Journal of Experimental Psychology:
Human Perception and Performance
2009, Vol. 35, No. 2, 520–529

0096-1523/09/$12.00 DOI: 10.1037/a0013552

Comprehension of Familiar and Unfamiliar Native Accents Under Adverse Listening Conditions

Patti Adank
Radboud University Nijmegen

Bronwen G. Evans
University College London

Jane Stuart-Smith
University of Glasgow

Sophie K. Scott
University College London

This study aimed to determine the relative processing cost associated with comprehension of an unfamiliar native accent under adverse listening conditions. Two sentence verification experiments were conducted in which listeners heard sentences at various signal-to-noise ratios. In Experiment 1, these sentences were spoken in a familiar or an unfamiliar native accent or in two familiar native accents. In Experiment 2, they were spoken in a familiar or unfamiliar native accent or in a nonnative accent. The results indicated that the differences between the native accents influenced the speed of language processing under adverse listening conditions and that this processing speed was modulated by the relative familiarity of the listener with the native accent. Furthermore, the results showed that the processing cost associated with the nonnative accent was larger than for the unfamiliar native accent.

Keywords: speech comprehension, native accents, nonnative accents, adverse listening conditions

Listeners frequently encounter speakers with a nonnative accent or an unfamiliar native accent. In these situations, they have to adapt to the phonological–phonetic variation in these accents. For nonnative accents, it is generally assumed that the variation arises from the interaction between the segmental and suprasegmental characteristics of a speaker's first (L1) and second (L2) languages (Best, 1994; Best, McRoberts, & Goodell, 200l; Flege, 1991). For example, at the segmental level, variation can occur when L2-learners produce phonetic contrasts absent in their native language, for instance the /l/-/r/ distinction (Yamada, 1995) or the /l/-/w/ distinction (Best & Strange, 1992) for Japanese learners of American English. At the suprasegmental level, L2 learners have been found to have difficulties producing L2-appropriate word stress (Guion, Harada, & Clark, 2004) and intonation patterns (Grabe, 2004; Trofimovich & Baker, 2006).

Patti Adank, Donders Institute for Brain, Cognition, and Behavior, Centre for Cognitive Neuroimaging, Radboud University Nijmegen, Nijmegen, the Netherlands; Bronwen G. Evans, Department of Speech, Hearing, and Phonetic Sciences, University College London, London, United Kingdom; Jane Stuart-Smith, Department of English Language, University of Glasgow, Glasgow, United Kingdom; Sophie K. Scott, Department of Cognitive, Perceptual, and Brain Sciences and Institute of Cognitive Neuroscience, University College London.

This research was supported by the Economic & Social Research Council (ESRC) under Project No. PTA-026-27-1218. We wish to thank Stuart Rosen for supplying the speech-shaped noise used to create the SNR stimuli and Robert Lawson for recording the Glaswegian participants.

Correspondence concerning this article should be addressed to Patti Adank, Donders Institute for Brain, Cognition, and Behavior, Centre for Cognitive Neuroimaging, Radboud University Nijmegen, Kapittelweg 29, Nijmegen 6525EN, the Netherlands. E-mail: patti.adank@donders.ru.nl

Such phonological–phonetic variation in L2-accented speech influences speech comprehension in native listeners. When listening to L2 speakers, native listeners make more errors and show longer response times (Clarke & Garrett, 2004; Munro & Derwing, 1995a, 1995b; Rogers, Dalby, & Nishi, 2004; Schmid & Yeni-Komshian, 1999; van Wijngaarden, 2001). For example, Clarke and Garrett (2004) used a cross-modal matching task with response time measurement: they presented American English listeners with sentences produced by an American English speaker and a Spanish-English bilingual speaker in two experiments and an American English speaker and a low-proficiency Chinese-English bilingual speaker in a third experiment. A sentence was played, and subsequently a visual probe was shown. Listeners had to indicate whether the visual probe matched the last word of the sentence. Clarke and Garrett found that processing for nonnative accented speech is initially slower than for native speech, but that listeners quickly adapt to the accent, within two to four sentence lengths. After the adaptation period, the processing deficit is reduced.

Native (regional) accents also exhibit phonological–phonetic variation at segmental (e.g., Adank, van Hout, & van de Velde, 2007, for Dutch; Clopper, Pisoni, & de Jong, 2005, for American English; Wells, 1982, for British English) and suprasegmental levels (Nolan & Grabe, 1996). In recent years, several studies have investigated whether this variation influences comprehension in the same way as variation in a nonnative accent (Cutler, Smits, & Cooper, 2005; Floccia, Goslin, Girard, & Konopczynski, 2006; Labov, Karen, & Miller, 1991; Major, Fitzmaurice, Bunta, & Balasubramanian, 2005). Together, the results of these studies indicate that listeners show less efficient speech processing for an unfamiliar native accent. For instance, Floccia et al. (2006) conducted a lexical decision experiment (their Experiment 1) in which

French listeners heard speech in their own variety of French, a familiar native accent, and an unfamiliar native accent. They found slower response times (33 ms) for sentences spoken in the unfamiliar native accent.

However, the processing cost associated with understanding an unfamiliar native accent seems to be difficult to pin down. Floccia et al. did not find a processing delay for the unfamiliar accent across all their experiments (e.g., their Experiment 2). An explanation for Floccia et al.'s results may be that the cost associated with processing phonological–phonetic variation in unfamiliar native accents is small to negligible in quiet conditions. Listeners may benefit from the redundancy in the acoustic signal in quiet and have relatively little difficulty with the small deviations in the realization of speech sounds in an unfamiliar accent. A similar ceiling effect has been found in an experiment comparing the processing speed of synthetic versus natural speech (Pisoni, Nusbaum, & Greene, 1985). Pisoni et al. found an interaction between signal-to-noise ratio (SNR) and the type of speech (i.e., natural or synthetic). They reported a small delay for synthetic speech when both types of speech were presented in quiet but substantially larger delays for synthetic speech in noise. Pisoni et al. hypothesized that greater cognitive effort is required for understanding synthetic speech in noise because there is less redundancy in this type of speech. This may go relatively unnoticed in quiet but may become more pronounced under adverse listening conditions. Thus, the increased processing cost may be due to noise-masking portions of phonetic cues relevant for comprehension.

