Accent intelligibility differences in noise across native and non-native accents: Effects of talker-listener pairing at acoustic-phonetic and lexical levels

Louise Stringer, Paul Iverson
University College London, London, UK

Correspondence address:
Louise Stringer (now at the University of York)
University of York, York, YO10 5GR

+44 1904 321592
louise.stringer@york.ac.uk

This is a green open-access version of this manuscript, published July 2019 in Journal of Speech, Language, and Hearing Research, 62(7), 2213-2226.
https://doi.org/10.1044/2019_JSLHR-S-17-0414

Abstract

Purpose: The intelligibility of an accent strongly depends on the specific talker-listener pairing. To explore the mechanism of this influence, we investigated the relationship between acoustic-phonetic similarity and accent intelligibility across native (L1) and non-native (L2) talker-listener pairings. We also used online measures to observe processing differences in quiet. ***Method:*** English (N=16) and Spanish (N=16) listeners heard Standard Southern British English, Glaswegian English and Spanish-accented English in a speech recognition task (in quiet and noise) and an EEG task (quiet only) designed to elicit the Phonological Mismatch Negativity (PMN) and N400 effect. Stimuli were drawn from the Non-native Speech Recognition sentences (Stringer & Iverson, in preparation). The acoustic-phonetic similarity between listeners' accents and the three accents was calculated using the ACCDIST metric (Huckvale, 2004, 2007). ***Results:*** Talker-listener pairing had a clear influence on accent intelligibility. This was linked to the phonetic similarity of the talkers and listeners, but similarity could not account for all findings. The influence of talker-listener pairing on processing in quiet was less clear, but it seems that a close match between talker-listener accents is required to elicit a PMN response. The N400 is more robust to accent mis-matches, with some relationship to intelligibility. ***Conclusion:*** These findings suggest that the influence of talker-listener pairing on intelligibility may be partly attributable to accent similarity in addition to accent familiarity. Online measures also show that differences in talker-listener accents can disrupt processing in quiet even where accents are highly intelligible.

The intelligibility of an accent depends heavily on the particular combination of talker and listener. Listeners with a standard native accent generally find their own accent more intelligible than a regional accent (e.g., Adank, Evans, Stuart-Smith & Scott, 2009; Clopper & Bradlow, 2008; Pinet, Iverson & Huckvale, 2011; Smith, Holmes-Elliott, Pettinato & Knight, 2014) or a non-native accent (e.g., Adank et al., 2009; Bent & Bradlow, 2003; Floccia, Butler, Goslin & Ellis, 2009; Pinet et al., 2011). However, a regional accent is usually as intelligible as a standard native accent for regional-accented listeners (e.g., Adank et al., 2009; Floccia, Goslin, Girard & Konopczynski, 2006; Sumner & Samuel, 2009), and lower-proficiency non-native listeners can find a non-native accent to be more intelligible than a native accent (Pinet & Iverson, 2010; Pinet et al., 2011). While this influence of talker-listener pairing on accent intelligibility is clear, the mechanism of this is not fully understood.

Accent intelligibility is often attributed to the listener's experience with a talker's accent; familiar accents tend to be more intelligible than less familiar ones. This is particularly exemplified by the asymmetry in intelligibility across standard-regional accent pairings. A standard-accented listener finds their own accent to be more intelligible than a regional accent, but they are equally intelligible for regional-accented listeners. For example, word identification is quicker (Adank et al., 2009) and more accurate (Smith et al., 2014) in Standard Southern British English (SSBE) than a Glaswegian accent for SSBE listeners, but no difference is seen between accents for Glaswegian listeners. General American listeners recognise words in a New York accent more slowly than in their own accent, but word recognition is equally fast in both accents for New Yorkers (Sumner & Samuel, 2009). Regional French listeners also recognise words in their accent and standard Parisian French equally quickly (Floccia et al., 2006). This is proposed to occur because standard-accented listeners may have little experience of a regional accent, but regional-accented listeners are likely to be familiar with the standard accent through widespread media exposure or personal interactions (Adank et al., 2009; Sumner & Samuel, 2009). Familiarity with non-native accents can also aid intelligibility, as Dutch listeners who regularly interact with German-accented Dutch talkers recognised words in this accent more quickly than inexperienced listeners. The inexperienced listeners also initially only showed priming effects for weakly and medium accented words, whereas experienced listeners were also primed by strongly accented words (Witteman, Weber & McQueen, 2013). For non-native listeners, a standard native language becomes more intelligible as proficiency in the target language develops, possible because non-native listeners' exposure to a native accent increases with proficiency. When talker and listener have the same L1, proficient listeners behave like native listeners, in that a standard native accent is more intelligible than their non-native accent. This has been found for standard English accents (SSBE or General American) and the listeners' accent for experienced French (Pinet et al., 2011), Spanish (Imai, Walley & Flege, 2005), Dutch (van Wijngaarden, Steeneken & Houtgast, 2002) and Chinese listeners (Hayes-Harb et al., 2008). Intermediate non-native listeners (with L1s listed above and also Korean) often find the standard accent and their own accent equally intelligible (Bent & Bradlow, 2003, Imai et al., 2005, Pinet & Iverson, 2010, van Wijngaarden, 2001, Xie & Fowler, 2013), and low proficiency non-native listeners tend to find their own accent to be most intelligible (Pinet & Iverson, 2010; Pinet et al., 2011; Van Wijngaarden et al., 2002; Xie & Fowler, 2013).

While these findings suggest that familiarity is a major determinant of accent intelligibility, recent evidence shows that the acoustic-phonetic similarity between talker-listener accents is also influential. Pinet et al. (2011) measured accent similarity in terms of vowel spectral properties and vowel duration, finding a strong positive relationship between similarity and intelligibility across numerous native and non-native talker-listener pairings. For example, SSBE listeners found their accent more intelligible than a Northern Irish accent, which was more intelligible than French- and Korean-accented English; this pattern was also found in terms of accent similarity. As it could be expected that listeners' familiarity with the accents also follows this pattern, these findings could suggest that effects previously attributed to familiarity may actually have stemmed from differences in accent similarity. These findings could be extrapolated more generally to standard-accented native listeners; regional accents differ systematically in phonetic realisation of phonemes from a standard accent (Wells, 1982, pp. 72-86), and non-native accents contain both systematic variations due to L1 influence (Walter, 2001) and unsystematic variations relating to proficiency (e.g., Flege, Bohn & Jang, 1997), so accent similarity may account for many of the findings described above. Non-native listeners' finding a native accent more intelligible as proficiency develops could also be accounted for by similarity, as their accents also become more acoustically-phonetically similar to a native accent as proficiency improves (Burgos et al., 2014, Pinet et al., 2011).

Accent similarity can also explain some patterns of intelligibility that cannot be accounted for by familiarity. Pinet et al. (2011) found SSBE and Korean-accented English to be equally intelligible for low-proficiency French listeners. Familiarity suggests that non-native listeners would find SSBE (which they are exposed to through teaching materials and the media) to be more intelligible than unfamiliar Korean-accented English. However, the French listeners' accent was equally distant from the two accents in terms of vowel spectral properties and duration, which may explain why they are equally intelligible for these listeners (Pinet et al., 2011). Studies focusing on specific vowel pairs have also found that spectrally similar vowels tend to be more intelligible than more distant vowels (Oder, Clopper & Ferguson, 2013; Witteman et al., 2013; Wright & Souza, 2013). Together, these findings show that the interaction of acoustic-phonetic features of the listener-talker accents may contribute to intelligibility in addition to familiarity.

For non-native talkers and listeners with different L1s, accent similarity also seems to be important. If the L1s are fairly similar, listeners may find a talker's accent to be as intelligible as their own; Chinese and Korean-accented English is equally intelligible for Chinese and Korean listeners (Bent & Bradlow, 2003). If talker's L1 is more different to the listener's, this accent may be less intelligible than their own, particularly if the talker and/or listener has low proficiency in the target language. This has been found for French and Korean-accented English for experienced (but not proficient) French listeners (Pinet et al., 2011), Dutch and Japanese-accented English for Dutch listeners (Weber, Broersma & Aoyagi, 2011) and Korean and Arabic-accented English for both groups of listeners (Stibbard & Lee, 2006). These findings likely arise from interaction of the L1s with the target language. For similar L1s (e.g., Korean and Chinese), phonetic features of the talker's L1 interact with the target language in a similar way to the listener's L1, producing similar accent features (Bent & Bradlow, 2003). Conversely, for quite different L1s (e.g., Korean and Arabic), the phonetic features of the talker's accent can be very different to those of the listener's (Stibbard & Lee, 2006). This again suggests that accent similarity may influence intelligibility.

