .0001
Trial x Listener Accent		n/a		56.68	2 *	< .0001
Trial x Test Accent		n/a		39.18	2 *	< .0001
Trial x Listener Accent x Test Accent		n/a		27.33	1	< .0001

Table 4.

Results of mixed-effects modelling. *: χ^2, df and significance levels are for the factor plus interaction terms including that factor.

Effect	χ^2	**df**	**p**
Test	6427.2	7*	< .0001
Listener Accent	84.6	6*	< .0001
Test Accent	124.8	6*	< .0001
Familiarization Voice	52.2	6*	< .0001
Test x Listener Accent	8.8	3*	.0328
Test x Test Accent	11.0	3*	.01149
Test x Familiarization Voice	44.7	3*	< .0001
Listener Accent x Test Accent	76.3	2*	< .0001
Listener Accent x Familiarization Voice	2.8	2*	.2421
Test Accent x Familiarization Voice	3.4	2*	.1821
Test x Listener Accent x Test Accent	2.3	1	.1295
Test x Listener Accent x Familiarization Voice	2.6	1	.1052
Test x Test Accent x Familiarization Voice	3.1	1	.0793

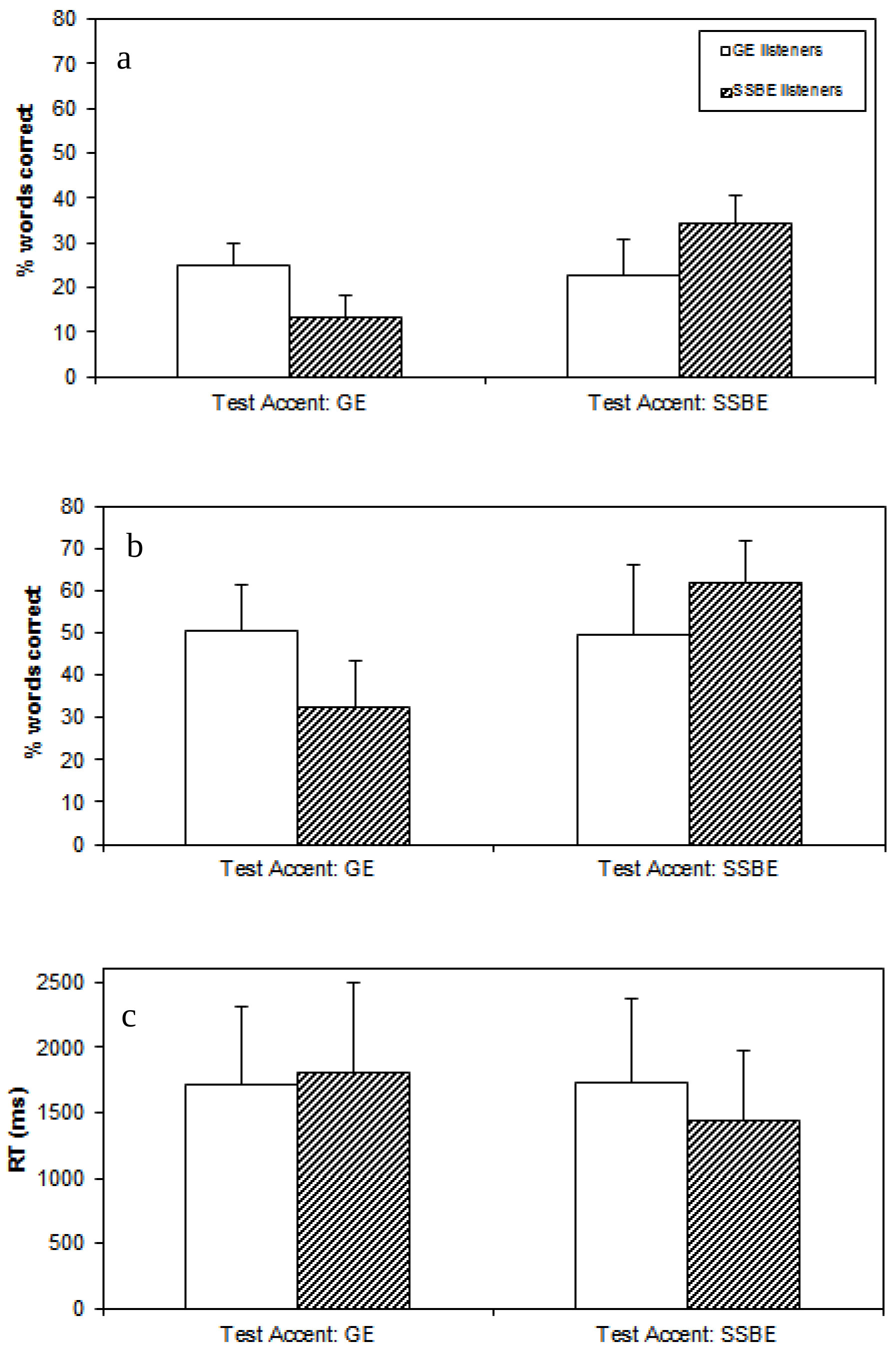

Figure 1. Results for experimental tasks, by Test Accent and Listener Accent. a, b: Percentage of words correctly reported in pre-test and post-test respectively. c: RT to respond correctly in familiarization period. Error bars represent 1 standard deviation.