Comparable effects of adverse listening conditions on speech comprehension have been reported for nonnative accents (e.g., Munro, 1998; Rogers et al., 2004; van Wijngaarden, 2001). Rogers et al. compared the intelligibility of native (English) and mildly accented nonnative (Chinese-accented English) sentences. The sentences were presented in quiet and at three SNRs. Intelligibility was measured as the proportion of correctly identified content words. The results for the quiet condition showed small differences between native and nonnative speech, but results for the noise showed considerably lower scores for the nonnative speech.

In the present study, we investigated whether a similar interaction—between adverse listening conditions and speaker accent such as reported for nonnative and synthetic speech—also occurs for unfamiliar native accents. The aim was to determine the relative processing cost associated with comprehending speech in an unfamiliar native accent compared with that of comprehending a familiar native accent in quiet and at three SNRs: if comprehending speech in an unfamiliar native accent is associated with a greater processing cost under adverse listening conditions, then the magnitude of that cost may be reliably estimated.

The processing cost was measured using a computerized version of the Speech and Capacity of Language Processing Test, or SCOLP (Baddeley, Emslie, & Nimmo-Smith, 1992). SCOLP is a written test originally used as a measure of the slowing down of a patient's cognitive processing following mild head injury (Hinton-Bayre, Geffen, & McFarland, 1997). In SCOLP, the participant verifies as many sentences as possible in 2 min. The sentences are all obviously true or false, and all consist of a mismatch of participant and predicate from true sentences (e.g., "Tomato soup is a liquid" vs. "Tomato soup is people"). As some of these combinations are rather peculiar, the test is sometimes referred to as the "Silly Sentences Test." Overall, it provides a sensitive and reliable measure of the speed of language comprehension, as errors tend to be low across most patient groups. In this set of experiments, we used an aural speeded sentence verification task (May, Alcock, Robinson, & Mwita, 2001) converted from the written version of the SCOLP test (Baddeley, Gardner, & Grantham-McGregor, 1995).

Experiment 1

Experiment 1 investigated the interaction between processing an unfamiliar native accent and adverse listening conditions. The experiment had a between-subjects design, with two listener groups. One listener group was presented with sentences spoken in a familiar and an unfamiliar native accent, while the second group was presumed to be familiar with both accents in which the sentences were spoken.

Two accents of British English were selected. The first accent was Southern Standard British English, here referred to as *Standard English* (SE). SE is a variety of Southern British English that is spoken widely in the Greater London area. The second accent was a variety of Scottish English, namely *Glaswegian English* (GE), which is spoken in Glasgow, Scotland. SE and GE differ considerably at the phonological–segmental level (cf. Stuart-Smith, 2004, for details).

Two groups of listeners were included in the experiment to test the effect of the listener's relative familiarity with GE. The first group included listeners from the Greater London area who spoke SE, referred to as *SE listeners*. The SE listeners were assumed to be familiar with SE and unfamiliar with GE, as a short survey among U.K. phoneticians confirmed the popular belief that GE would be a highly unintelligible native accent for SE listeners. The second group included listeners from Glasgow, referred to as *GE listeners*. The GE listeners were expected to be equally familiar with SE and GE. SE functions as the socioeconomically dominant variety of English across the United Kingdom, is available to middle-class GE speakers through their increased geographical mobility, and is widely used in U.K. national broadcasting media. The assumption that speakers of a regional accent may be as familiar with the standard variety as with their own variety seems appropriate given recent results for General American English (Clopper & Bradlow, in press). Clopper and Bradlow compared the intelligibility of General American with a variety of regional accents of American English and found higher intelligibility for general American.

The stimuli were presented for the two accents in quiet and at three SNRs: $+3$ decibels (dB), 0 dB, and -3 dB, thus creating eight experimental conditions. Ninety-six true/false sentences were presented to the participants, that is, 12 sentences per experimental condition. The sentences were counterbalanced across conditions, so that all 96 sentences were presented in all conditions across all subjects within a listener group. All listeners heard sentences in all four conditions. This was repeated for the SE and GE listener groups, thus ensuring that all sentences were presented in all conditions for both groups. Furthermore, not more than 2 sentences of one speaker were presented in succession, as Clarke and Garrett (2004) found that familiarization occurs after as few as 2 sentences from a speaker.

Method

Participants. The SE listener group consisted of 24 participants (13 men, 11 women; age range, 19–39 years; average age, 27.3 years). They were recruited from the Greater London area and screened for their familiarity with SE and GE. They were all native speakers of English who had lived in southern England all their lives. Overall, they claimed to be unfamiliar with Glaswegian or any other Scottish accents.[1]

The GE listener group consisted of 24 listeners (17 men and 7 women; age range, 19–54 years; average age, 25.4 years). They were native speakers of English and had lived in Glasgow all their lives. Before the experiment, they were asked about their familiarity with SE and GE. All stated that they were familiar with Southern Standard British English and Scottish English, specifically Glaswegian. All were paid for their participation.

Materials. Recordings were made of four SE speakers and four GE speakers. All speakers were male, middle class, and between 20 and 46 years old. Only male speakers were selected because including both genders would have introduced unwanted variation related to the gender differences in larynx size and vocal tract length (cf. Peterson & Barney, 1952).

For every speaker, recordings were made of the 100 sentences of Version A of the SCOLP test (Baddeley et al., 1992). The sentence was presented on the screen of a notebook computer, and speakers were instructed to quietly read the sentence and subsequently to pronounce the sentence as a declarative statement. All sentences were recorded once. However, if the speaker made a mistake, the interviewer went back 2 sentences, and the speaker was instructed to repeat both.

