The effect of regional and non-native accents on word recognition processes: A comparison of EEG responses in quiet to speech recognition in noise

Louise Stringer, Paul Iverson
Department of Speech, Hearing and Language Sciences,
University College London, London, UK
l.stringer.11@ucl.ac.uk, p.iverson@ucl.ac.uk

10.21437/Interspeech.2014-555

Abstract

Previous work has shown that both the speed and accuracy of word recognition can be reduced if a talker has a regional or non-native accent, particularly under noisy conditions. This study investigated whether the reduced intelligibility of some accents in noise are related to aspects of speech processing in quiet. Our goal was to see if difficulties processing accented speech under adverse listening conditions can be revealed in quiet using electrophysiological methods. Participants heard English sentences in a standard, regional or non-native accent. Behavioural measures found listeners had a strong intelligibility advantage for the standard accent when mixed with noise, replicating previous work. However, differences in processing the accents were less clear using the EEG measures. We found a significant Phonological Mapping Negativity (PMN) elicited by phonological anomalies, as well as an N400 effect, which is related to semantic anomalies, but although both responses were somewhat larger for the standard accent, there was more overlap in the EEG responses for different accents than for speech-in-noise recognition. More work is needed, but it is plausible that some of the difficulty that listeners have with accents under noise may be related to the same accent processing difficulties that occur in quiet.

Index Terms: Speech perception, accented speech, EEG

1. Introduction

Listening to speech in an unfamiliar regional or non-native accent can impair speech perception processes. Listeners may be slower to recognise words [1, 2], and the accuracy of word recognition can also be impaired when listening to non-native [3] and regional accents [4]. However, while an unfamiliar accent can be much less intelligible than a familiar, standard accent in noisy conditions, when listening to speech in quiet, regional and non-native accents may have little or no impact on the accuracy of speech recognition [5]. As word recognition accuracy in quiet is generally at or near ceiling, this means it is not clear whether the processing difficulties associated with unfamiliar regional and non-native accents arise specifically in the presence of background noise, or if accented speech impedes speech processing in general. EEG methods could be useful in this context as they directly measure processes underlying speech perception, instead of relying on a behavioural response to give an indirect measure of processing difficulties. As such, EEG methods could reveal processing difficulties caused by regional and non-native accents in quiet conditions that may be obscured by high performance levels in behavioural measures.

Two EEG responses related to word recognition processes may be particularly useful to investigate the effect of accent on speech perception processes. These are the Phonological Mapping Negativity (PMN) and the N400 effect. The PMN effect is a negative-going response occurring around 270-300ms after the onset of a critical word, and is related to phonological processing stages of word recognition [6]. The PMN can be seen as a 'goodness-of-fit' measure between the expected phonological form of a word, as predicted based on context, and the actual acoustic-phonetic input received, with a better fit giving a smaller response [7]. The N400 effect occurs slightly later, with a negative peak approximately 400ms after critical word onset. The N400 is related to lexical integration, and is elicited by semantic mismatches between expected and heard words. Again, larger mismatches give larger responses [8]. The PMN and N400 effects are generally quantified by calculating the difference in amplitude of responses to unexpected, anomalous words relative to responses to expected, predictable words. If unfamiliar accents cause speech processing difficulties in quiet as well as in noise, this could be reflected in differences in PMN and N400 effects for regional and non-native accents relative to a standard accent in addition to reduced intelligibility of the accents in noise.

Previous work suggests that PMN and N400 effects may be influenced by regional and non-native accents, but findings are inconsistent. Regional accents may lead to increased PMN effects compared to a listener's own accent [9], but this may not occur if the regional accent is familiar to a listener [10], and the N400 seems to be unaffected by regional accents [9]. Non-native accents may elicit smaller PMN and N400 effects than a listener's accent [9], but the N400 effect has also found to be larger for non-native accents [11], or equivalent to the N400 for a standard accent [12]. Although findings have so far been inconsistent, it seems that unfamiliar accents may influence PMN and N400 effects under some circumstances.

