Cognition 179 (2018) 163–170

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents

Jieun Song*, Paul Iverson

Department of Speech, Hearing and Phonetic Sciences, University College London, Chandler House, 2 Wakefield Street, London WC1N 1PF, United Kingdom

ARTICLE INFO

Keywords:
Listening effort
Non-native language (L2) speech recognition
Cognitive load
Entrainment to the speech envelope
N400
Accent processing

ABSTRACT

Speech communication in a non-native language (L2) can feel effortful, and the present study suggests that this effort affects both auditory and lexical processing. EEG recordings (electroencephalography) were made from native English (L1) and Korean listeners while they listened to English sentences spoken with two accents (English and Korean) in the presence of a distracting talker. Neural entrainment (i.e., phase locking between the EEG recording and the speech amplitude envelope) was measured for target and distractor talkers. L2 listeners had relatively greater entrainment for target talkers than did L1 listeners, likely because their difficulty with L2 speech recognition caused them to focus more attention on the speech signal. N400 was measured for the final word in each sentence, and L2 listeners had greater lexical processing in high-predictability sentences than did L1 listeners. L1 listeners had greater target-talker entrainment when listening to the more difficult L2 accent than their own L1 accent, and similarly had larger N400 responses for the L2 accent. It thus appears that the increased effort of L2 listeners, as well as L1 listeners understanding L2 speech, modulates their auditory and lexical processing during speech recognition. This may provide a mechanism to compensate for their perceptual challenges under adverse conditions.

1. Introduction

Understanding speech in a non-native language (L2) can be effortful because one's perceptual and linguistic representations are typically not fully tuned to the L2 (e.g., Flege, 1992; Iverson et al., 2003). However, it is not clear what effects this additional listening effort and cognitive load have on the processes underlying L2 speech recognition. Cognitive load could be expected to interfere with L2 speech recognition; an unrelated visual search task can reduce L1 listeners' reliance on acoustic detail in speech (Mattys, Brooks, & Cooke, 2009; Mattys & Palmer, 2015) as well as reduce auditory cortical responses to non-speech tones (Molloy, Griffiths, Chait, & Lavie, 2015). Similarly, lexical-semantic processing can be disrupted under high cognitive load or in the presence of noise (e.g., Aydelott, Dick, & Mills, 2006; Carey, Mercure, Pizzioli, & Aydelott, 2014; Obleser & Kotz, 2011). However, listening effort can also be thought of as facilitating speech perception, in that it allows L1 listeners to modulate their processing to fit the demands of the listening situation, both by enhancing their representation of the acoustic signal through greater focused attention (e.g., Ding & Simon, 2012) and searching more thoroughly among lexical competitors when the signal is thought to be less reliable (e.g., McQueen & Huettig, 2012). That is, some of the additional effort and load experienced by L2 listeners may be a product of compensatory mechanisms that help overcome L2 perceptual and comprehension difficulties.

The present study investigated speech recognition for attended target speakers in the presence of distractor speakers, for L1 and L2 listeners and speech, using measures of neural entrainment and lexical processing along with behavioral measures of speech comprehension. Understanding speech in two-talker situations is thought to be difficult because of auditory masking, the executive control required to select and suppress information streams, and the interference from the linguistic content of competing speech (e.g., Brungart, 2001). Behavioral research has demonstrated that L1 listeners are more accurate than L2 speakers at understanding speech in this environment (Cooke, Garcia Lecumberri, & Barker, 2008). The reduction of phonetic information due to masking, and the increased cognitive and perceptual loads of two-talker conditions, likely combine with the more general perceptual and cognitive difficulties that listeners have with L2 speech (e.g., see Lecumberri, Cooke, & Cutler, 2010; Stowe & Sabourin, 2005 for a review). However, speech recognition is also affected by the similarity between the talker's and listener's accents rather than being purely driven by overall proficiency; L1 listeners are more accurate with L1-accented speech than with L2 accents, but L2 listeners can sometimes

* Corresponding author.
E-mail address: jieun.song@ucl.ac.uk (J. Song).

https://doi.org/10.1016/j.cognition.2018.06.001
Received 9 November 2016; Received in revised form 31 May 2018; Accepted 4 June 2018
Available online 26 June 2018

be more accurate with L2 accents or at least find L1 and L2 accents to have comparable intelligibility (e.g., Bent & Bradlow, 2003; Pinet, Iverson, & Huckvale, 2011; Van Wijngaarden et al., 2002).