The GE speakers were recorded in Glasgow, and the SE speakers were recorded in London. The GE speakers were recorded in a sound-treated room, using an AKG SE300B microphone (AKG Acoustics, Vienna, Austria), which was attached to an AKG N6-6E preamplifier, on a Tascam DA-P1 DAT recorder (Tascam Div., TEAC Corp., Tokyo, Japan). Each stimulus was transferred directly to hard disk using a Kay Elemetrics DSP sonagraph (Kay Elemetrics, Lincoln Park, NJ). Because our investigative team speaks in Southern English accents, we arranged for the GE recordings to be conducted by a native GE interviewer to avoid the possibility of speech accommodation toward Southern English (Trudgill, 1986). The recordings of the SE speakers were made in an anechoic room, using a Brüel and Kjær 2231 sound level meter (Brüel and Kjær Sound & Vibration Measurement, Nærum, Denmark) as a microphone/amplifier. This microphone was fitted with a 4165 microphone cartridge and its A/C output was fed to the line input of a Sony 60ES DAT recorder (Sony Corp., Tokyo) and the digital output from the DAT recorder fed to the digital input of a Delta 66 sound card (M-Audio UK, Watford, UK) in the Dell Optiplex GX280 personal computer (Dell Corp., Fort Lauderdale, FL). The SE recordings were conducted by a native SE interviewer. The difference in recording conditions between the two speaker groups was not noticeable in the recordings, and it is thus unlikely that intelligibility of the two accents was affected.

Next, all sentences were saved into their own file with beginning and end trimmed at zero crossings (trimming on or as closely as possible to the onset and offset of initial and final speech sounds) and resampled at 22,050 Hz. Subsequently, the speech rate differences across all eight speakers were equalized, so that every sentence had the same length across all eight speakers. This was necessary to ensure straightforward interpretation of the dependent variable (i.e., to be able to express the results in milliseconds). First, for each of the 96 sentences, the average duration across all speakers was calculated. Second, we used the Pitch Synchronous Overlap Add Method, or PSOLA (Moulines & Charpentier, 1990), as implemented in the Praat software package (Boersma & Weenink, 2003), to digitally shorten or lengthen the sentence for each speaker separately. The effect of the shortening or lengthening was in some cases just audible, but it was expected that any effects due to this manipulation were small to negligible, as the manipulations were relatively small and were carried out across all sentences for all speakers in the experiment. Table 1 shows the average percentages of lengthening or shortening performed with PSOLA per accent. Each stimulus was peak-normalized at 99% of its maximum amplitude. Finally, speech-shaped noise was added at the three SNRs.[2] This speech-shaped noise was based on an approximation to the long-term average speech spectrum for combined male and female voices (cf. Byrne et al., 1994, their Table 2). The root-mean-square levels per one third of the octave band were converted into spectrum level and plotted on an octave scale. A three-line approximation was used to capture the major part of the shape from 60 Hz to 9 kHz. This consisted of a low-frequency portion rolling off below 120 Hz at 17.5 dB/octave and a high-frequency portion rolling off at 7.2 dB/octave above 420 Hz, with a constant spectrum portion in between. Per sentence, the noise sound file was cut at a random position from a longer (6-s) segment of speech-shaped noise, so that the noise varied randomly across sentences. The speech-shaped noise had the same duration as the sentence and started and ended with the onset and offset of the sentence. The root mean squares of the sentence and the noise were determined and scaled as to fit the SNR level and finally were combined through addition. Finally, using Praat, we peak-normalized and scaled the intensity of the sound file to 70 dB sound pressure level (SPL).

Procedure. The SE listeners were tested in London, and the GE listeners were tested in Glasgow. All listeners were tested individually in a quiet room while facing the screen of a notebook computer. They received written instructions. The listeners responded using the notebook's keyboard. Half of the participants were instructed to press the *q* key with their left index finger for true responses and to press the *p* key with their right index finger for false responses. The response keys were reversed (i.e., *p* for true and *q* for false) for the other half of the participants. Listeners were not screened for handedness. The stimuli were presented over headphones (Philips SBC HN110; Philips Electronics, Eindhoven, the Netherlands) at a sound level that was kept constant for all

[1] One of the participants had spent some months in Glasgow a decade before. His results were included as they showed the same tendencies as the majority of the participant population.

[2] It was decided to use speech-shaped noise instead of multi-speaker babble, which is often used in intelligibility studies because the accent of the talkers in the babble mixture could influence speech processing (van Heukelem & Bradlow, 2005).

participants.[3] Stimulus presentation and the collection of the responses were performed with Cogent 2000 software (Cogent 2000 team, Wellcome Trust, London, UK), running under Matlab (Mathworks, Cambridge, UK). The response times were measured relative to the end of the audio file, following the computerized SCOLP task in May et al. (2001).

Each trial proceeded as follows. First, the stimulus sentence was presented. Second, the program waited for 3.5 s before playing the next stimulus, allowing the participant to make a response. If the participant did not respond within 3.5 s, the trial was recorded as *no response*. The participants were asked to respond as quickly as they could and told that they did not have to wait until the sentence had finished (allowing for negative response times, as response time was calculated from the offset of the sound file).

Ten familiarization trials were presented prior to the start of the experiment. The familiarization sentences had been produced by a male SE speaker. This speaker was not included in the actual experiment, and neither were the 10 familiarization sentences. The experiment's duration was 15 min, without breaks.

Results

Errors. The error scores were based on the percentage of incorrect responses per participant per SNR condition. The data of 4 participants from the SE listener group were excluded from further analysis, as they did not perform the task correctly.[4,5] The data from three GE participants were excluded, as more than 20% of their responses were slower than 3.5 s. Table 2 shows the average error percentages of the 20 remaining SE participants and the 21 remaining GE participants. Before performing any statistical analyses, we converted the percentages per participant to rationalized arcsine units, or RAUs, (Studebaker, 1985), which is customary for proportional scales (Clarke & Garrett, 2004; Rogers et al., 2004). Transforming the raw proportions to RAU ensures that the mean and variance of the data are relatively uncorrelated and that the data are on a linear and additive scale (cf. Studebaker, 1985).