Although accent similarity may be able to account for a lot of differences in intelligibility, it has limitations. In particular, similarity cannot account for asymmetry in intelligibility for standard-regional accent pairings, as the two accents should be equally distant across the pairing. Similarity would then suggest that talkers and listeners find each other's accents equally intelligible, but familiarity gives regional listeners the advantage over standard listeners. Short-term exposure can also allow some adaptation to unfamiliar accents. Witteman et al. (2013) found that hearing a story in German-accented Dutch was sufficient for the inexperienced listeners to show a priming effect for strongly accented words, although they continued to respond more slowly than experienced listeners. Native listeners' recognition of a non-native accent can also improve during the course of testing, in terms of speed (Clarke & Garrett, 2004) and accuracy (Bradlow & Bent, 2008), and listeners show rapid perceptual learning of ambiguous phoneme realisations, such as tokens between /s/ and /ʃ/ (Kraljic & Samuel, 2005) or between /f/ and /s/ (McQueen, Cutler & Norris, 2006). While not an accent in itself, accommodating systematic phoneme variations like this could be analogous to adapting to the variation in accented speech. As the similarity between talker-listener accents will not change during such short time frames, these rapid adaptations also suggest that listeners are able to use knowledge of an accent to boost intelligibility.

The influence of talker-listener pairing on accent intelligibility is well-established, but differences in intelligibility are generally elicited in noise, and may not be observed if speech is presented in quiet. That is, a range of accents can be highly intelligible in quiet, but intelligibility differences emerge as phonetic information is removed in a degraded signal. For example, SSBE and Glaswegian accents are equally intelligible in quiet for SSBE-listeners (Adank et al., 2009), and Pinet et al.'s (2011) findings suggest that SSBE and proficient French listeners find SSBE, Northern-Irish, French and Korean accents all to be highly intelligible in quiet, even though they show distinct tuning to the SSBE accent in noise. This could suggest that the influence of talker-listener pairing arises specifically as an interaction with noise, but it seems more likely that differences in accent processing do occur in quiet but are difficult to observe behaviourally using outcome-based measures (e.g., word recognition accuracy scores, response times) which may not be sensitive enough to identify subtle disruptions in processing that are not severe enough to prevent successful word recognition. To explore the influence of talker-listener pairing in quiet, online electrophysiological measures may be useful. There are two event-related potential components (ERPs) of interest; the Phonological Mapping Negativity (PMN) and the N400 effect, which occur around 300ms and 400ms after critical word onset, respectively. In the context of lexical competition, Connolly & Phillips (1994) propose that the PMN relates to phonological mapping of input to the stored representations of activated candidates during lexical selection, and that the N400 reflects semantic integration of the recognised word with the context. The ERPs are elicited by input which violates either phonological (PMN) or semantic (N400) expectations about upcoming words, with greater violations of expectations leading to more negative responses (e.g., Connolly & Phillips, 1994; Diaz & Swaab, 2007; Kutas & Hillyard, 1984; Newman & Connolly, 2009).

Recent studies suggest that talker-listener pairing does influence word recognition processing at varying levels in quiet. Brunellière and Soto-Faraco (2013) found that Catalan speakers showed PMN and N400 effects in response to their own accent and a regional accent. The PMN seemed to be smaller for the regional accent, and the N400 seems equivalent for both accents, but relative response sizes were not specifically analysed. They also found that when a sentence's final word was semantically expected but appeared in the 'other' accent, switching from the listeners' accent to the regional accent elicited a PMN response, but the reverse did not. They suggest this shows that phonological representations activated during lexical selection are less specified when processing a regional accent, but that semantic expectations are developed in both home and regional accents. Goslin, Duffy and Floccia (2012) presented listeners with neutral sentences in their own, regional and non-native accents. For the PMN, the biggest response was for a regional accent, followed by their own accent and smallest for a non-native accent. For the N400, responses were equivalent for regional and their own accents, and again smallest for a non-native accent. The authors suggest this shows that regional accents disrupt phonological mapping processes, while non-native accents cause more long-lived disruption that must be compensated for at a later lexical integration stage of word recognition. There seems to be some similarity in the findings of these two studies for regional accent processing, but ERP findings for non-native accent processing are less consistent; in response to semantic violations, equivalent N400 effects were elicited for Turkish-accented and native Dutch (Hanulíková, van Alphen, van Gogh & Weber, 2012), and a larger N400 effect was elicited in response to a mixed group of non-native Spanish accents than a native Spanish accent (Romero-Rivas, Martin & Costa, 2015).

Differences in PMN and N400 effects found in these studies could arise from a number of factors. Goslin et al. (2012) and Romero-Rivas et al. (2015) analysed ERPs in response to mixed groups of accents; as the specific talker-listener pairing is highly influential on accent processing, mixing pairs could obscure differences in processing. Methodology also differed across studies, which makes direct comparison difficult; Goslin et al. (2012) presented only neutral sentences, which may not create strong enough violations of expectations to reliably elicit the PMN and N400, Brunellière and Soto-Faraco (2013) did not specifically compare size of responses across the talker-listener pairing, and Hanulíková et al.'s N400 investigation (2012) was only a small part of a study into syntactic processing. Finally, accent intelligibility could also be a source of variation, as studies with artificially degraded speech show that more intelligible speech elicits larger N400 effects (Aydelott, Dick & Mills, 2006; Boulenger, Hoen, Jacquier & Meunier, 2011; Obleser & Kotz, 2011; Strauß, Kotz & Obleser, 2013). Less intelligible speech limits access to semantic information, leading to weaker semantic expectations about upcoming words and in turn smaller N400 effects (Aydelott et al., 2006). Phonological expectations may be affected in a similar way, as a PMN effect has been found for clear, but not degraded speech (Strauß et al., 2013). None of the studies above formally compared ERP magnitude with accent intelligibility, but Goslin et al. (2012) did report intelligibility ratings. The same pattern was found as for the N400 effect; the listeners' own accent and the regional accents were rated as being equally intelligible, with non-native accents being less intelligible (Goslin et al., 2012).

The current study further investigated the mechanism of influence of the talker-listener pairing comparing intelligibility in noise to talker-listener phonetic similarity, and to online measures of accent processing at phonological and lexical levels in quiet. The study included two listener groups (SSBE and Spanish-accented listeners) and three talker groups (SSBE, Glaswegian English and Spanish-accented English). A word recognition task was used to establish accent intelligibility for each talker-listener pair, and the acoustic-phonetic similarity of accent pairs in terms of vowel spectral properties and vowel duration were calculated. Based on previous findings (Pinet et al., 2011), we expected to find a positive relationship between accent similarity and intelligibility. To explore processing in quiet, listeners were also presented with sentences in the same three accents in an ERP task designed to elicit the PMN and N400 effect, allowing us to investigate the influence of talker-listener pairing on different processing stages. Based on findings relating to degraded speech (e.g., Aydelott et al., 2006) and comparisons with intelligibility ratings (Goslin et al. 2012), we expected larger PMN and N400 responses for more intelligible accents.

Methods

Participants

16 native, monolingual English listeners, and 16 native Spanish listeners completed the study. Participants were right handed and reported no known hearing, language or learning impairments. English participants (mean age 25.25 years, s.d. 4.20 years) had an SSBE accent and grew up in southern England. Due to technical problems, one English participant's data were excluded from analysis. Spanish participants (mean age 19.38 years, s.d. 2.02 years) were raised in northeast Spain, and none had ever lived in an English-speaking country. All were students in an English Studies degree at the University of the Basque Country, spoke English at an upper-intermediate or advanced level (as rated by their university course) and began learning English between the ages of 5 and 7. They spoke some Basque, but all are dominant Spanish speakers. Testing took place at University College

London (English listeners) and the University of the Basque Country (Spanish listeners).

Stimuli and Apparatus

Four talkers (2 male, 2 female) of SSBE, Glaswegian English (GE), and Spanish-accented English (SpE) recorded the complete set of Non-native Speech Recognition (NNSR) sentences (Stringer & Iverson, in preparation). SSBE and SpE talkers had the same language background described above for English and Spanish listeners, respectively. Recordings were made in a recording booth in London (SSBE talkers and one GE talker), Glasgow (remaining GE talkers) or northeastern Spain (SpE talkers) at a sampling rate of 44100 Hz and 24 bits per sample. Speech-shaped noise was generated for each talker based on the smoothed long-term spectrum of their speech recordings. Sentences were embedded in this noise at +3dB, 0dB and −3dB signal-to-noise ratios (SNRs).