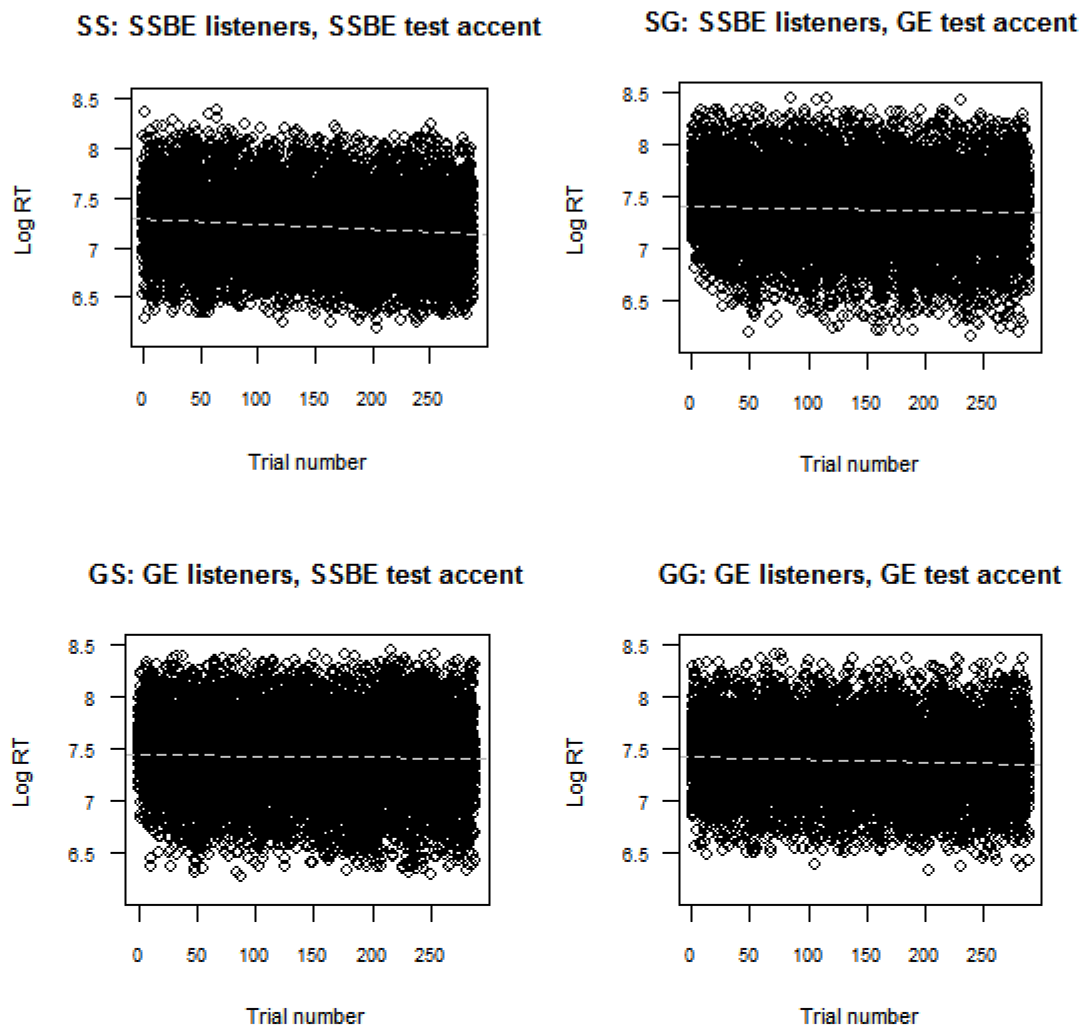

Figure 2. Evolution of reaction time as a function of experimental trial during the familiarization phase, according to Listener Accent and Test Accent.

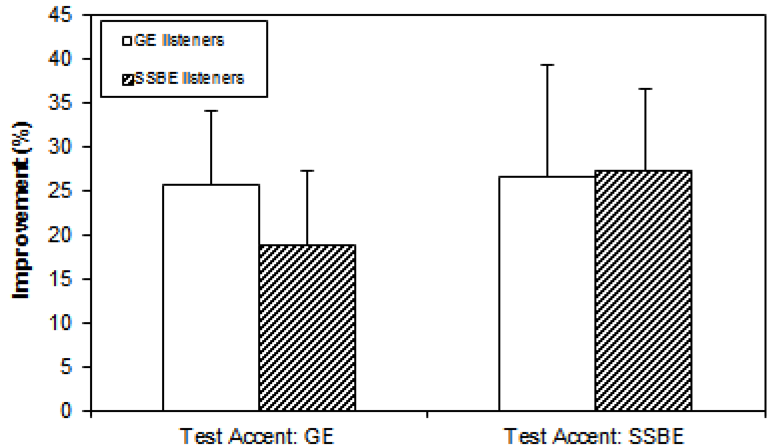

Figure 3. Improvement from pre- to post-test, by Listener Accent and Test Accent. Error bars represent 1 standard deviation.

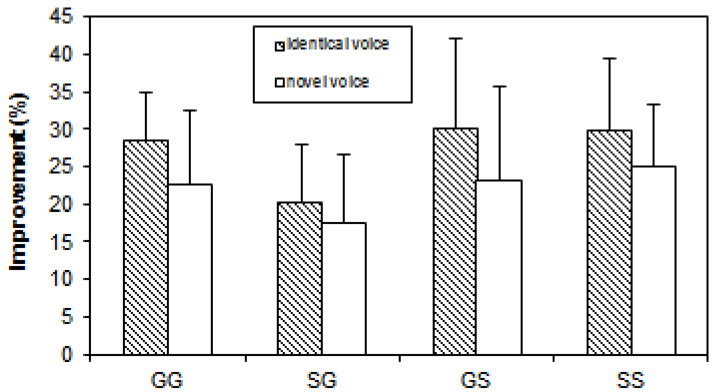

Figure 4. Improvement from pre- to post-test, by Test Accent, Listener Accent, and Familiarization Voice. Error bars represent 1 standard deviation.