We used EEG to examine how well listeners' auditory processing tracked the acoustics of target and distractor speakers. Previous neural entrainment work has demonstrated that low-frequency neural oscillations in the auditory cortex (1–8 Hz) become phase-locked to the speech amplitude envelope (e.g., Ahissar et al., 2001; Luo & Poeppel, 2007). In two-talker situations, attention can selectively enhance the neural entrainment to the target talker over the distractor, reflecting speech segregation and selection in complex auditory scenes (Ding & Simon, 2012; Kerlin, Shahin, & Miller, 2010; Zion Golumbic et al., 2013). Previous studies have also shown that neural entrainment can be higher when speech is more intelligible, in experiments that used altered acoustic signals such as vocoded speech or added background noise (e.g., Ding, Chatterjee, & Simon, 2014; Peelle, Gross, & Davis, 2013; Gross et al., 2013; Howard & Poeppel, 2010). It has been thought that this link between entrainment and intelligibility occurs because higher-level linguistic processing can aid lower-level auditory tracking of speech (e.g., listeners can predict the onset of upcoming words; Peelle & Davis, 2012). In the present study, one could thus expect that L1 listeners would have higher target-talker entrainment than L2 listeners, both because they find the stimuli to be more intelligible and because their underlying linguistic representations and processes are better optimized for L1 speech. However, it is also possible that the greater difficulty of L2 listeners may force them to focus more attention to the acoustic signal, thereby producing relatively greater entrainment to the target talker than the distractor.

We simultaneously assessed lexical processing using the N400 response. The N400 has been linked to the ease of lexical access, with a greater response for more difficult words (Federmeier, 2007; Kutas & Federmeier, 2000). Any factors that affect lexical access, such as context, word frequency, or repetition, can thus affect N400 amplitude (e.g., Van Petten & Kutas, 1990; for a review, Lau et al., 2008). The N400 has also been linked to the ease of semantic integration of the word with its previous sentence context (e.g., smaller N400 for more congruent words), a process that seems to begin before lexical selection is complete (e.g., Hagoort, 2008). We are considering N400, in the present study, to be an indication of effort at the lexical level. This is accurate in the very broad sense that N400 is greater when lexical processing is more difficult, but it is also plausible in the narrow sense of effort being dependent on the degree that an individual is trying to concentrate on a task (e.g., McGarrigle et al., 2014); there is some evidence that greater attention to the semantic content of a stimulus can increase lexical processing and the N400 (Bonte, Parviainen, Hytönen, & Salmelin, 2006; Mirman, McClelland, Holt, & Magnuson, 2008).

One could expect that listeners might need more lexical processing for more difficult spoken accents, but previous studies have produced highly inconsistent findings. Goslin, Duffy, and Floccia (2012) found reduced N400 responses for foreign-accented speech, whereas Romero-Rivas, Martin, and Costa (2015) found increased N400 responses for foreign-accented speech in initial blocks, which reduced with further exposure; Hanulíková, van Alphen, van Goch, and Weber (2012) found no differences. N400 results for L2 listeners have been similarity inconsistent (e.g., Hahne, 2001; Hahne & Friederici, 2001; Stringer, 2015). It may be that these relationships are complex because N400 can increase with additional lexical processing, but can also decrease when the intelligibility of the signal drops below critical levels (e.g., Obleser & Kotz, 2011; Obleser, Wise, Dresner, & Scott, 2007). Stimulus and listener differences between studies may thus have effects on N400 magnitude that are difficult to understand on their own, although they may become more interpretable in the context of other behavioral and neural measures.

No previous work has linked cortical entrainment to N400. In behavioral work, speech perception under difficult conditions has been previously investigated as a tradeoff between the relative amount of attention focused on acoustic detail versus the reliance on lexical structure, implying that it can be difficult to focus on both levels simultaneously, although this may be more a matter of measurement methodology (e.g., Mattys, et al., 2009; Mattys, White, & Melhorn, 2005). However, it is plausible that listening effort can have more general rather than selective effects, with increased concentration on a task increasing auditory and lexical processing simultaneously. There is evidence too that the degree of cortical entrainment and lexical processing are both linked to higher intelligibility, and in this sense, both may be greater when listening is *less* effortful (e.g., Ding, et al., 2014; Obleser & Kotz, 2011; Obleser et al., 2007; Peelle, et al., 2013).