After transforming the data to RAUs, we performed a three-factor mixed-model analysis of variance (ANOVA) with the transformed error rates as the dependent variable and with accent (SE or GE) and SNR (quiet, +3 dB, 0 dB, and −3 dB SNR) as within-subject factors and with listener group as a between-subjects factor. The results showed a main effect of SNR, $F(1, 117) = 129.86$, $p < .05$, generalized $\eta^2 = .61$ (Bakeman, 2005). A second main effect was found for accent, $F(1, 39) = 41.29$, $p < .05$, generalized $\eta^2 = .34$, indicating that listeners showed different RAUs for the two accents. The SNR × Accent interaction showed a significant effect, $F(3, 117) = 3.26$, $p < .05$, generalized $\eta^2 = .01$; this indicates that RAUs varied per accent depending on the noise level. Finally, an effect was found for the Listener Group × Accent interaction, $F(1, 39) = 17.18$, $p < .05$, generalized $\eta^2 = .28$, which indicates that the RAUs for both listener groups varied depending on the accent. The Listener Group × SNR interaction was not significant, $F(1, 117) = 1.58$, $p = .20$, generalized $\eta^2 < .01$, indicating that the different noise levels affected listeners in both groups in the same way. Finally, the three-way interaction among accent, SNR, and listener group was also not significant, $F(1, 117) = 0.56$, $p = .64$, generalized $\eta^2 < .01$. We performed a series of paired t tests on the RAUs across all SNRs and the two speaker accents for the two listener groups to determine the locus of the interaction between SNR and accent. The results for the SE listeners showed differences between GE and SE sentences for 0-dB and +3-dB SNR at a corrected significance level (Bonferroni correction, $p < .025$). No effects were found for the GE listener group. In sum, the analysis of the errors shows an interaction between accent and SNR: SE listeners made more errors for GE sentences at moderate SNRs.

Table 1
Mean Percentage (and Standard Deviation) of Sentence Lengthening or Shortening in Experiments 1 and 2

Experiment	SE *M*	SE *SD*	GE *M*	GE *SD*	SpE *M*	SpE *SD*
1	0.6	9.8	2.2	13.2	—	—
2	2.7	9.5	4.7	9.2	−4.4	8.6

Note. Experiment 1 involved two accents: Standard English (SE) and Glaswegian English (GE); Experiment 2 involved an additional accent: Spanish-accented English (SpE). Sentences were lengthened or shortened (indicated by minus sign) with Pitch Synchronous Overlap Add Method (PSOLA). Dash = not applicable.

Table 2
Mean Percentage (and Standard Deviations) of Errors for the SE and GE Accents in the SE and GE Listener Groups in Experiment 1

	SE listeners				GE listeners			
	SE speakers		GE speakers		SE speakers		GE speakers	
Noise condition	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Quiet	4	5	8	8	4	4	2	4
+3 dB	13	9	29	11	17	9	21	11
0 dB	26	15	40	13	30	13	31	13
−3 dB	37	20	38	11	36	16	39	15
Total	20	18	29	18	21	17	23	18

Note. SE = Standard English; GE = Glaswegian English; dB = decibels.

[3] For the first participant, the sound settings were adjusted to a comfortable level. All other participants were given the option of changing this setting to a level that was more comfortable for them, but all of them stated that the initially chosen level was comfortable.

[4] When responding to the sentences at the two poorest SNRs, the excluded participants responded as soon as the sentence started. Consequently, they showed response times that were on average shorter for the two poorest SNRs than for the sentences in quiet and at +3 dB, while their performance was at chance. When questioned about this strategy after the experiment, they reported they could not properly understand the sentence in the 0 dB and −3 db SNR conditions and therefore randomly guessed and responded as soon as they heard the noise. The error percentages for these participants for the 0 dB and −3 dB SNR were close to chance level.

[5] It was verified that counterbalancing of the sentences in the two experiments was not affected by the exclusion of the 4 participants in Experiment 1 and the 4 in Experiment 2.

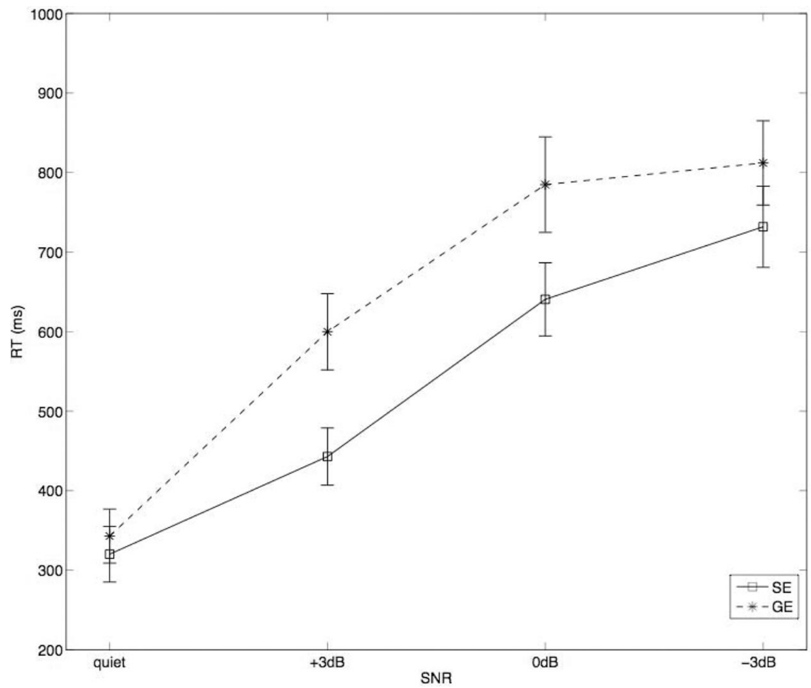

Figure 1. Average response times in milliseconds for the Standard English listener group (Experiment 1) for Standard English (SE; solid line with squares) and Glaswegian English (GE; dashed line with asterisks). Error bars depict 1 standard error.

Response times. Only correct responses were included in the analysis of the response times. All values larger than the average plus 2.5 standard deviations per noise level as calculated across all participants in both groups were considered to be outliers and excluded from analysis. Figure 1 shows the average response times per accent/SNR for the SE listeners, and Figure 2 shows the results for the GE listeners.