This study investigated whether the processing difficulties caused by regional and non-native accented speech in noise also occur in quiet by using EEG methods in addition to a behavioural speech recognition task. Participants with a standard English accent listened to English sentences spoken in their standard accent, a regional accent and a non-native accent. In a behavioural speech recognition task, sentences were presented with background noise, and in an EEG task, PMN and N400 responses were calculated for sentences presented in quiet. We expect the regional and non-native accents to be less intelligible than the standard accent when heard in noise. If regional and non-native accents also cause processing difficulties in quiet, differences in the magnitude of PMN and N400 effects could be expected for these accents compared to those to a familiar, standard accent. If no differences in magnitudes are found, this could suggest that the difficulties seen for accented speech in noise may arise specifically under adverse listening conditions.

2. Materials and Methods

2.1. Participants

Ten native English speakers (5 female, mean age = 25.4 years, SD = 4.40 years, range = 20-31 years) took part in the study. All were from South East England, with a standard Southern British English accent, were right-handed and reported no known hearing, language or learning impairment. Participants received payment for their time.

2.2. Stimuli

2.2.1. Talkers

Materials were recorded by 4 talkers (2 male, 2 female) of a standard accent (Standard Southern British English), a regional accent (Glaswegian English), and a non-native accent (Spanish-accented English). Talkers of the standard and regional accents were native, monolingual English speakers, raised in either the southeast of England or Glasgow, respectively. Non-native talkers were native Spanish speakers from the Basque region of Spain, who spoke English at an upper-intermediate or advanced level (average age of acquisition = 6.25 years, SD = 1.26 years, range = 5−8 years). Recordings were made in a soundproof booth in either London, Glasgow or Spain at a sampling rate of 44100Hz, with 24bits per sample, and then normalised to the same mean intensity.

2.2.2. Sentence Materials

Materials consisted of 432 matched sentence triplets in three conditions, which were designed for this study to allow the use of matched, but not identical materials in the speech recognition and EEG tasks. The three sentence conditions differed in the level of contextual constraint in the sentence and/or the congruity of the final keyword. Predictable sentences had a highly constrained sentence context completed by a congruent final keyword. This keyword is the word most likely to complete the specific context generated in the sentence, so it is highly predictable. Neutral sentences had the same congruent keyword, but the sentence context was weakly constrained. As the context generated is less specific, the final keyword is no longer easy to predict. Anomalous sentences contained the same highly constrained sentence context as the predictable sentences, but were completed by an incongruent keyword unrelated to the context. Strongly and weakly constrained sentence contexts were matched for content word count, total word count and syllable count. Congruent and incongruent final keywords were nouns between one and five syllables long, and each pair were matched on syllable count, lexical stress placement, lexical frequency and phonological neighbourhood density. To ensure all keyword pairs were immediately acoustically distinguishable, there was no initial phonological overlap between the predictable and anomalous words. Complete sentence length was 6-10 words and 6-16 syllables.
In the speech in noise recognition task participants heard all 432 neutral sentences. Sentences were presented in quiet, or mixed with speech-shaped noise at +3dB, 0dB or −3dB signal-to-noise ratios. An equal number of sentences were presented in each noise and accent condition, which was counterbalanced across the experiment. In the EEG task, participants heard 216 predictable and 216 anomalous sentences, presented in quiet. An equal number of sentences were heard in each accent within each sentence condition. The condition and accent each sentence appeared in was counterbalanced across the experiment.

2.3. Procedure

The speech recognition and EEG tasks took place in different sessions between one and four days apart. All participants completed the EEG task first, and both sessions took place in a dimly lit, soundproof booth. In the EEG session, before each sentence participants heard a short beep, where they were instructed to blink. After a silence of 1s, a sentence was presented, followed by a silence of 0.75s and a second longer beep. Participants were asked to try not to blink during this time. To ensure participants attended to the sentences, they were asked to decide if the final word of each sentence matched the context, and pressed a corresponding button (labelled “yes” or “no”) on a keyboard when they heard the second beep. The next sentence was presented after the button was pressed, and short breaks were given after every 50 sentences. In the speech-in-noise recognition task participants heard sentences in quiet or in noise presented at a comfortable listening volume. Participants repeated the words they understood, and the experimenter recorded the number of keywords correctly identified per sentence.