The present study compared these levels of processing under focused attention by playing English (L1) and Korean (L2) listeners pairs of simultaneous English sentences spoken in two different accents (English and Korean) and presented to separate ears. EEG was recorded while listeners were instructed to selectively attend to one of the talkers. Neural entrainment was measured as the amount of phase coherence between EEG signals and the amplitude envelope of the speech from the target and distractor talkers. We used sentences that differed in terms of the predictability of the final word, which allowed us to simultaneously assess lexical processing. Subjects were instructed to press a button on catch trials (semantically anomalous sentences in the target ear), and the accuracy of the button response was used as a behavioral measure of their speech recognition performance.

2. Methods

2.1. Subjects

Twenty-three native speakers of British English (12 female) and 21 native speakers of Korean (14 female) participated in the experiment. One British and two Korean subjects were excluded from the analyses because of noisy recordings (i.e., bad channels or less than 50% of trials passing artifact rejection). All subjects were right-handed adults under 35 years old ($M_{English}$ = 21.8 y, M_{Korean} = 26.5 y) without self-reported hearing or neurological impairments. Korean speakers reported that they started learning English at school in South Korea at an average age of 10 years (5–14 y), and that they had not lived in English-speaking countries before they became adults. Their average length of residence in English-speaking countries was 1 year (1–31 months).

2.2. Materials

English sentences were recorded by female native speakers of Southern British English and Korean (one each). The Korean speaker studied English at school in Korea and had lived in the U.K for one year. The stimuli consisted of 720 pairs of sentences presented simultaneously in different ears, with a different talker in each ear, and with sentences matched in duration. The average duration of the British sentences was originally 0.44 s shorter than that of the Korean speaker, so the sentences of the British speaker were lengthened and those of the Korean speaker were shortened by 10% using an overlap-add procedure (Boersma & Weenink, 2016). All of the stimuli had 44,100 16-bit samples per second. The stimuli were counterbalanced between subjects with order randomized.

The sentences varied in the predictability of the final word to allow for measurement of N400. We used an existing corpus of N400 stimuli designed for L2 learners (Stringer, 2015), and expanded the number of sentences by editing another L2 sentence corpus (Calandruccio & Smiljanic, 2012) to vary final-word predictability. High cloze probability sentences comprised 42.5% of the corpus, consisting of strongly constraining sentence contexts and congruent final words (mean 93% cloze probability; e.g., *Beef and milk come from cows*). Another 42.5% of the stimuli were low cloze probability sentences (cloze probability < 40%; e.g., *The man draws pictures of cows*). The remaining 15% of the

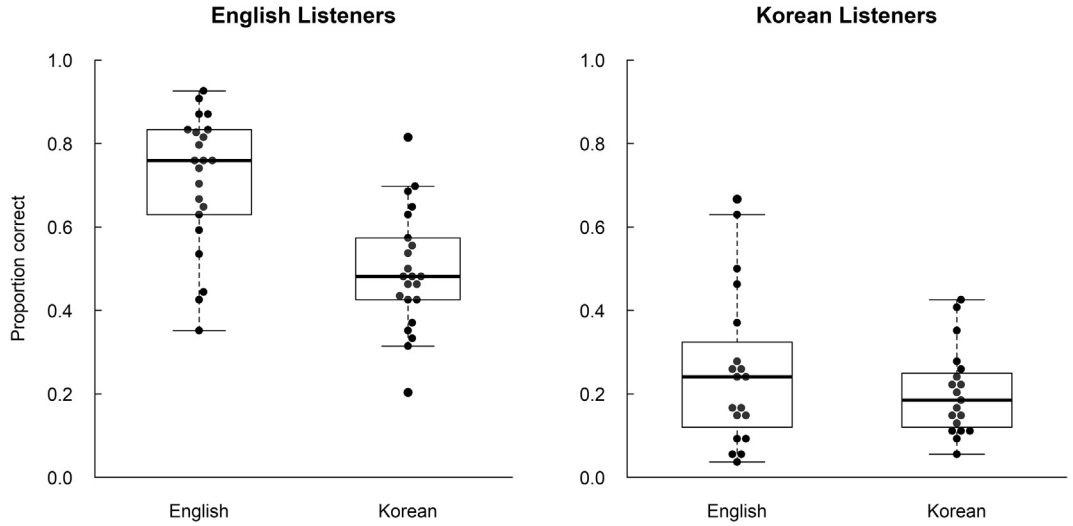

Fig. 1. Combined boxplot and beeswarm plots of the proportion of correctly identified anomalous sentences by speaker accent (English and Korean) for English (L1) and Korean (L2) listeners. English listeners' performance was significantly better than that of Korean listeners, but was poorer for the Korean accent compared to the English accent. Korean listeners had similar levels of performance for both accents.