As a first step, we performed an ANOVA with the response times as the dependent variable with the same design as for the errors. Main effects were found for SNR, $F(2.24, 85.04) = 137.74$, $p < .05$, Huynh-Feldt-corrected for nonsphericity, generalized $\eta^2 = .58$, and accent, $F(1, 38) = 10.28$, $p < .05$, generalized $\eta^2 = .17$. The SNR × Accent interaction was marginally significant, $F(3, 114) = 2.27$, $p = .08$, generalized $\eta^2 = .01$. A final effect was found for Listener Group × Accent, $F(1, 38) = 21.84$, $p < .05$, generalized $\eta^2 = .29$. No effect was found for Listener Group × SNR, $F(1, 114) = 0.56$, $p = .64$, generalized $\eta^2 < .01$, indicating that the different noise levels similarly affected all listeners. The Listener Group × SNR × Accent interaction was not significant, $F(3, 114) = 1.51$, $p = .22$, generalized $\eta^2 < .01$. Finally, a series of paired t tests was carried out per SNR across both speaker accents and for both listener groups separately. The results confirmed the observation from Figure 1 that the SE listeners were slower (Bonferroni-corrected significance level, $p < .025$) when listening to GE sentences at +3 dB and 0 dB SNR. The difference in response times between the familiar and unfamiliar accents in quiet was not significant. No effect was found for −3 dB SNR, indicating that the difference between the two accents disappears when the SNR deteriorates. The t tests for the GE listeners showed no significant effects.

Discussion

The results of Experiment 1 can be summarized as follows. When sentences were presented in moderate adverse listening conditions (+3 dB and 0 dB SNR), SE listeners were slower to give correct responses to sentences spoken in the unfamiliar native accent, while GE listeners made an equal number of errors and were equally fast for both accents. The processing delay for SE listeners for GE sentences may be explained by the fact that SE listeners have been exposed to—and interacted with—other SE speakers all their lives, while they are largely unfamiliar with GE speakers. This delay may reflect the additional processing SE listeners performed for dealing with differences between SE and GE (Stuart-Smith, 2004). It appears that the SE listeners processed these differences effectively in quiet, as no significant differences in processing speed were found for the two accents. It seems plausible that the longer response times and the higher number of errors for the GE sentences for the SE listeners at moderate SNRs were caused by their relative unfamiliarity with GE and were not attributable to other differences between the data sets (e.g., recording conditions, speaker idiosyncrasies), as the GE listeners (who were equally familiar with both accents) showed no difference in processing speed or number of errors for both accents.

Experiment 2

Experiment 1 showed that sentence processing in an unfamiliar native accent was delayed under adverse listening conditions. It is not clear whether this delay is of the same magnitude as the slowing down associated with processing speech

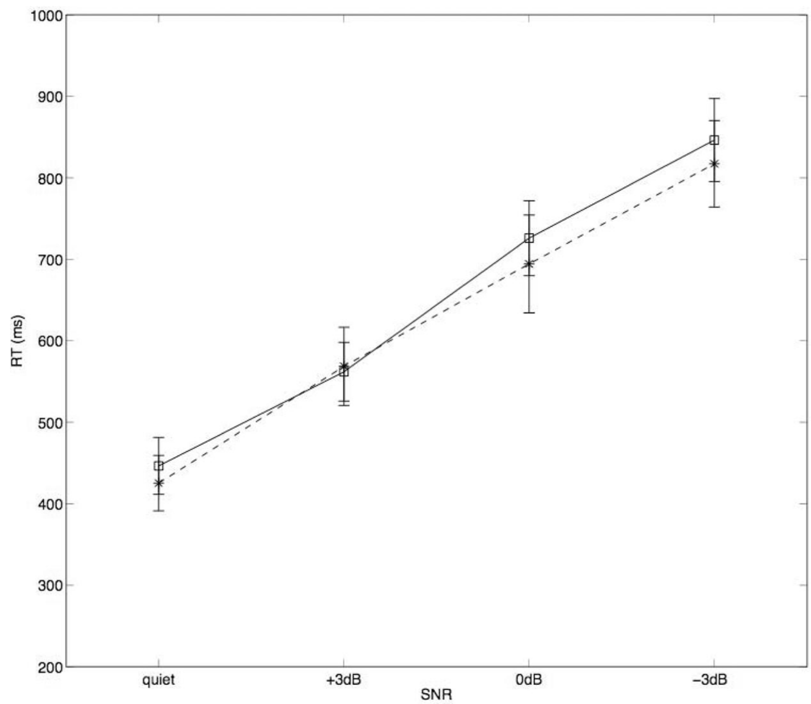

Figure 2. Average response times in milliseconds for the Glaswegian English listener group (Experiment 1) for Standard English (SE; solid line with squares) and Glaswegian English (GE; dashed line with asterisks). Error bars depict 1 standard error.

in a nonnative accent under adverse conditions. In Experiment 2, a group of SE listeners performed the same sentence verification task as in Experiment 1, only here they were presented with speech from three speaker groups: SE, GE, and Spanish-accented English (SpE). The set-up of Experiment 2 was the same as that in Experiment 1 and was conducted in London, with SE listeners.

Method

Participants. Twenty participants (7 men, 13 women; age range, 19–35 years; average age, 26.8 years) took part in the experiment. They were native speakers of English and had lived in the south of England all their lives. All stated that they were not overly familiar with Glaswegian or other Scottish English accents, nor with Spanish-accented English. All were paid for their participation. None of them had participated in Experiment 1.

Materials. Six speakers were included in Experiment 2: two SE speakers and two GE speakers used in Experiment 1, plus two SpE speakers. We recorded two male Spanish-accented English speakers using the same set-up as was used for the SE speakers in Experiment 1. Both speakers were from Latin America and had learned English as a second language at school from age 12. They had been living in the United Kingdom for an average of 3 years. They were judged by the experimenters to speak with a moderately heavy Spanish accent. The recordings were conducted by a native SE interviewer. Next, the average duration across all six speakers was calculated for each of the 96 sentences. The sentence was digitally shortened or lengthened for each speaker separately with the same procedure as in Experiment 1 (cf. Table 1). From there on, the stimuli were processed as in Experiment 1. The procedure was similar to Experiment 1; only in this experiment, 8 sentences were presented per condition instead of 12.