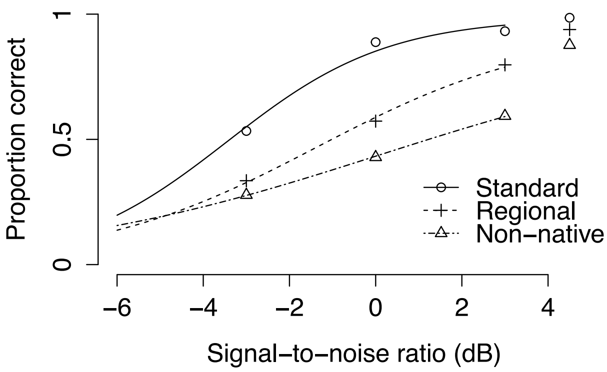

Figure 1: Psychometric function showing speech-in-noise recognition accuracy for each accent. Recognition in quiet is shown to the right.

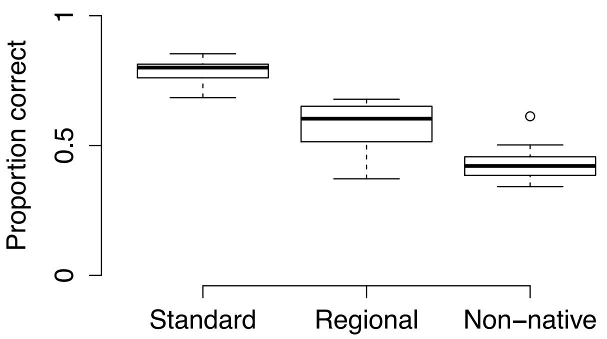

Figure 2: Speech-in-noise recognition accuracy averaged across noise levels

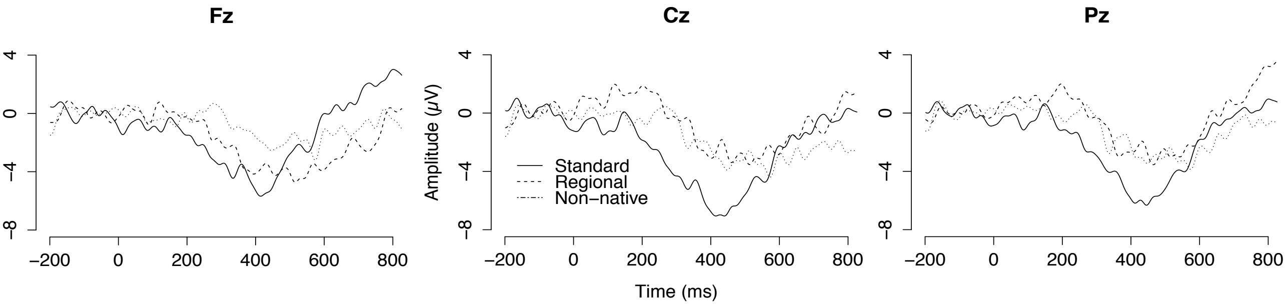

Figure 3: Responses to anomalous words relative to predictable words at three midline electrodes

2.4. EEG Methods

EEG recordings were made from 64 Ag-AgCl active electrodes (BioSemi) arranged according to the 10/20 system, along with EOG recordings made from electrodes placed above and below the left eye and electrodes adjacent to the external canthus of each eye. Data were online referenced to left mastoid and filtered with a low-pass cut-off of 100Hz and a high-pass cut-off of 0.16Hz. Data were re-referenced offline to an electrode placed on the tip of the nose, high-pass filtered with a cut-off of 0.5Hz and then low-pass filtered with a cut-off of 30Hz before being downsampled to 512Hz. Artifacts related to eye-movements in continuous data were corrected for using ICA. Epochs time-locked to the onset of the final word of each sentence were extracted (200ms pre-word onset, 800ms post-word onset), and any epochs still containing artifacts exceeding a threshold of ±150μV were rejected. Remaining epochs were then averaged over each accent and sentence condition.