The NNSR sentences comprise 432 triplets of sentences in three semantic conditions; semantically predictable (strongly constrained sentence + congruous final keyword), semantically neutral (weakly constrained sentence + congruous final keyword) and semantically anomalous (strongly constrained sentence + incongruous final keyword). An example triplet is shown in Table 1. Sentence context was established by two or three ‘pointer’ keywords (with the final keyword, this gives three or four keywords per sentence). Pointer keywords are the words which provide context, with remaining words carrying little semantic information. Weakly constrained sentences were created by replacing pointer keywords in strongly constrained sentences with more general words. Sentence types are matched on mean syllable, word and pointer keyword counts. Final keywords are restricted to nouns. Congruous/incongruous pairs are immediately phonetically distinguishable and matched on syllable count, stress pattern, lexical frequency, phonological neighbourhood density and phonological Levenshtein distance. Semantically neutral sentences were presented in the speech-in-noise recognition task, and semantically predictable and semantically anomalous sentences in the ERP task.

EEG data recordings were made from 64 Ag-AgCl active electrodes (BioSemi) arranged according to the 10/20 system, along with electrodes placed above and below the left eye and electrodes adjacent to the external canthus of each eye. Data were collected at a sampling rate of 2048Hz, online referenced to the left mastoid and filtered with a low-pass cut-off of 100Hz and a high-pass cut-off of 0.16Hz. Stimuli were presented through Etymotic ER-1 insert earphones.

Procedure

Speech-in-noise recognition task. 432 semantically neutral NNSR sentences were presented binaurally over headphones at a comfortable listening volume. Listeners verbally repeated words they understood, and the experimenter recorded the number of keywords correctly identified per sentence. Sentences were presented in quiet and three noise levels, with each condition containing an equal number of sentences from each talker. Sentences appeared in a random order and were not repeated. Across the experiment, sentence distribution was counterbalanced between participants so each sentence appeared in all accents and noise conditions. Short breaks were given throughout the task, which took around one hour in total.

Accent similarity analysis. After the speech-in-noise recognition task, listeners recorded 48 semantically predictable NNSR sentences using the above procedure. The acoustic-phonetic similarity of accents for each talker-listener pair was calculated in terms of vowel spectral properties and vowel duration. Vowel similarity was focused on as it is practical to measure, and can be reliably used to identify English accents (Huckvale, 2004). Vowels (excluding schwa) were identified with automatic phonetic alignment of a phonological transcription to the recording using the HTK Hidden Markov Modelling Toolkit (1989). Alignments were hand-checked for errors.

Vowel spectral similarity was calculates using the ACCDIST metric (Huckvale, 2004, 2007). For each talker and listener, Mel-frequency cepstral coefficients (MFCCs) were calculated across the first and second half of each vowel. MFCCs of vowels appearing in repeated tokens of words were averaged across the tokens, but vowels occurring in different contexts (e.g., 'large'/'stars') were not averaged. An intra-speaker vowel spectral distance table was computed based on the Euclidian distance between the MFCC vectors of each pair of vowels. This use of relative, rather than absolute distances between vowels normalises speaker-specific differences in production (Huckvale, 2007). Accent similarity for each talker-listener pair was given by the correlation between their vowel spectral distance tables. For each listener, accent similarity was averaged across the four talkers of each accent to give a representative measure of the similarity of their accent to each accent group as a whole.

Vowel duration similarity was also calculated. For each talker and listener, a list of the duration of 341 individual vowel tokens (excluding schwa) was compiled, and the correlation between lists calculated for each talker-listener pair to give a measure of accent similarity. Accent similarity for each listener was averaged across talkers in an accent group in the same way as described above.

The average similarity of accents within talkers of each accent and the two listener groups were also calculated by finding the spectral and durational correlation between talker-talker or listener-listener pair within a group as described above, and then averaging the correlation across the group.

ERP task. The ERP task took place one to four days before the speech-in-noise recognition task. Participants heard either the predictable or anomalous sentence (216 of each) from the 432 sentence triplets of the NNSR sentences, with an equal number of sentences from each talker in each condition. For each participant, sentences were presented in a random order and were not repeated. The distribution of semantic and accent conditions was counterbalanced so that each possible combination occurred across the experiment. All sentences were presented in quiet.

Participants first heard a short beep, where they were instructed to blink. After 1s silence, a sentence was played, followed by 0.75s silence and another longer beep. Participants were asked not to blink again until they heard this second beep. To ensure participants attended to the sentences, they decided if the final word was congruous and pressed a button labelled “yes” or “no” on a keyboard held on their lap at the second beep. The next sentence was presented after this response. A short initial training task was given, with 4 sentences in each accent (these sentences were not repeated in the main task). Short breaks were given after every 50 sentences, and the task took around 45 minutes.

EEG analysis. Unless otherwise specified, data were processed using SPM8 (Litvak et al., 2011). Data were re-referenced offline to an electrode on the tip of the nose, high-pass filtered with a cut-off of 0.5Hz, low-pass filtered with a cut-off of 30Hz and down-sampled to 512Hz. Eye-movement artifacts were corrected using independent component analysis (EEGLAB, Delorme & Makeig 2004). ERPs were extracted in 1000ms epochs time-locked to final keyword onset, including a 200ms pre-stimulus baseline. Trials containing artifacts exceeding a threshold of ±150μV were rejected (an average of 10.87 trials/English listener and 11.31 trials/Spanish listener). Remaining trials were averaged over each combination of accent and sentence condition for each participant. Grand means for the two groups were calculated for each accent and sentence condition combination by averaging responses over participants at each sample during the 1000ms epoch. Finally, difference waveforms for each listener group and accent were calculated by subtracting responses to predictable sentences from those to anomalous sentences.

For each participant, the PMN and N400 effect were calculated using responses measured at the Cz electrode. Cz was selected as the PMN has a frontal-central distribution (e.g.: Newman, Connolly, Service & McIvor, 2003) and the N400 is usually concentrated over centro-parietal sites (see Kutas & Federmeier, 2011 for review).The PMN for each accent was found by averaging response amplitude for each semantic condition across the 200-350ms time window (post final-word onset), and subtracting mean amplitude for semantically predictable sentences from that of anomalous sentences. The N400 effect was calculated based on each participant's average N400 latency across all accents, because this ERP often has a longer latency for non-native listeners than native listeners (e.g.: Hahne, 2001; Weber-Fox & Neville, 1996). Latency was found by subtracting responses to predictable sentences from those to anomalous sentences at each sample within the 350-500ms time window (post final-word onset), then averaging across accents to give a mean N400 effect. The time point of the most negative-going amplitude within the window was used as that participant's N400 latency. The N400 effect for each accent was calculated by averaging response amplitude for each semantic condition at this latency, and then subtracting mean amplitude for semantically predictable sentences from that of anomalous sentences.

Results

Accent Intelligibility

Figure 1 displays word recognition accuracy in each accent as a function of noise level, showing a clear influence of talker-listener pairing on accent intelligibility. Average accent intelligibility (in noise only) is given in Table 2. To investigate further, word recognition accuracy was entered into a logistic mixed-effects model as a binomial variable showing recognition accuracy (containing the number of correctly identified keywords and the total number of keywords for each sentence, this format was used for the accuracy variable in all analyses). Fixed effects of listener group, accent and noise level (including interaction terms) were included, along with by-listener, by-speaker and by-item random intercepts. The model was fit using the glmer function in the lme4 package (Bates, Maechler, Bolker & Walker, 2014) in the R environment (R Core Team, 2017; all logistic mixed-effects models described were fit using this algorithm, and all analysis was conducted using R).