stimuli were semantically anomalous sentences that were used as catch trials (e.g., *Beef and milk come from bays*). The three types of sentences were randomized within each block.

2.3. Procedure

During EEG recording, subjects selectively attended to a target ear/talker and pressed a button whenever they heard a semantically anomalous sentence in that ear. The experiment had 8 blocks of 90 simultaneous sentences, with the target talker and ear alternating between blocks. The duration of the inter-stimulus silence intervals was randomly jittered from 1.5 to 1.7 s.

2.4. EEG recording and analysis

EEG was recorded through a Biosemi Active Two system with 64 (Ag/AgCl) electrodes mounted on an elastic cap and 7 external electrodes (left and right mastoids, nose, two vertical and horizontal EOG electrodes). Recordings were made with a sampling rate of 2048 Hz. Electrode impedances were kept within the range of ± 25 kΩ. The stimuli were presented via Etymotic ER-1 insert earphones.

After recording, the EEG signals were referenced to the average of the left and right mastoids. Noisy channels were interpolated. The data were high-pass filtered at 0.1 Hz and low-pass filtered at 40 Hz using Butterworth filters as implemented in the ERPlab toolbox (Lopez-Calderon & Luck, 2014) of EEGlab (Delorme & Makeig, 2004). Independent Component Analysis was applied to correct for eye blinks and horizontal eye movements. All pre-processing procedures, except for filtering, were performed in Matlab using the Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011).

2.4.1. Coherence analysis

We fit multivariate Temporal Response Functions (mTRF; Crosse, Di Liberto, Bednar, & Lalor, 2016; O'Sullivan et al., 2015) in backward models relating the EEG data for each subject to the Hilbert envelopes of the target and distractor sentences. The mTRFs were trained over −150 to 500 ms time lags across all of the 64 channels; target and distractor talkers were modeled separately and the signals were filtered (2–8 Hz) to specifically model the delta-theta range. This produced a set of channel weights over time that linearly mapped the EEG responses back to the original auditory envelopes, with these mTRFs being analogous to traditional auditory evoked potentials (i.e., peaks that mirror P1-N1-P2 responses; Crosse et al., 2016). A leave-one-out approach was taken to avoid over-fitting the data; the response for each sentence was predicted based on a model trained on all other sentences, omitting each sentence from its own training set. The data was assessed in terms of the coherence between the predicted envelopes derived from the EEG data and the actual envelopes that were presented; the data was divided into 2-s Hann windows with 50% overlap, and coherence was calculated from the cross-spectral density of the FFT of the predicted and original signals, divided by the power spectrum of each signal. This thus produced an assessment of how closely the EEG data was phase-locked to the original amplitude envelopes of targets and distractors at a range of frequencies (0.5 Hz resolution).

2.4.2. N400 analysis

The data was segmented into epochs time-locked to the final-word onsets (200 ms pre-stimulus and 1000 ms post-stimulus intervals). Trials with amplitude exceeding ± 150 μV were rejected, and the rejection rate averaged across subjects was 12.6%. A non-parametric cluster-based permutation analysis (Maris & Oostenveld, 2007) was performed to investigate the distribution of the N400 response across the scalp, comparing the high cloze and low cloze conditions for each electrode in a 200–650 ms time window. There were significant differences in a large cluster (on average, 56 of the 64 electrodes) across the entire time window, $p < 0.001$. The main statistical analysis was then performed using a more narrow time window (300–500 ms) and smaller midline electrode set (Fz, FCz, Cz, CPz and Pz) in the interest of avoiding including other potentials (e.g., phonological mismatch negativity); this smaller selection fit within the larger significant cluster.