Results

Errors. The error scores were calculated as in Experiment 1 and converted to RAUs (cf. Table 3 for the raw percentage error). Four participants were excluded from the analysis, as they performed the task incorrectly (cf. Experiment 1). First, we performed a two-factor mixed-model ANOVA with RAU-transformed error

Table 3
Mean Percentage (and Standard Deviations) of Errors for the SE, GE, and SpE Accents in Experiment 2

	SE		GE		SpE	
Noise condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Quiet	3	8	2	5	9	7
+3 dB	17	7	18	13	32	16
0 dB	18	14	21	15	42	19
−3 dB	28	19	38	21	47	16
Total	16	15	20	19	32	21

Note. SE = Standard English; GE = Glaswegian English; SpE = Spanish-accented English; dB = decibels.

rates as the dependent variable and with accent (SE, GE, and SpE) and SNR as factors. The results showed main effects of SNR, $F(3, 45) = 77.53$, $p < .05$, generalized $\eta^2 = .66$, and accent, $F(1.85, 27.79) = 29.11$, $p < .05$, Huynh-Feldt-corrected, generalized $\eta^2 = .49$. The SNR × Accent interaction was not significant, $F(6, 90) = 1.54$, $p < .17$, generalized $\eta^2 < .01$). Next, three series of t tests were conducted for all SNRs across all three accents. All levels of significance were (Bonferroni) corrected for multiple comparisons, setting the significance level to 0.017. There were no differences between SE and GE, while all four SNRs differed significantly between SE and SpE, and differences were found between GE and SpE at all SNRs. Listeners thus made more errors for SpE than for GE and SE.

Response times. Figure 3 shows the average response times in milliseconds per accent and per SNR. The results from the ANOVA (only correct responses faster than 2.5 s were included) showed main effects of SNR, $F(2.7, 40.51) = 55.45$, $p < .05$, Huynh-Feldt-corrected, generalized $\eta^2 = .62$, and accent, $F(2, 30) = 25.66$, $p < .05$, generalized $\eta^2 = .49$. The SNR × Accent interaction was also significant, $F(3.11, 46.67) = 3.75$, $p < .05$, Huynh-Feldt-corrected, generalized $\eta^2 = .01$. The paired t tests (corrected significance level of $p < .017$) showed effects between SE and GE at 0 dB SNR (and an effect just not significant at the corrected level for +3 dB, $p = .026$), effects for quiet, +3 dB and 0 dB SNR for SE and SpE, and an effect at −3 dB for GE and SpE.

Discussion

The aim of Experiment 2 was to compare the relative processing cost for comprehending speech in an unfamiliar native accent under adverse listening conditions with comprehending speech in a nonnative accent. The results show, first, that listeners made more errors when verifying sentences produced by SpE speakers, compared with sentences produced by SE and GE speakers. Second, response times for SpE were slower than for SE and GE, and response times were slower for GE than for SE for moderate SNRs. Third, it appears that the delay for verifying SpE sentences compared with SE sentences was larger than the delay for processing GE sentences compared with SE sentences. Across all SNRs, GE sentences were on average processed 88 ms slower than SE sentences, while SpE sentences were processed 114 ms slower than SE sentences and 26 ms slower than GE sentences. Finally, the results indicated that the delay in processing for the unfamiliar native accent increased in noise (albeit only at moderate significance levels), while this was not the case for the delay associated with the nonnative accent.

The results for the SE and GE sentences resemble the results for the SE listeners in Experiment 1: SE listeners in Experiment 1 also showed longer response times for GE sentences than for SE sentences. However, there is one discrepancy: listeners made more errors when verifying GE sentences in Experiment 1 but not in Experiment 2. This may have been caused by the selection of the speakers for Experiment 2. In Experiment 2, only two of the original four GE speakers were used. Perhaps the phonetic–phonological differences between the GE and SE accents for the two selected speakers were less prominent than for the other two GE speakers. It could thus be that there was less accent-related variation present in Experiment 2, which may have improved performance for GE.

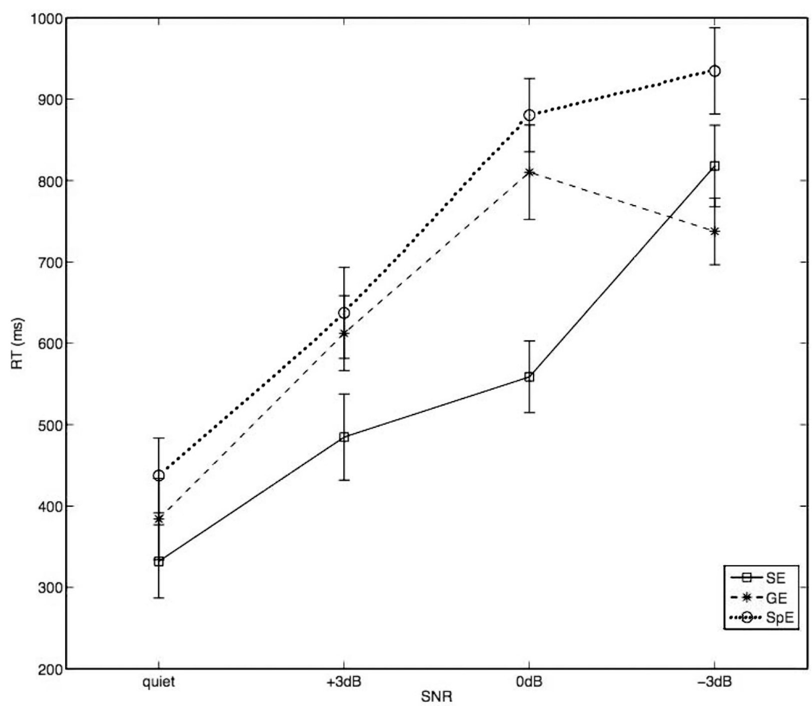

Figure 3. Average response times in milliseconds for the Standard English listener group (Experiment 2) for Standard English (SE; solid line with squares), Glaswegian English (GE; dashed line with asterisks), and Spanish-accented English (SpE; dotted line with circles). Error bars depict 1 standard error.