Significant main effects were found of listener group (χ^2=82.70, df=1, p<0.001), accent (χ^2=16.99, df=2, p<0.001) and noise level (χ^2=2135.72, df=3, p<0.001). This shows English listeners were more accurate than Spanish listeners, the three accents were not equally intelligible, and increased noise levels led to lower intelligibility. All interactions were also significant; listener group*accent (χ^2=384.36, df=2, p<0.001), listener group*noise level (χ^2=118.93, df=3, p<0.001), accent*noise level (χ^2=96.205, df=6, p<0.001), and the three-way interaction term (χ^2=52.08, df=6, p<0.001). The key variables of listener group and accent were investigated with pairwise contrasts (Bonferroni-corrected) of accent intelligibility (for each group separately). This was done using the mcposthoc function in the R package 'LMERConvenienceFunctions' (Tremblay & Ransjin, 2015, also used for other posthoc tests unless stated). These showed that across all noise levels, for English listeners, SSBE>GE>SpE (all p<0.05) and for Spanish listeners, SSBE=SpE>GE (p<0.01).
To explore the three-way interaction, Tukey post-hoc comparisons were run using the glht function in the multcomp package (Hothorn, Bretz & Westfall, 2008). This required constructing another model where listener group, accent and noise level factors were re-coded as a single fixed effect variable with a discrete level for each possible combination of the original variable levels. Different patterns of intelligibility are seen as noise level increases. In quiet, speech recognition accuracy is equivalent for most talker-listener pairings, although GE for Spanish listeners is less intelligible than most other pairings. In noise, English listeners show the intelligibility pattern SSBE>GE>SpE; this is beginning to emerge at +3dB, and is very clear at 0dB and +3dB. Spanish listeners consistently show the pattern SSBE=SpE>GE in noise. As noise increases, accent intelligibility for each talker-listener pair

is significantly lower at each noise level (all $p<0.01$), with two exceptions: high noise levels (-3dB) are required to disrupt the intelligibility of SSBE for English listeners compared to quiet conditions, and the intelligibility of GE for these listeners is also robust in light noise (+3dB) compared to quiet.

These findings show that English listeners are highly tuned into their own SSBE accent; in adverse conditions, it remains highly intelligible until a high level of background noise is present. They are also somewhat tuned to the GE accent, as the intelligibility of this accent was not affected by light noise. Spanish listeners are equally tuned to the SSBE and SpE accents, but struggled to recognise the GE accent even in quiet.

Accent Acoustic-Phonetic Similarity

The average acoustic-phonetic correlation within talker accent and listener groups is given in Table 3. All groups are more similar in terms of vowel duration than vowel spectral properties, likely because vowel duration can vary in one dimension, whereas vowel spectral properties can vary in many more ways. GE talkers and Spanish talkers and listeners are less homogenous than the SSBE talkers and listeners, particularly in terms of spectral properties. This could arise from a number of factors; although asked to speak naturally, GE talkers may have modulated their speech while interacting with the non-GE experimenter, and one GE talker also lived in London. There is more variation within non-native speech (Flege et al., 1997), which can explain why the Spanish talkers and listeners are less homogeneous compared to native groups (although Spanish talkers and listeners show similar levels of variation within the groups).

The average correlation between accents across talker-listener pairings are shown in Table 2. To establish patterns of accent similarity, measures of similarity between each listener-talker accent combination based on vowel spectral and durational similarity were entered into separate linear mixed-effects models with talker accent and listener group as fixed effects (also including their interaction term), and by-listener random intercepts. The model was fit using the lmer function from the R package ‘LmerTest’ (Kuznetsova, Brockhoff & Christensen, 2016; all linear mixed-effects models were fit using this algorithm).

For vowel spectral similarity, there were significant effects of listener group ($F(1,32)=29.33$, $p<0.0001$), accent ($F(2,64)=88.21$, $p<0.0001$) and the listener group*accent interaction ($F(2,64)=136.46$, $p<0.0001$). Posthoc pairwise contrasts (Bonferroni-corrected) showed that spectral similarity between accents differed for the two listener groups; SSBE>GE ($p<0.001$)>SpE ($p=0.0173$) for English listeners and SpE>SSBE ($p<0.001$)>GE ($p=0.0105$) for Spanish listeners.

For durational similarity, there were also significant effects of listener group ($F(1,31)=32.17$, $p<0.0001$), accent ($F(2,64)=323.85$, $p<0.0001$) and the listener group*accent interaction ($F(2, 64)=209.76$, $p<0.0001$). For English listeners, pairwise contrasts (Bonferroni-corrected) showed the same pattern of durational similarity as for spectral similarity, SBBE>GE>SpE (all differences $p<0.0001$). For Spanish listeners, the pattern differed slightly; SSBE and SpE were equally durationally similar to the listeners’ accents, and both were more similar than GE ($p<0.001$).

The intelligibility of each accent (in noise only) for each listener is plotted against the spectral and durational similarity between talker-listener vowels in Figure 2. These plots suggest there is a positive relationship between talker-listener accent similarity and intelligibility. To explore further, word recognition accuracy was entered into two logistic mixed-effects models, containing the fixed effect of either vowel spectral similarity or vowel durational similarity. Both models also contained fixed effects of listener group and noise level, the interaction between the similarity measure and listener group, and by-accent, by-listener, by-speaker and by-item random intercepts (by-accent random intercepts were included instead of a fixed effect of accent to avoid over-fitting the model, as the acoustic-phonetic similarity itself is a measure of accent).

For the spectral similarity model, significant effects were found for vowel spectral similarity (χ^2=148.6, df=1, p<0.001), listener group (χ^2=65.9, df=1, p<0.001), noise level (χ^2=2166.4, df=3, p<0.001) and the vowel spectral similarity*listener group interaction (χ^2=100.9, df=1, p<0.001). For the durational similarity model, model, all effects were also significant: vowel durational similarity (χ^2=196.9, df=1, p<0.001), listener group (χ^2=7.8, df=1, p=0.005), noise level (χ^2=2173.4, df=3, p<0.001) and the vowel durational similarity*listener group interaction (χ^2=47.2 df=1, p<0.001). The significant effects of the similarity measures further support the hypothesis that talker-listener acoustic-phonetic similarity accounts for some variance in accent intelligibility. The interactions with listener group show that this relationship differs for the two listener groups. Figure 2 suggests this relationship is stronger for English listeners, which was confirmed by the correlations between each similarity measure and word recognition accuracy for the two listener groups (calculated using the polyserial function from the CTT package in R, Willse, 2014). For both types of similarity, English listeners show a stronger relationship between accent similarity and intelligibility: vowel spectral similarity, rho = 0.33 (English listeners) and 0.28 (Spanish listeners), vowel durational similarity, rho = 0.34 (English listeners) and 0.24 for Spanish listeners. Again, this is likely to be due to the higher variability within the Spanish listeners.

ERP Responses and Accent Intelligibility

Figure 3 shows grand mean difference wave responses at Cz over the 1000ms epoch for each accent and listener group. EEG responses appear to differ between listener groups and across accents within groups, suggesting that talker-listener pairing is able to influence processing in quiet conditions. The PMN and N400 were analysed to investigate this influence at different processing levels.

PMN (200-350ms). PMN amplitudes at Cz are shown in Figure 4. These PMN amplitudes were entered into a linear mixed-effects model with the fixed terms of listener group and accent, the listener group*accent interaction and by-participant random intercepts. Significant effects were found of accent ($F(2, 93)=6.04$, $p=0.0034$) and listener group ($F(1,93)=4.94$, $p=0.0286$), with English listeners showing larger responses than Spanish listeners. However, the listener group*accent interaction was not significant. The PMN is a negative-going response, with amplitudes in response to anomalous input more negative that those to predictable input, so the difference between the two amplitudes would be expected to be negative. However, for listener, only the SSBE-English listener pairing showed a sizeable PMN response, with all others very small or positive (Table 2). This shows that a PMN was not reliably elicited in this study, and therefore any link between PMN response amplitude and accent intelligibility was not further investigated.

N400 (350-500ms). N400 effect amplitudes at Cz are shown in Figure 4. These amplitudes were entered into a linear mixed-effects model fixed effects of accent and listener group and by-participant random intercepts. Significant effects were found of accent ($F(2, 93)=8.20$, $p=0.0005$) and listener group ($F(1, 93)=8.17$, $p=0.0052$), but the interaction was not significant. English listeners again showed larger responses than Spanish listeners, but both groups exhibit the same general pattern in average N400 amplitude for the accents (Table 2). This study is interested in the relative size of N400 for accents within each listener group, rather than overall differences between the groups, so pairwise comparisons (Bonferroni-corrected) were conducted, but only contrasts within groups were considered. For both groups, SSBE > GE (English listeners, $p=0.0267$; Spanish listeners, $p=0.0010$), but SSBE = SpE. For English listeners SpE=GE, but for Spanish listeners, SpE>GE ($p=0.0393$). This seems to arise because an N400 was not reliably elicited in response to GE for Spanish listeners, as shown by the positive average N400 amplitude for this talker-listener pairing.

To investigate the relationship between lexical level processes and accent intelligibility, speech recognition scores were entered into a logistic mixed-effects model with the fixed effects of N400 response amplitude at Cz (anomalous - predictable responses), listener group and noise level, the interaction between the N400 amplitude and listener group and by-accent, by-listener, by-speaker and by-item random intercepts. Significant effects on speech recognition accuracy were found for listener group (χ^2=85.47, df=1, p<0.001), noise level (χ^2=2108.78, df=3, p<0.001), N400 effect amplitude (χ^2=5.31, df=1, p=0.021) and the listener group*N400 interaction (χ^2=15.59, df=1, p<0.001). The significant effect of N400 effect amplitude on intelligibility suggests that talker-listener pairing influences lexical level processes, even in quiet.