3. Results

English listeners were more accurate at the behavioral response (i.e., the proportion of correctly identified anomalous sentences), and had a larger intelligibility advantage for English- than Korean-accented speech (Fig. 1; $M_{\text{English accent}} = 0.71$, $M_{\text{Korean accent}} = 0.49$). Korean listeners had more similar intelligibility for both accents although they still found English-accented speech more intelligible ($M_{\text{English accent}} = 0.26$, $M_{\text{Korean accent}} = 0.20$). A logistic mixed-model analysis was performed on the behavioral results with listener group and speaker accent as independent variables, and random intercepts for each subject and sentence stimulus. The results verified that there were main effects of listener group, $\chi^2(1) = 70.83$, $p < 0.001$, and speaker accent, $\chi^2(1)$

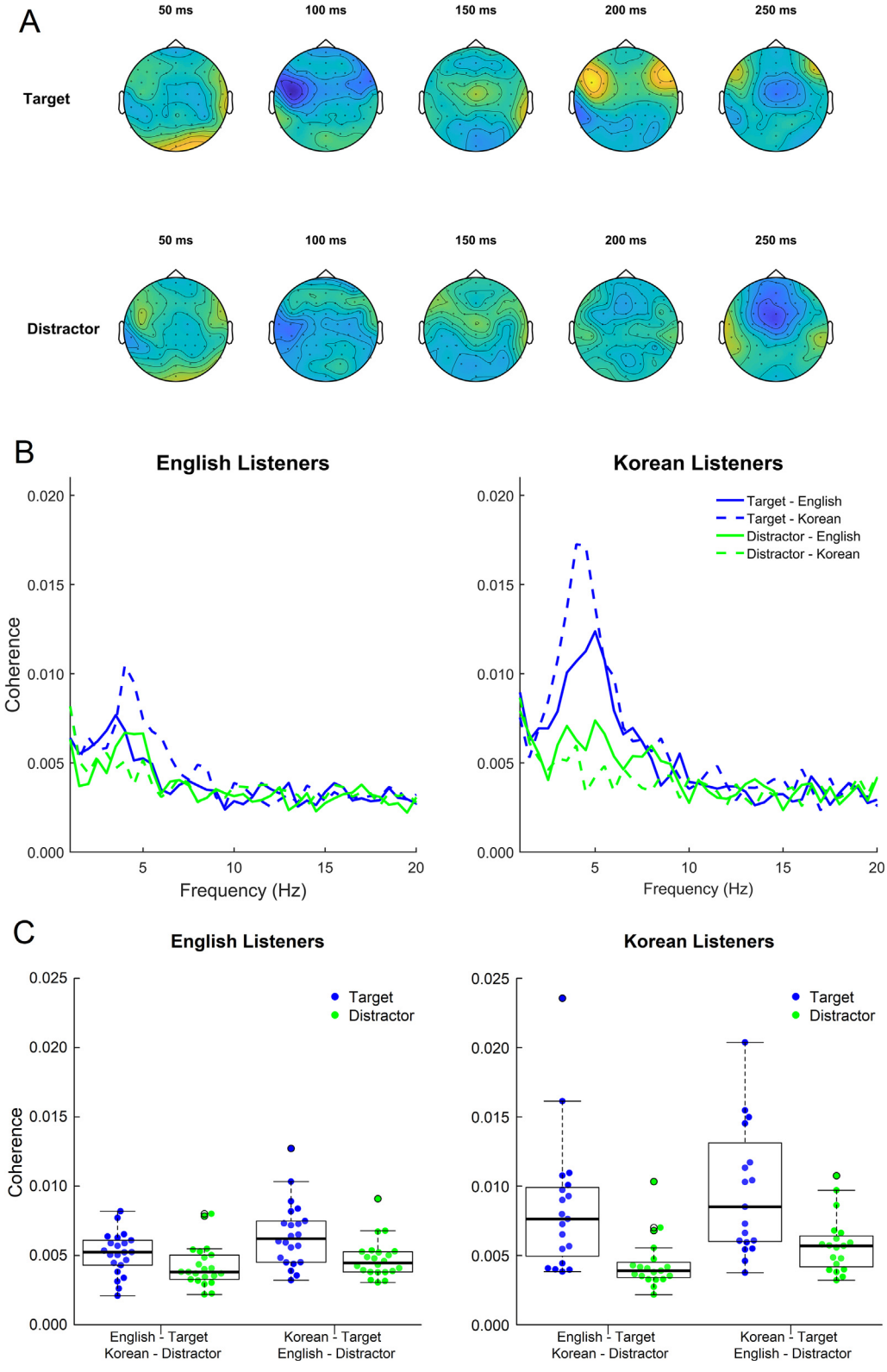

Fig. 2. Results of the coherence analysis for English (L1) and Korean (L2) listeners. (A) Scalp topographies of the average channel weights of the target and distractor decoders over time lags between 50 and 250 ms. The differences in weights for the target and distractor decoders were evident at 100 and 200 ms. (B) Coherence values plotted as a function of frequency. (C) Combined boxplot and beeswarm plots of individual coherence values. Koreans had significantly greater entrainment to the target than did the English, suggesting that the greater listening effort for L2 speech enhanced the selectivity of their auditory processing. Coherence was also greater when the target was Korean-accented speech for both groups of listeners.

$= 27.98$, $p < 0.001$, as well as a significant interaction between these two variables, $\chi^2(1) = 20.63$, $p < 0.001$. Although the accuracy may appear low, the false alarms were much lower (i.e., button presses to non-anomalous sentences; $M_{English} = 0.02$, $M_{Korean} = 0.03$), suggesting very conservative response biases. The task was also difficult because distinguishing low and anomalous sentences requires finer-grained

processing of meaning than high vs. anomalous and there was only 1.5–1.7 s between sentences.

Despite the fact that Koreans found this task harder, they had greater neural entrainment to target talkers than did English listeners (Fig. 2), opposite to previously reported positive relationships between speech entrainment and intelligibility (e.g., Peelle et al., 2013). That is, both listener groups had coherence peaks in the delta-theta range (2–8 Hz), with the target having greater coherence than the distractor, but with this effect larger for L2 than L1 listeners. This likely occurred because L2 listeners had to focus more attention on the target speech signal due to their recognition difficulty. A mixed-model analysis was conducted with coherence values averaged in the delta-theta range as the dependent variable; listener group, speaker