General Discussion

The purpose of the present study was to determine the relative processing cost of comprehending speech in an unfamiliar native accent under adverse listening conditions. As this processing cost could not always be reliably estimated in quiet listening conditions (e.g., Floccia et al., 2006), we investigated the interaction between adverse listening conditions and sentences in an unfamiliar native accent in two experiments.

In Experiment 1, listeners whose language variety was Standard English (SE) or Glaswegian English (GE) performed a sentence verification task in which they were presented with sentences at various SNRs in SE or GE. The SE listeners were assumed to be familiar with SE and unfamiliar with GE, while the GE listeners were assumed to be equally familiar with both accents. The results for the SE listeners showed that they made more errors and showed slower response times at moderate SNRs for the GE sentences. The results for the GE listener group in Experiment 1 showed that they made an equal number of errors and responded equally fast for both accents. The finding that the performance of the GE listeners was not affected by the accent of the speaker confirms that the processing delay for the GE sentences by the SE listener group was due to the relative unfamiliarity of the SE listeners with the Glaswegian accent. SE listeners thus benefited from their relative familiarity with SE.

Experiment 2 was a comparison of the relative cost of processing speech in an unfamiliar native accent and in a nonnative accent under moderately adverse listening conditions in SE listeners only. The results showed a pattern in the response times that also had been found in Experiment 1 for the SE listeners: listeners processed GE sentences slower than SE sentences at moderately adverse SNRs. Second, the results showed that the processing delay associated with listening to an unfamiliar native accent is less prominent than the delay associated with listening to a nonnative accent, as the delays for SpE compared with SE were larger. Processing of SpE sentences was also slower than GE sentences. These results fit the argumentation in Clarke and Garrett (2004) that the phonological–phonetic variation in nonnative accents represents an extreme form of the variation in native accents (Nygaard & Pisoni, 1998). When listening to a nonnative accent, listeners may thus have to adapt more than when listening to a native accent, which could in return be reflected in a lower processing cost for the native accent.

Accent Processing in Quiet and Under Adverse Listening Conditions

No effects were found for processing the unfamiliar native accent in quiet. This result shows again that the cognitive processing cost cannot easily be estimated in quiet conditions (cf. Floccia et al., 2006). However, in both experiments, an interaction was found between the unfamiliar accent and moderately poor SNRs (+3 dB and 0 dB SNR for Experiment 1, and 0 dB SNR for Experiment 2): listeners slow down considerably for these SNRs for the unfamiliar accent. A similar interaction has been found in experiments comparing the processing speed for synthetic versus natural speech (e.g., Pisoni et al., 1985). In conclusion, it seems justified to assume that processing an unfamiliar native accent in noise is delayed compared with processing a familiar native accent in noise.

Familiarity With a Native Accent and Speech Comprehension

Experiment 1 indicates that familiarity with a native accent benefits speech comprehension, as SE listeners responded slower when listening to GE, an unfamiliar native accent for SE listeners. GE listeners, on the other hand, responded equally fast for GE, their native accent, and for SE, a familiar native accent.

Previous research on accent adaptation has suggested that the ability to adapt to an unfamiliar native accent may require long-term experience of interacting with speakers of that accent. Evans and Iverson (2007) investigated vowel perception and production among university students from the north of England, as they adapted their accent from regional to educated (i.e., SE) norms. Participants were tested in their production and perception at regular intervals over a period of 2 years in a battery of tests. At each testing session, they read a short passage and a set of experimental words. They also completed two perceptual tasks; they found best exemplar locations for words embedded in either northern- or southern-English-accented carrier sentences and identified words in noise spoken in either a northern- or southern-English accent. The results demonstrated that participants changed their spoken accent to sound more southern after attending university, though there were individual differences; some participants changed their accent more than others and some produced more southern vowels overall (i.e., at each testing time). These individual differences in production affected perceptual processing. Specifically, individuals who had a more southern-English accent overall were better at identifying SE-accented speech than those who had a more northern-English accent overall. This was unexpected, as all participants had been born and raised in the same community and had similar experience with SE; although they were all highly familiar with SE through the media, they had little experience of interacting with SE speakers before going to university. On the basis of Evans and Iverson's results, one could hypothesize that familiarity with a native accent does not come from being exposed to it through the media alone but that interaction with speakers of that accent (or even adapting one's own speech to that accent) is also required. However, our results do not provide support for this hypothesis, as GE listeners were equally fast for GE and SE. The GE listeners had been born and raised in Glasgow, and although they were highly familiar with SE through the media, they had had little experience of interacting with SE speakers on a regular basis. One possibility is that the Glaswegian listeners had had enough experience with SE, both through the media and through interacting with SE speakers, to enable them to adapt easily to SE speech. Also, the Glaswegian listeners were recruited through Glasgow University, where SE is frequently encountered. Furthermore, Glasgow is a large city where listeners frequently come into contact with speakers of different regional accents. It is thus possible that through their contact with the university and experience of living in a multidialectal environment, these listeners had gained enough experience of interacting with SE speakers. The present results thus suggest that it is not necessary to interact with speakers of a different native accent on a regular basis in order to be highly familiar with the accent.

Nevertheless, how much or what kind of exposure is required to obtain equally efficient processing for a familiar and an unfamiliar native accent is presently unclear. It would be interesting to establish whether explicit short-term training with an unfamiliar native accent would speed up comprehension. Before training, the results should resemble those of the SE listeners in Experiment 1, but after training, the delay for the unfamiliar accent should disappear, and results should resemble those of the GE listeners in Experiment 1. One study has already shown a similar effect of explicit training for nonnative-accented speech (Bradlow & Bent, 2008), but another study on familiar and unfamiliar native accents of Dutch did not show an effect of short-term exposure on the speed of word processing (Adank & McQueen, 2007).