Discussion

As anticipated, speech recognition accuracy results show a clear influence of talker-listener pairing on accent intelligibility; English listeners are highly tuned to their own SSBE accent, with intelligibility only being adversely affected at high noise levels. They also find GE as intelligible as SSBE in light noise, but at higher noise levels the intelligibility of GE falls to match that of SpE. Spanish listeners find SSBE and their own SpE accents equally intelligible at all noise levels, with GE consistently less intelligible, even in quiet.

Both vowel spectral qualities and duration significantly affected accent intelligibility, suggesting that acoustic-phonetic similarity between talker-listener accents can influence intelligibility. For English listeners, both measures of accent similarity mirrored accent intelligibility: their accent was most similar to SSBE, followed by GE and finally SpE. This is consistent with Pinet et al.’s findings (2011) for SSBE listeners who showed same patterns for intelligibility and similarity. For Spanish listeners the relationship between accent similarity and intelligibility was weaker due to greater variation within the group, but durational similarity did match the pattern of intelligibility of the accents. However, the SpE accent was more spectrally similar to the Spanish listeners' accent than SSBE, even though both were equally intelligible. This could reflect greater difficulties in acquiring spectral properties of English vowels, as Spanish has only five vowels (Martínez-Celdrán, Fernández-Planas, & Carrera-Sabaté, 2003), or could suggest that exposure to SSBE has helped the accent to become more intelligible than predicted by accent similarity. Similarity also cannot explain other patterns often explained by familiarity including the asymmetry seen across standard-regional pairings, as found by Adank et al. (2009) and Sumner and Samuels (2009) and rapid adaptation to novel accents (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Kraljic & Samuel, 2005; McQueen et al., 2006; Witteman et al., 2013). However, similarity can account for some differences that familiarity cannot, such as inexperienced French listeners finding SSBE and Korean-accented English to be equally intelligible (Pinet et al., 2011).

Taken together, these findings suggest that accent similarity and familiarity may interact to determine the intelligibility of an accent. The acoustic-phonetic similarity between talker-listener accents could establish a baseline intelligibility, where more similar accents are easier to map onto stored abstract representations for the listener's own accent. Exposure may then allow listeners to improve mapping the accent to stored representations, increasing the intelligibility of familiar accents. This interaction can account for asymmetry across standard-regional pairings; regional listeners' familiarity with the standard accent allows efficient mapping to their own representations, but standard listeners lack enough experience of the regional accent to do this, and instead the similarity of the accent is more important in determining its intelligibility. Where listeners show rapid adaptation to an accent after short-term exposure, the initial intelligibility may be influenced by the similarity between the accent and their own, and the experience gained of the accent throughout the experiment leads to more efficient mapping to their stored representations, allowing quicker or more accurate responses. Further research is needed to explore this hypothesis, using a greater variety of talker-listener pairings (including a regional listener group and talker accents that differ in familiarity for listener groups) and tracking adaptation over the experiment.

This study also investigated the influence of talker-listener pairing on online measures of accent processing in quiet. We considered two ERPs related to word recognition, the PMN and N400. The PMN is involved in phonological processing, particularly mapping input onto representations of activated lexical candidates (Connolly & Phillips, 1994). A significant effect of PMN amplitude on intelligibility was found, but a reliable PMN response was only elicited for the SSBE talker-English listener pairing. As the GE accent was also highly intelligible in quiet for English listeners, this shows that the talker-listener pairing can disrupt phonological processing even if this is not observed behaviourally. Brunellière and Soto-Faraco (2013) propose that accent variation modulates listeners' expectations of upcoming words; in a regional accent, listeners form broad phonological expectations about upcoming words, but only generate fine-grained phonological expectations for their own accent. Our findings suggest that native listeners only formed phonological expectations for their own accent, and non-native listeners did not form strong expectations at all. This difference to Brunellière and Soto-Faraco's findings (2013) may stem from the talker-listener pairings in the studies; Brunellière and Soto-Faraco's (2013) standard and regional accents show fairly minor differences in the application of vowel reduction, while the SSBE, GE and SpE accents in the current study differ considerably. There was also quite high variation within the Spanish talker and listener groups, leading to a low level of similarity for this pairing. Based on these findings, it may then be possible that a PMN is only elicited where the talker's speech is very similar to the listener's, in order to allow efficient mapping to representations.

The N400 is involved in later stages of lexical processing relating to semantic integration (Connolly & Phillips, 1994). This ERP was more robust to mismatches in talker-listener pairing, with only the Spanish listener-GE talker failing to elicit an N400 effect. A significant effect was found of N400 on intelligibility, suggesting that talker-listener pairing also influences lexical levels of processing in quiet. For all but one talker-listener pairing, the same patterns were found for N400 amplitude as for the intelligibility of the accents in noise, which is consistent with findings showing more intelligible artificially degraded speech elicits larger N400 effects (Aydelott et al., 2006; Boulenger et al., 2011; Obleser & Kotz, 2011; Strauß et al. 2013). However, for English listeners, the N400 for the SpE accent was not significantly different to either SSBE or GE despite being the least intelligible accent, showing that the influence of talker-listener pairing at the lexical integration stage of word recognition is not solely due to accent intelligibility. Socio-linguistic factors may also influence processing; listeners are more tolerant of variation and errors in non-native than native speech (Hanulíková et al., 2012; Schmid & Yeni-Komshian, 1999) and process it in less detail (Lev-Ari & Keysar, 2012). Word recognition can also be delayed if listeners are less confident of correctly identifying the input (McQueen & Huettig, 2012; Trude, Tremblay & Brown-Schmidt, 2013), possibly to avoid incorrectly recognising words and then needing to apply costly repair processes. Factors such as these may then modulate listeners' expectations about accented speech in addition to accent intelligibility and so influence the N400.

The findings of this study and the inconsistency of findings across EEG studies of accent processing suggests that the specific-talker listener pairing is highly influential on online word recognition processes. Apart from the current study, EEG studies have defined accents only in terms of standard/home, regional or non-native categories, but the importance of the specific talker-listener pairing on accent processing suggests that these broad categories are too general; what applies to one study's standard-regional pairing may not apply to different standard-regional pairings in other studies. Goslin et al. (2012) and

Romero-Rivas et al. (2015) also included multiple accents within their regional and non-native groupings. As the intelligibility of accents within these groups can differ greatly, combining accents in this way may also have obscured processing differences relating to the particular talker-listener pairing.

Conclusion

This study investigated two aspects of the mechanism through which talker-listener pairing influences accent intelligibility; the relationship between accent similarity and intelligibility and online measures of phonological and lexical level processes. Findings provided further support for Pinet et al.'s hypothesis (2011) that the acoustic-phonetic similarity of accents can account for variance in intelligibility. It seems likely that similarity interacts with accent familiarity; similarity may provide a baseline level of intelligibility that exposure can improve on. In terms of online processing, a high level of similarity between talker-listener accents seems to be required to elicit the PMN response, showing that the talker-listener pairing can disrupt processing even for accents that are highly intelligible in quiet. At later lexical integration processes, it seems that the talker-listener pairing is influential, but factors further than similarity and intelligibility also influence processing.

Acknowledgements

This work was funded by the EU FP7 Marie Curie Initial Training Network INSPIRE, under grant agreement number FP7-PEOPLE-2011-290000. Part of this work was presented in "The effect of regional and non-native accents on word recognition processes: A comparison of EEG responses in quiet to speech recognition in noise", a presentation at INTERSPEECH 2014, Singapore, 14-18 September 2014. Thank you to María Luisa García Lecumberri and Edurne Petrirena at the University at the Basque Country for their generous assistance with organising stimuli recording and intelligence testing.

References

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), 520–9. doi:10.1037/a0013552

Aydelott, J., Dick, F., & Mills, D. L. (2006). Effects of acoustic distortion and semantic context on event-related potentials to spoken words. *Psychophysiology*, *43*(5), 454–64. doi:10.1111/j.1469-8986.2006.00448.x

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.<doi:10.18637/jss.v067.i01>.