3. Results

3.1. Speech in noise recognition accuracy

Due to technical problems, one participant was excluded from the speech recognition analysis. The mean proportion of words correctly identified in each accent as a function of noise level is illustrated in Figure 1. This shows that the standard, regional and non-native accents have different levels of intelligibility in noise, but similar intelligibility in quiet. To explore the influence of accent on word recognition further, for each participant the proportion of words correctly identified was averaged across the three noise levels, excluding sentences presented in quiet (Figure 2). This average speech-in-noise accuracy levels was entered into a linear-mixed model with accent as a fixed effect, and participant and experimental script version as random effects. The effect of accent on speech recognition accuracy was highly significant (F(2, 17.97)=94.458, p<0.001), with post-hoc tests showing that word recognition was more accurate for the standard accent (78.45% correct), followed by the regional accent (56.84% correct) and the non-native accent (43.33% correct). Scores for the three accents are all significantly different to teach other (all p<0.001).

3.2. EEG Measures

Data from a second participant was excluded from the EEG analysis, again for technical reasons. Grand average waveforms for the three accents suggest that accent may influence the magnitude of the PMN and N400 effects (Figure 3). To explore this further, the mean amplitude of responses for each accent and sentence condition was calculated over the time windows 200-350ms (corresponding to the PMN) and 350-500ms (corresponding to the N400) for each participant and electrode. In line with the literature, mean amplitudes for accent and sentence conditions were averaged across the midline electrodes (Fz, FCz, Cz, CPz, Pz and POz) for each time window. The magnitude of the PMN and N400 effects for the three accents were calculated by subtracting this midline mean amplitude for the predictable sentences from the mean response to anomalous sentences (Figure 4). The size of the PMN and N400 effects for the different accents were then analysed using a linear mixed-effect model for each time window, with the fixed effect of accent and random effects of participant and experimental script version

3.2.1. PMN (200-350ms)

Accent had a significant effect on the size of the PMN effects observed (F(2, 17.99)=5.543, p=0.013). Post-hoc tests showed the PMN effect for the standard accent was significantly larger than for the non-native accent (t(9)=3.313, p=0.009). The PMN effect for the regional accent fell between that of the other two accents, but was not significantly smaller than the PMN for the standard accent (t(9)=-1.943, p=0.084, n.s.) or larger than the PMN for the non-native accent (t(9)=1.370, p=0.204, n.s).

3.2.2. N400 (350-500ms)

The N400 effect was also influenced by accent (F(2, 17.99)=3.5907, p=0.049), with post-hoc tests showing the N400 effect for the standard accent was significantly larger than for the non-native accent (t(9)=2.655, p=0.026). Again, the N400 for the regional accent was in between these accents, but was not significantly smaller than the N400 for the standard accent (t(9)=-1.644, p=0.135, n.s.) or larger than the N400 for the non-native accent (t(9)=1.011, p=0.339, n.s.).

4. Discussion

As expected, we found the regional and non-native accents to be less intelligible to our listeners than the standard accent when heard in background noise. While the regional and non-native accents were also slightly less intelligible in quiet, the effect of accent on word recognition accuracy was much more pronounced in noise, which reproduces previous

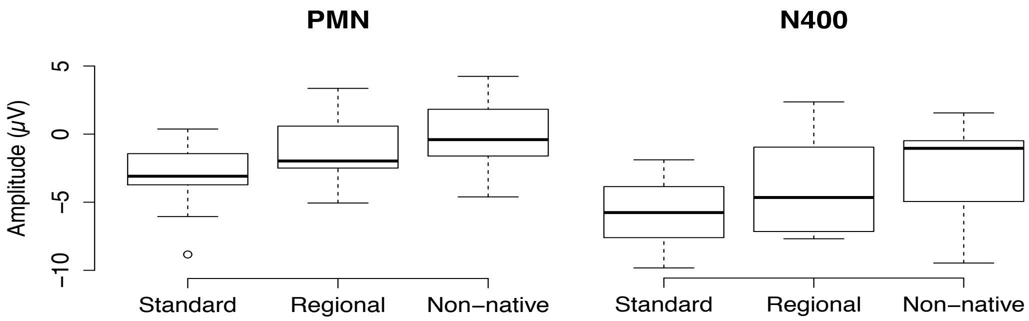

Figure 4: *Average PMN and N400 effect across midline electrodes*

findings [5]. This shows that the regional and non-native accents caused processing difficulties for our listeners in noisy conditions, but the speech recognition task could not reveal whether they also experienced these difficulties in quiet.