accent, and target (i.e., target vs. distractor) as independent variables; and with by-subject random intercepts. The interaction between listener group and target was significant, $\chi^2(1) = 18.53$, $p < 0.001$, as well as the main effects of listener group, $\chi^2(1) = 9.09$, $p = 0.003$, and target, $\chi^2(1) = 62.62$, $p < 0.001$. There was also a significant interaction between speaker accent and target, $\chi^2(1) = 13.29$, $p < 0.001$; entrainment to the target was greater when listeners attended to the Korean speaker than the English. However, there was no significant three-way interaction including listener background, $p = 0.332$, suggesting that both groups attended more to the Korean accent even though English speakers had greater difficulty with this accent in the behavioral test.

Average channel weights for the target and distractor decoders are displayed in Fig. 2 over a range of time lags (50 – 250 ms), with the positive or negative magnitude indicating which channel and time points were more critical to the mapping between EEG and speech signals (Crosse et al., 2016). The differences between the target and distractor decoders were most evident at negative weights near 100 ms and positive weights near 200 ms, with the attended decoder having increased weights around bilateral frontotemporal electrodes. Although the channel weights are not readily interpretable in terms of neural sources (e.g., Haufe et al., 2014), our obtained weights seem related to the N1-P2 auditory ERPs that are often found in EEG recordings.

A mixed-model analysis was performed for final-word N400 responses in target-talker sentences, with N400 amplitudes as the dependent variable; listener group, speaker accent and sentence type as independent variables; and with by-subject random intercepts. The results demonstrated a typical N400 effect (Fig. 3), with greater amplitudes in low than high cloze probability sentences, $\chi^2(1) = 100.33$, $p < 0.001$, suggesting that individuals had more effortful lexical processing when the final word was less predictable. English listeners had significantly greater context-related differences between low- and high-cloze sentences than did Korean listeners; the interaction between sentence type and listener group was significant, $\chi^2(1) = 12.22$, $p < 0.001$. Korean listeners had similar N400 amplitudes for both accents, but English listeners had larger N400 amplitudes for the Korean accent; the interaction between listener group and speaker accent, $\chi^2(1) = 3.83$, $p = 0.050$, and the main effect of speaker accent were significant, $\chi^2(1) = 6.45$, $p = 0.011$. Lexical processing thus mirrored the intelligibility of these sentences, with English listeners needing additional lexical processing to compensate for the more-difficult Korean accent, and Korean listeners needing effortful processing for both sentence types and accents.