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *114*(3), 1600. doi:10.1121/1.1603234

Boulenger, V., Hoen, M., Jacquier, C., & Meunier, F. (2011). Interplay between acoustic/phonetic and semantic processes during spoken sentence comprehension: An ERP study. *Brain and language*, *116*, 51–63. doi:10.1016/j.bandl.2010.09.011

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

Brunellière, A., & Soto-Faraco, S. (2013). The speakers' accent shapes the listeners' phonological predictions during speech perception. *Brain and language*, *125*(1), 82–93. doi:10.1016/j.bandl.2013.01.007

Burgos, P., Cucchiarini, C., van Hout, R., & Strik, H. (2014). Phonology acquisition in Spanish learners of Dutch: error patterns in pronunciation. *Language Sciences*, *41*, 129–142. https://doi.org/10.1016/j.langsci.2013.08.015

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

Clopper, C. G., & Bradlow, A. R. (2008). Perception of dialect variation in noise: intelligibility and classification. *Language and speech*, *51*(Pt 3), 175–98. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2744323&tool=pmcentrez&rendert

ype=abstract

Connolly, J., & Phillips, N. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*(3), 256–266. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/jocn.1994.6.3.256

Delorme, A. & Makeig, S. (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics, *Journal of Neuroscience Methods* 134:9-21

Diaz, M., & Swaab, T. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain Research*, (1146), 85–100. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1853329/

Flege, J., Bohn, O., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470. Retrieved from http://www.sciencedirect.com/science/article/pii/S0095447097900528

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), 1276–93. http://doi.org/10.1037/0096-1523.32.5.1276

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and language*, *122*(2), 92–102. doi:10.1016/j.bandl.2012.04.017

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, *30*(3), 251–66. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11523274

Hanulíková, A., van Alphen, P. M., van Gogh, M. M., & Weber, A. (2012). When One Person's Mistake Is Another's Standard Usage: The Effect of Foreign Accent on Syntactic Processing. *Journal of cognitive neuroscience*, *24*(4), 878–887. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/jocn_a_00103

Hayes-Harb, R., Smith, B. L., Bent, T., & Bradlow, A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *Journal of Phonetics*, *36*(4), 664–679. https://doi.org/10.1016/j.wocn.2008.04.002

Hothorn, T., Bretz, F. & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. Biometrical Journal 50(3), 346--363.

HTK Hidden Markov Modelling toolkit (1989). http://htk.eng.cam.ac.uk/

Huckvale, M. (2004). ACCDIST: a Metric for Comparing Speakers' Accents. In *Proceedings of the International Conference on Spoken Language Processing*. Jeju, Korea. Retrieved from http://discovery.ucl.ac.uk/12139/1/12139.pdf

Huckvale, M. (2007). *ACCDIST: an accent similarity metric for accent recognition and diagnosis*. *Speaker Classification II*. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-74122-0_20

Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *The Journal of the Acoustical Society of America*, *117*(2), 896. https://doi.org/10.1121/1.1823291

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*(2), 141–178. https://doi.org/10.1016/j.cogpsych.2005.05.001

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(August), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163. Retrieved from http://www.nature.com/nature/journal/v307/n5947/abs/307161a0.html

Kuznetsova, A. Brockhoff, P. B. & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-33. https://CRAN.R-project.org/package=lmerTest

Lev-Ari, S., & Keysar, B. (2012). Less-Detailed Representation of Non-Native Language: Why Non-

Native Speakers' Stories Seem More Vague. *Discourse Processes*, *49*(7), 523–538. https://doi.org/10.1080/0163853X.2012.698493

Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., Barnes, G., Oostenveld, R., Daunizeau, J., Flandin, G., Penny, W., Friston, K. (2011). EEG and MEG data analysis in SPM8. *Computational intelligence and neuroscience*, *2011*, 852961. doi:10.1155/2011/852961

Martínez-Celdrán, E., Fernández-Planas, A. M., & Carrera-Sabaté, J. (2003). Castilian Spanish. *Journal of the International Phonetic Association*, *33*(2), 255–259. http://doi.org/10.1017/S0025100303001373

McQueen, J. M., & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *The Journal of the Acoustical Society of America*, *131*(1), 509–517. https://doi.org/10.1121/1.3664087

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*(6), 1113–1126. https://doi.org/10.1207/s15516709cog0000_79

Newman, R. L., & Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: evidence for bottom-up and top-down effects on spoken word processing. *Biological Psychology*, *80*(1), 114–21. http://doi.org/10.1016/j.biopsycho.2008.04.008

Newman, R. L., Connolly, J. F., Service, E., & McIvor, K. (2003). Influence of phonological expectations during a phoneme deletion task: evidence from event-related brain potentials. *Psychophysiology*, *40*(4), 640–647. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14570171

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–95. http://doi.org/10.1037/0033-295X.115.2.357

Obleser, J., & Kotz, S. a. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage*, *55*(2), 713–23. doi:10.1016/j.neuroimage.2010.12.020

Oder, A. L., Clopper, C. G., & Ferguson, S. H. (2013). Effects of dialect on vowel acoustics and intelligibility. *Journal of the International Phonetic Association*, *43*(1), 23–35. http://doi.org/10.1017/S0025100312000333

Pinet, M., & Iverson, P. (2010). Talker-listener accent interactions in speech-in-noise recognition: effects of prosodic manipulation as a function of language experience. *The Journal of the Acoustical Society of America*, *128*(3), 1357–65. http://doi.org/10.1121/1.3466857

Pinet, M., Iverson, P., & Huckvale, M. (2011). Second-language experience and speech-in-noise recognition: effects of talker-listener accent similarity. *The Journal of the Acoustical Society of America*, *130*(3), 1653–62. doi:10.1121/1.3613698

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Romero-Rivas, C., Martin, C. D., & Costa, A. (2015). Processing changes when listening to foreign-accented speech. *Frontiers in Human Neuroscience*, *9*(March), 167. http://doi.org/10.3389/fnhum.2015.00167

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

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), 590–608. doi:10.1080/17470218.2013.822009

Stibbard, R. M., & Lee, J.-I. (2006). Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *The Journal of the Acoustical Society of America*, *120*(1), 433. https://doi.org/10.1121/1.2203595

Strauß, A., Kotz, S. A., & Obleser, J. (2013). Narrowed Expectancies under Degraded Speech : Revisiting the N400. *Journal of Cognitive Neuroscience*, *25*(8), 1383–1395. doi:10.1162/jocn

Stringer, L. & Iverson, P. (In preparation). Non-native Speech Recognition Sentences: a new materials set for non-native speech perception research

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://doi.org/10.1016/j.jml.2009.01.001

Tremblay, A. & Ransjin, J. (2015). LMERConvenienceFunctions: Model Selection and Post-hoc

Analysis for (G)LMER Models. R package version 2.10. https://CRAN.R-project.org/package=LMERConvenienceFunctions

Trude, A. M., Tremblay, A., & Brown-Schmidt, S. (2013). Limitations on adaptation to foreign accents. *Journal of Memory and Language*, *69*(3), 349–367. https://doi.org/10.1016/j.jml.2013.05.002

van Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *The Journal of the Acoustical Society of America*, *111*(4), 1906. https://doi.org/10.1121/1.1456928

Weber, A., Broersma, M., & Aoyagi, M. (2011). Spoken-word recognition in foreign-accented speech by L2 listeners. *Journal of Phonetics*, *39*(4), 479–491. https://doi.org/10.1016/j.wocn.2010.12.004

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational Constraints on Functional Specializations for Language Processing: ERP and Behavioral Evidence in Bilingual Speakers. *Journal of Cognitive Neuroscience*, *8*(3), 231–56. http://doi.org/10.1162/jocn.1996.8.3.231

Wells, J. C. (1982). *Accents of English: An Introduction*. Cambridge, UK: Cambridge University Press

Willse, J. T. (2014). CTT: Classical Test Theory Functions. R package version 2.1. https://CRAN.R-project.org/package=CTT

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception & Psychophysics*. http://doi.org/10.3758/s13414-012-0404-y

Wright, R., & Souza, P. (2012). Comparing Identification of Standardized and Regionally Valid Vowels. *Journal of Speech, Language, and Hearing Research*, *55*(February 2012), 182–193. http://doi.org/10.1044/1092-4388(2011/10-0278)b

Xie, X., & Fowler, C. a. (2013). Listening with a foreign-accent: The interlanguage speech intelligibility benefit in Mandarin speakers of English. *Journal of Phonetics*, *41*(5), 369–378. https://doi.org/10.1016/j.wocn.2013.06.003

Figure 1: Accent recognition accuracy as a function of noise level for native and non-native listeners