The EEG data recorded while listening to the accented speech in quiet found that both the PMN and N400 effects were influenced by accent, although results are not conclusive. The relative magnitudes of the PMN and N400 effects seem to reflect speech-in-noise intelligibility levels, with the largest effects observed for the standard accent, followed by the regional accent and the smallest effects for the non-native effect. However, only the PMN and N400 effects for the standard and non-native accents were significantly different here. This could suggest that non-native accents contain sufficient acoustic-phonetic variation to cause processing difficulties in quiet, but that the variation in regional accents is less disruptive and can be processed similarly to standard accents in quiet. This could suggest that the two accent types may affect speech processing in different ways. However, the small sample size of this study means these results are preliminary and should be interpreted with care. Further testing is required to explore whether regional accents also influence PMN and N400 effects, as suggested here by some findings approaching significance.

If further testing does find reduced PMN and N400 effects for regional as well as non-native accents, both accent types could cause subtle processing difficulties in quiet as well as in noisy conditions. As the PMN and N400 effects are related to the goodness-of-fit of input to expected phonological and semantic forms, processing difficulties may arise if accents influence the expectations made by listeners, whether based on the acoustic-phonetic properties of accented speech, or how listeners perceive accented talkers.

If the acoustic-phonetic variation in regional and non-native accent is likened to a form of signal degradation, these results may be similar to other work that found artificially degraded speech can reduce the N400 effect [13-16]. This reduction seems to be particularly related to attenuated responses for anomalous words [14]. Although it could be expected that listeners rely more on lexical information in adverse conditions, these findings suggest that limited access to phonetic and semantic information (due to signal degradation, or in this case, accent variation) may mean listeners form weaker predictions of upcoming words. This could attenuate PMN and N400 effects as anomalous words are now less incongruent and so elicit smaller responses.

There is also some evidence that listeners modulate their expectations of upcoming words based on talker qualities as well as reduced access to linguistic information. N400 effects have been found for sentences without semantic anomalies, but which would not be expected to be said by a particular talker (e.g.: a child saying “I should quit smoking”), showing that expectations are also based on knowledge of a speaker [17]. In addition to age and gender, a speaker’s accent may also affect the predictions made by listeners. A recent study found that the same syntactic errors elicit a P600 response if spoken in a native accent, but not in a non-native accent. This could suggest that speech in a non-native accent may be expected to be more likely to contain errors than speech in a native accent, so deviations from the syntactically correct forms are less unexpected and do not trigger a response [12]. Phonological anomalies may also be less unexpected for speech in a regional or non-native accent, leading to reduced PMN responses, and semantic anomalies in a non-native accent may also be more expected, causing a smaller N400 effect.

In the context of this experiment, if accented speech elicits smaller PMN and N400 responses compared to a standard accent, this could suggest that listeners form weaker predictions about the phonological and semantic forms of upcoming words for regional and non-native accents. This could be because acoustic-phonetic variation in accents acts as a form of signal degradation, limiting access to linguistic information, or because listeners form weaker predictions based on a talker’s background. This influence of accent on listener expectations could also be linked to the intelligibility of different accents. Predictable words are easier to recognize than words which are less predictable [18], but this benefit of semantic predictions may be less strong if the talker has a regional accent [4]. This suggests that weaker expectations, as related to smaller PMN and N400 effects, may be related to processing difficulties affecting speech in noise intelligibility.

In conclusion, non-native accents elicited reduced PMN and N400 effects compared to a standard accent, but it seems regional accents may also affect these phonological and semantic stages of word recognition. Further testing is required, but it seems that accented speech may cause subtle processing difficulties in quiet that are related to the difficulties seen in the presence of background noise.