In conclusion, the present study indicates that familiarity with the speaker's accent benefits listeners under adverse listening conditions. In showing that listening to an unfamiliar native accent influences the speed of language processing in adverse listening conditions, this study contributes to a growing body of research on the perceptual consequences of phonological–phonetic variation related to the speaker's accent.

References

Adank, P., & McQueen, J. M. (2007). The effect of an unfamiliar regional accent on the speed of word processing. In J. Trouvain and W. J. Barry (Eds.), *Proceedings of the XVIth International Congress of Phonetic Sciences, Saarbrücken, Germany,* 1925–1928.

Adank, P., van Hout, R., & van de Velde, H. (2007). An acoustic description of the vowels of northern and southern standard Dutch: II. Regional varieties. *Journal of the Acoustical Society of America, 121,* 1130–1141.

Baddeley, A. D., Emslie, H., & Nimmo–Smith, I. (1992). *The Speed and Capacity of Language Processing (SCOLP) Test.* Bury St. Edmunds, UK: Thames Valley Test Co.

Baddeley, A. D., Gardner, J. M., & Grantham-McGregor, S. (1995). Cross-cultural cognition-developing tests for developing countries. *Applied Cognitive Psychology, 9,* S173–S195.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37,* 379–384.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation hypothesis. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception* (pp. 167–224). Cambridge, MA: MIT Press.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America, 109,* 775–794.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20,* 305–330.

Boersma, P., & Weenink, D. (2003). *Praat: Doing phonetics by computer.* Retrieved August 11, 2008, from http://www.fon.hum.uva.nl/praat

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106,* 707–729.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., et al. (1994). An international comparison of long-term average speech spectra. *Journal of the Acoustical Society of America, 96,* 2108–2120.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented speech. *Journal of the Acoustical Society of America, 116,* 3647–3658.

Clopper, C. G., & Bradlow, A. (in press). Perception of dialect variation in noise: Intelligibility and classification. *Language and Speech.*

Clopper, C. G., Pisoni, D. B., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America, 118,* 1661–1676.

Cutler, A., Smits, R., & Cooper, N. (2005). Vowel perception: Effects of non-native language vs. non-native dialect. *Speech Communication, 47,* 32–42.

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America, 121,* 3814–3816.

Flege, J. E. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. In C. Ferguson and T. Huebner (Eds.), *Crosscurrents in second language acquisition and linguistic theories,* Philadelphia, PA: John Benjamins.

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,* 1276–1293.

Grabe, E. (2004). Intonational variation in urban dialects of English spoken in the British Isles. In P. Gilles & J. Peters (Eds.), *Regional variation in intonation* (pp. 9–31). Tuebingen, Germany: Niemeyer/Linguistische Arbeiten [Linguistic Work].

Guion, S. G., Harada, T., & Clark, J. J. (2004). Early and late Spanish-English bilinguals' acquisition of English word stress patterns. *Bilingualism: Language and Cognition, 7,* 207–226.

Hinton-Bayre, A. D., Geffen, G., & McFarland, K. (1997). Mild head injury and speed of information processing: A prospective study of professional rugby league players. *Journal of Clinical and Experimental Neuropsychology, 19,* 275–289.

Labov, W., Karen, M., & Miller, C. (1991). Near-mergers and the suspension of phonemic contrast. *Language Variation and Change, 3,* 33–74.

Major, R. C., Fitzmaurice, S. M., Bunta, F., & Balasubramanian, C. (2005). Testing the effects of regional, ethnic, and international dialects of English on listening comprehension. *Language Learning, 55,* 37–96.

May, J., Alcock, K. J., Robinson, L., & Mwita, C. (2001). A computerized test of speed of language comprehension unconfounded by literacy. *Applied Cognitive Psychology, 15,* 433–443.

Moulines, E., & Charpentier, F. J. (1990). Pitch-synchronous waveform processing technique for text-to-speech synthesis using diphones. *Speech Communication, 9,* 453–467.

Munro, M. J. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Learning, 20,* 139–154.

Munro, M. J., & Derwing, T. M. (1995a). Foreign accent comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45,* 73–97.

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of foreign-accented speech. *Language and Speech, 38,* 289–306.

Nolan, F., & Grabe, E. (1996). Preparing a voice line-up. *Forensic Linguistics, 3,* 74–94.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics, 60,* 355–376.

Peterson, B., & Barney, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24,* 175–184.

Pisoni, D., Nusbaum, H., & Greene, B. (1985). Perception of synthetic speech generated by rule. *Proceedings of IEEE, 73,* 1665–1676.

Rogers, C. L., Dalby, J., & Nishi, K. (2004). Effects of noise and proficiency level on intelligibility of Chinese-accented English. *Language and Speech, 47,* 139–154.

Schmid, P. M., & Yeni-Komshian, G. H. (1999). The effects of speaker accent and target predictability on perception of mispronunciations. *Journal of Speech, Hearing, and Language Research, 42,* 56–64.

Stuart-Smith, J. (2004). The phonology of Scottish English. In C. Upton (Ed.), *Varieties of English: I. Phonology*. Berlin: Mouton de Gruyter.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech, Hearing, and Language Research, 28,* 455–462.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28,* 1–30.
Trudgill, P. (1986). *Dialects in contact.* Oxford, UK: Basil Blackwell.
van Heukelem, K., & Bradlow, A. R. (2005). Tuning in and tuning out: Speech perception in native- and foreign-talker babble. *Journal of the Acoustical Society of America, 118,* 1900.
van Wijngaarden, S. J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication, 35,* 103–113.
Wells, J. C., (1982). *Accents of English.* [Three volumes and cassette]. Cambridge, UK: Cambridge University Press.
Yamada, R. (1995). Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 305–320). Timonium, MD: York Press.

Received February 22, 2007
Revision received June 3, 2008
Accepted June 26, 2008 ■