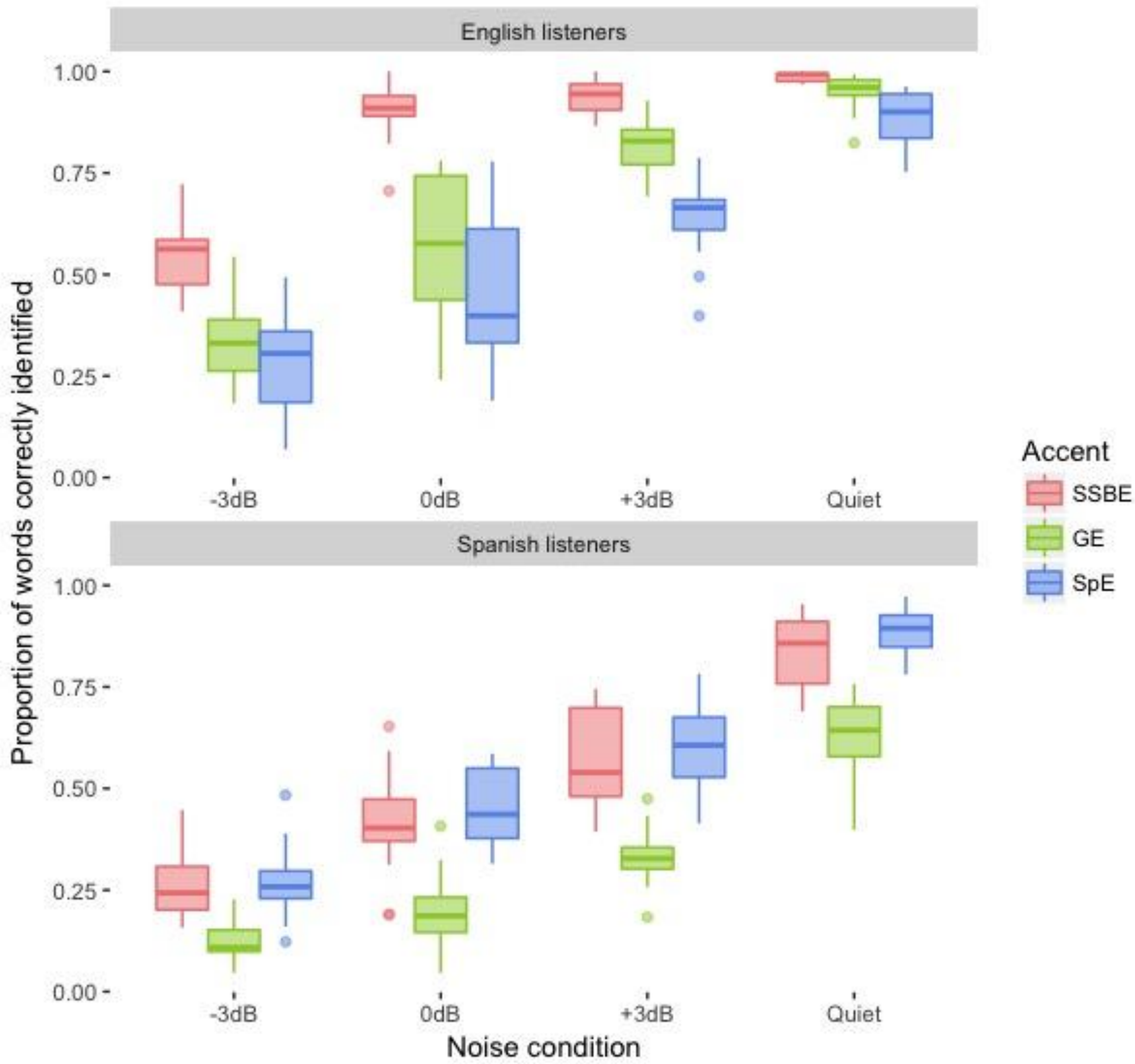
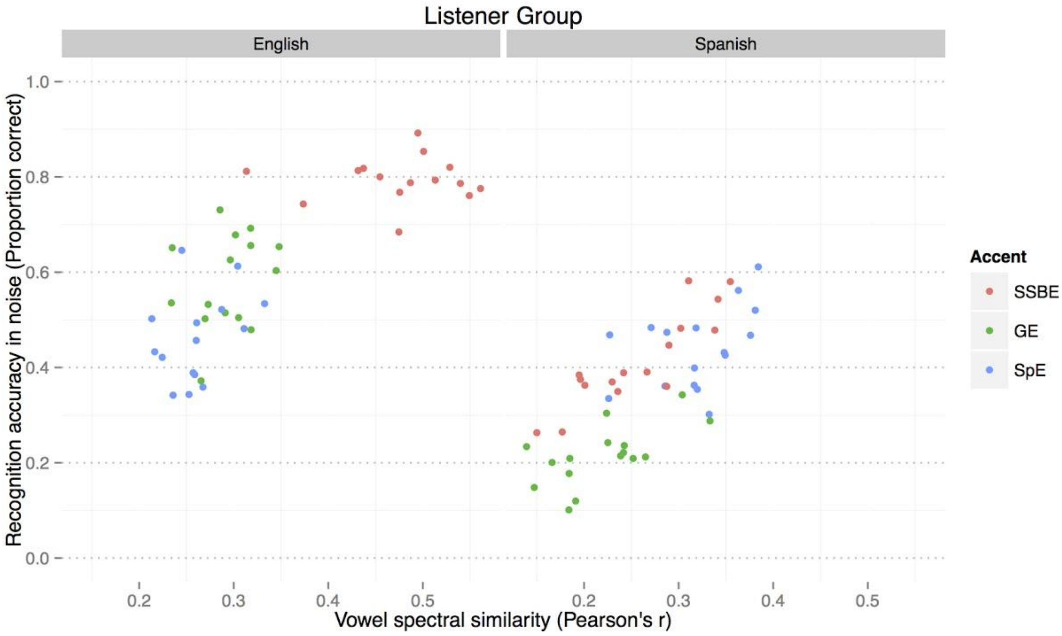


Figure 2: The relationship between accent intelligibility in noise and acoustic-phonetic similarity across talker-listener pairs in terms of vowel spectral similarity and vowel duration similarity



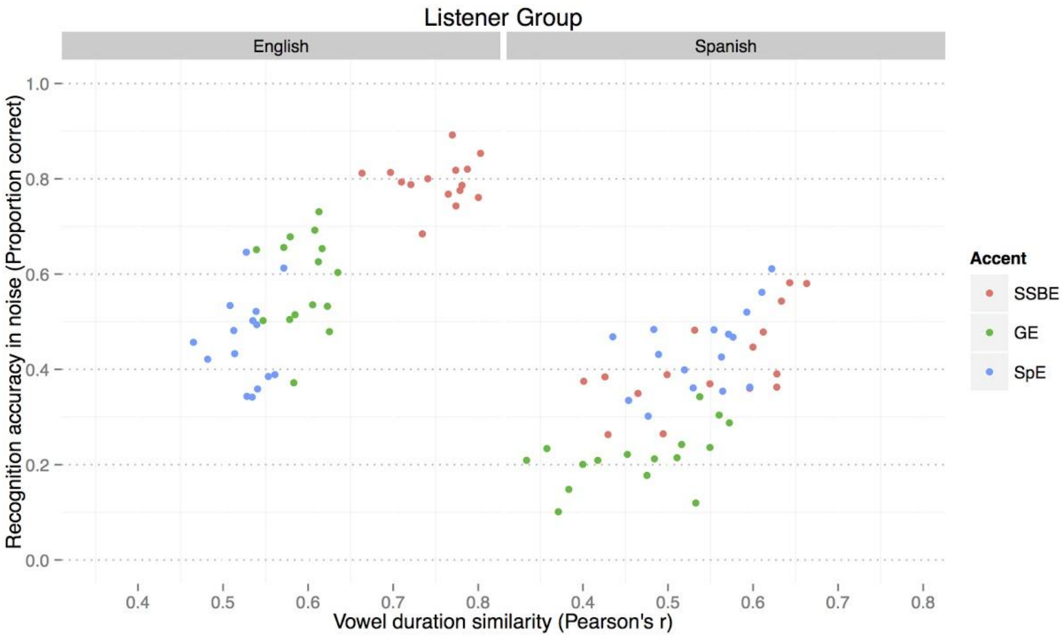


Figure 3: Grand-average waveforms at Cz showing differences in responses to anomalous and predictable final words across the talker-listener pairings