5. Acknowledgements

This study was funded with support from the European Commission under Contract FP7-PEOPLE-2011-290000.
Thank you to María Luisa García Lecumberri and Rachel Smith for their kind assistance with organising stimuli recordings.

6. References

[1] P. Adank, B. G. Evans, J. Stuart-Smith, and S. K. Scott, "Comprehension of familiar and unfamiliar native accents under adverse listening conditions.," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 35, no. 2, pp. 520–9, Apr. 2009.

[2] C. Floccia, J. Goslin, F. Girard, and G. Konopczynski, "Does a regional accent perturb speech processing?," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 32, no. 5, pp. 1276–93, Oct. 2006.

[3] T. Bent and A. R. Bradlow, "The interlanguage speech intelligibility benefit," *J. Acoust. Soc. Am.*, vol. 114, no. 3, p. 1600, 2003.

[4] C. G. Clopper, "Effects of dialect variation on the semantic predictability benefit," *Lang. Cogn. Process.*, vol. 27, no. 7–8, pp. 1002–1020, 2012.

[5] M. Pinet, P. Iverson, and M. Huckvale, "Second-language experience and speech-in-noise recognition: effects of talker-listener accent similarity.," *J. Acoust. Soc. Am.*, vol. 130, no. 3, pp. 1653–62, Sep. 2011.

[6] J. Connolly and N. Phillips, "Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences," *J. Cogn. Neurosci.*, vol. 6, no. 3, pp. 256–266, 1994.

[7] R. L. Newman and J. F. Connolly, "Electrophysiological markers of pre-lexical speech processing: evidence for bottom-up and top-down effects on spoken word processing.," *Biol. Psychol.*, vol. 80, no. 1, pp. 114–21, Jan. 2009.

[8] M. Kutas and S. Hillyard, "Brain potentials during reading reflect word expectancy and semantic association," *Nature*, vol. 307, pp. 161 – 163, 1984..

[9] J. Goslin, H. Duffy, and C. Floccia, "An ERP investigation of regional and foreign accent processing.," *Brain Lang.*, vol. 122, no. 2, pp. 92–102, Aug. 2012.

[10] A. Brunellière and S. Soto-Faraco, "The speakers' accent shapes the listeners' phonological predictions during speech perception.," *Brain Lang.*, vol. 125, no. 1, pp. 82–93, Feb. 2013.

[11] C. Romero-Rivas, C. D. Martin, and A. Costa, "On-line adaptation in spoken sentence comprehension: Processing foreign-accented speech," *Cortex.*, In Press

[12] A. Hanulíková and P. van Alphen, "When One Person's Mistake Is Another's Standard Usage: The Effect of Foreign Accent on Syntactic Processing," *J. Cogn. …*, vol. 24, no. 4, pp. 878–887, 2012.

[13] J. Aydelott, F. Dick, and D. L. Mills, "Effects of acoustic distortion and semantic context on event-related potentials to spoken words.," *Psychophysiology*, vol. 43, no. 5, pp. 454–64, Sep. 2006.

[14] V. Boulenger, M. Hoen, C. Jacquier, and F. Meunier, "Interplay between acoustic/phonetic and semantic processes during spoken sentence comprehension: An ERP study," *Brain Lang.*, vol. 116, pp. 51–63, 2011.

[15] J. Obleser and S. a Kotz, "Multiple brain signatures of integration in the comprehension of degraded speech.," *Neuroimage*, vol. 55, no. 2, pp. 713–23, Mar. 2011.

[16] A. Strauß, S. A. Kotz, and J. Obleser, "Narrowed Expectancies under Degraded Speech : Revisiting the N400," *J. Cogn. Neurosci.*, vol. 25, no. 8, pp. 1383–1395, 2013.

[17] J. J. A. Van Berkum, D. van den Brink, C. M. J. Y. Tesink, M. Kos, and P. Hagoort, "The Neural Integration of Speaker and Message," *J. Cogn. Neurosci.*, vol. 20, no. 4, pp. 580–591, 2008.

[18] D. N. Kalikow, K. N. Stevens, and L. L. Elliott, "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *J. Acoust. Soc. Am.*, vol. 61, no. 5, pp. 1337–1351, 1977.