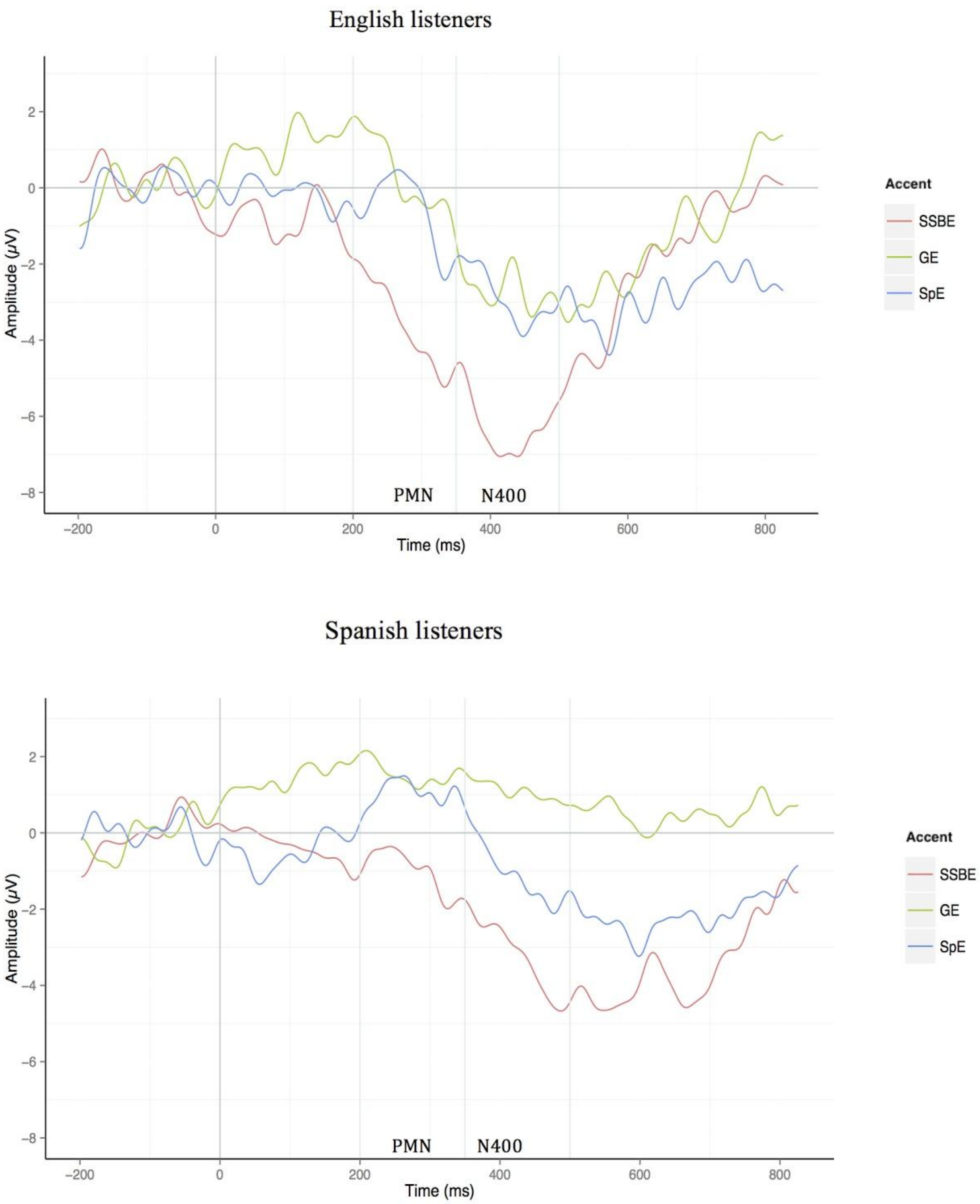


Figure 4. Phonological mapping negativity (PMN) and N400 responses at Cz across talker–listener pairings. SSBE = Standard Southern British English; GE = Glaswegian English; SpE = Spanish-accented English.

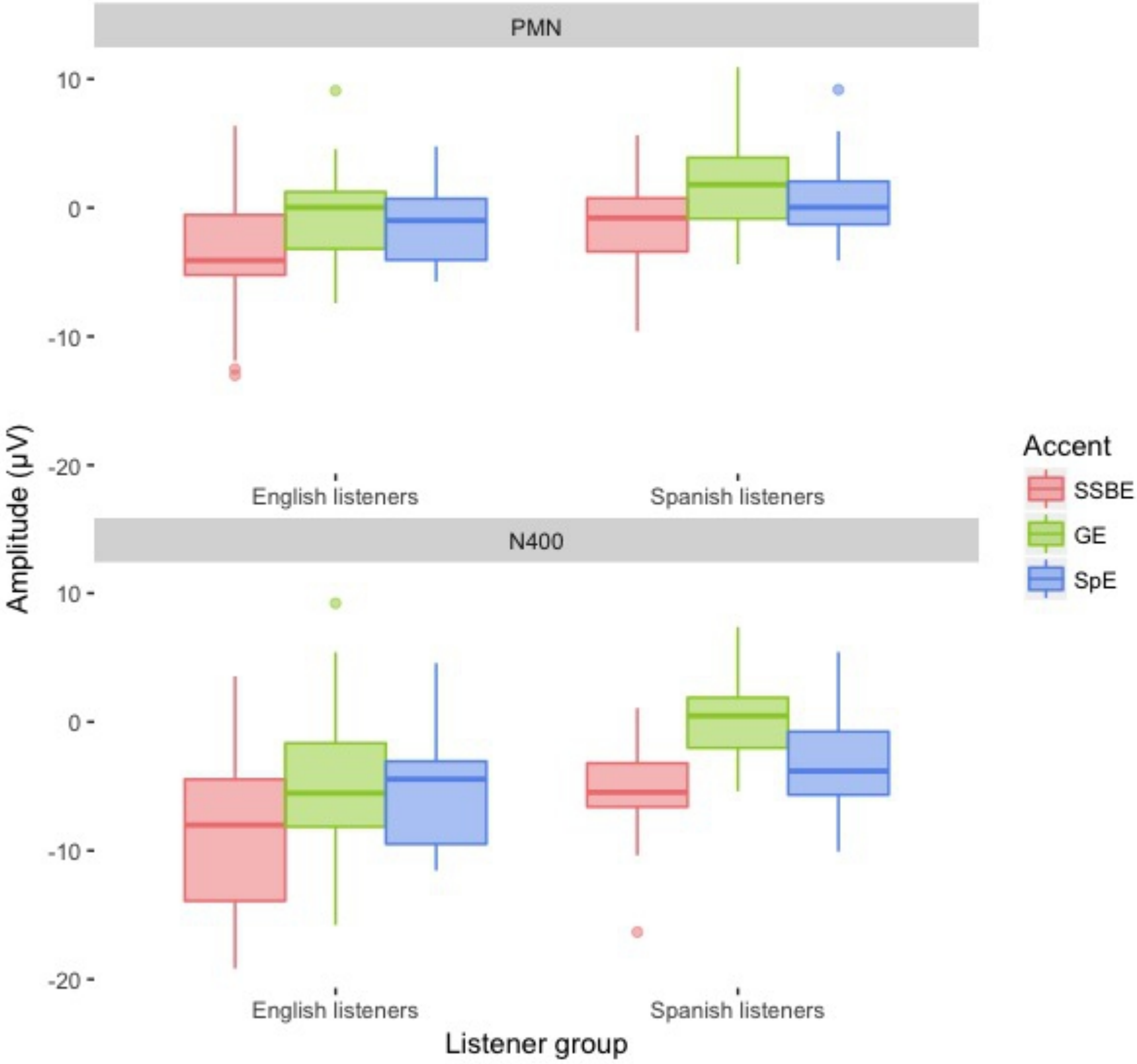


Table 1: An example sentence triplet from the Non-native Speech Recognition sentences. Content overlapping across semantic conditions is shown in bold and final keywords are capitalised.

Semantic condition	**Sentence context**	**Final keyword**	**Example**
Predictable	Strongly constrained	Congruous	**The dolphins are swimming in the SEA.**
Neutral	Weakly constrained	Congruous	The children are playing in the **SEA.**
Anomalous	Strongly constrained	Incongruous	**The dolphins are swimming in the** ROAD.

Table 2: Summary of speech recognition accuracy, accent similarity and EEG findings across talker-listener pairings.

Listener group	Accent	Average intelligibility in noise (% correct)	Spectral vowel similarity (Pearson's r)	Durational vowel similarity (Pearson's r)	PMN amplitude (μV)	N400 amplitude (μV)
English	SSBE	0.80	0.48	0.75	-3.77	-8.01
	GE	0.58	0.30	0.60	0.05	-3.90
	SpE	0.46	0.26	0.53	-0.66	-5.20
Spanish	SSBE	0.41	0.26	0.55	-1.31	-5.59
	GE	0.22	0.22	0.47	1.74	0.48
	SpE	0.44	0.32	0.54	0.80	-3.21