Individual-differences correlations compared average behavioral accuracy, target-talker selectivity (i.e., difference in coherence between target and distracting talkers), and the N400 effect (i.e., difference in N400 between HP and LP conditions). There was a significant correlation between the N400 effect and behavioral accuracy across all subjects, $r = 0.53$, $p < 0.001$, and separately within the English, $r = 0.50$, $p = 0.019$ and Korean listener groups, $r = 0.49$, $p = 0.032$. Behavioral accuracy was significantly correlated with target-talker selectivity (coherence) when calculated across all subjects, $r = -0.36$, $p = 0.019$, but not within English, $r = 0.06$, $p = 0.779$, or Korean groups, $r = -0.15$, $p = 0.541$; this appears to reflect group-level differences between English and Korean listeners, with Korean listeners having larger target-talker selectivity, and lower behavioural accuracy than English listeners. Target-talker selectivity was not significantly correlated with N400 effect for the entire group of subjects, $r = -0.28$, $p = 0.072$, or within English, $r = -0.41$, $p = 0.057$, or Korean groups, $r = -0.02$, $p = 0.936$. The results thus suggest that target-talker entrainment is not a simple function of ease of lexical processing, and speech intelligibility may be more directly related to N400 than entrainment.

4. Discussion

It has long been obvious that individuals need to listen harder during L2 speech comprehension. Our results demonstrate that this increased effort produces an adaptive change in the neural processing of speech by L2 listeners, enhancing their neural tracking of attended speech streams. L1 listeners appeared to likewise enhance their neural tracking through focused attention for difficult L2 accents. Lexical processing also increased under L2 listening and accents, but in this case it was less clear whether the increases in N400 were an automatic response to lexical difficulty or to some extent resulted from the listeners exerting greater concentration on semantic content under difficult conditions.

Previous work had found greater selective entrainment for more intelligible speech (e.g., Kong, Somarowthu, & Ding, 2015; Rimmele, Zion Golumbic, Schröger, & Poeppel, 2015), and it had been hypothesized that this occurs because listeners exploit higher-level linguistic information to aid lower-level auditory tracking of the speech envelope (e.g., Peelle et al., 2013; Rimmele et al., 2015). The present work demonstrates that cortical entrainment to speech can also be greater when intelligibility is lower, even when heard by L2 listeners who have less developed higher-level linguistic processes for that language. This previous work on intelligibility and entrainment has mostly varied intelligibility through acoustic degradation (e.g., vocoders or added noise) or the level of the distractor talker (e.g., Kong et al., 2015), so it is possible in these previous studies that the greater entrainment was more linked to greater signal clarity (e.g., natural spectral-temporal modulations) than to better speech comprehension (c.f., Ding & Simon, 2014). In our study, the speech signals were not acoustically degraded and there were interactions with the listener groups, thereby more directly revealing effects of attention independent from main effects of the signal.

However, it remains uncertain why Korean listeners had increased entrainment to the Korean accent (i.e., no significant 3-way interaction between target, accent, and listener group) despite having been able to understand both accents with relatively similar accuracy. Even though the differences in their behavioral-response accuracy for the two accents were small, it is possible that the Korean-accented speech was still more difficult for them to understand. Accent similarity is thought to promote intelligibility, such that low-proficiency L2 listeners can have higher intelligibility for L2 speech, even though this effect diminishes or reverses for more proficient listeners (e.g., Pinet et al., 2011; Van Wijngaarden et al., 2002; Imai, Walley, & Flege, 2005). However, L2 speech can also be more variable, with effects of fluency (e.g., pausing or slowing when producing less familiar words) and inconsistent phonetic realizations of words (e.g., L2 “errors” that are not always made); the match between the speech and a listener’s expectations can thus be less reliable than with more-consistent L1 speech. Korean listeners may have needed greater listening effort at an auditory level in response to these L2 speech issues, in order to achieve the same levels of accuracy that they had with L1 speech. With only two speakers, it is always possible that there were speaker differences that were unrelated to L1/L2 status (e.g., our L1 speaker was clear and articulate in comparison to other L1 speakers we possibly could have recorded), but the L1 and L2 stimuli were carefully controlled in terms of duration and recording quality; sentences with pauses and reading errors were re-recorded.

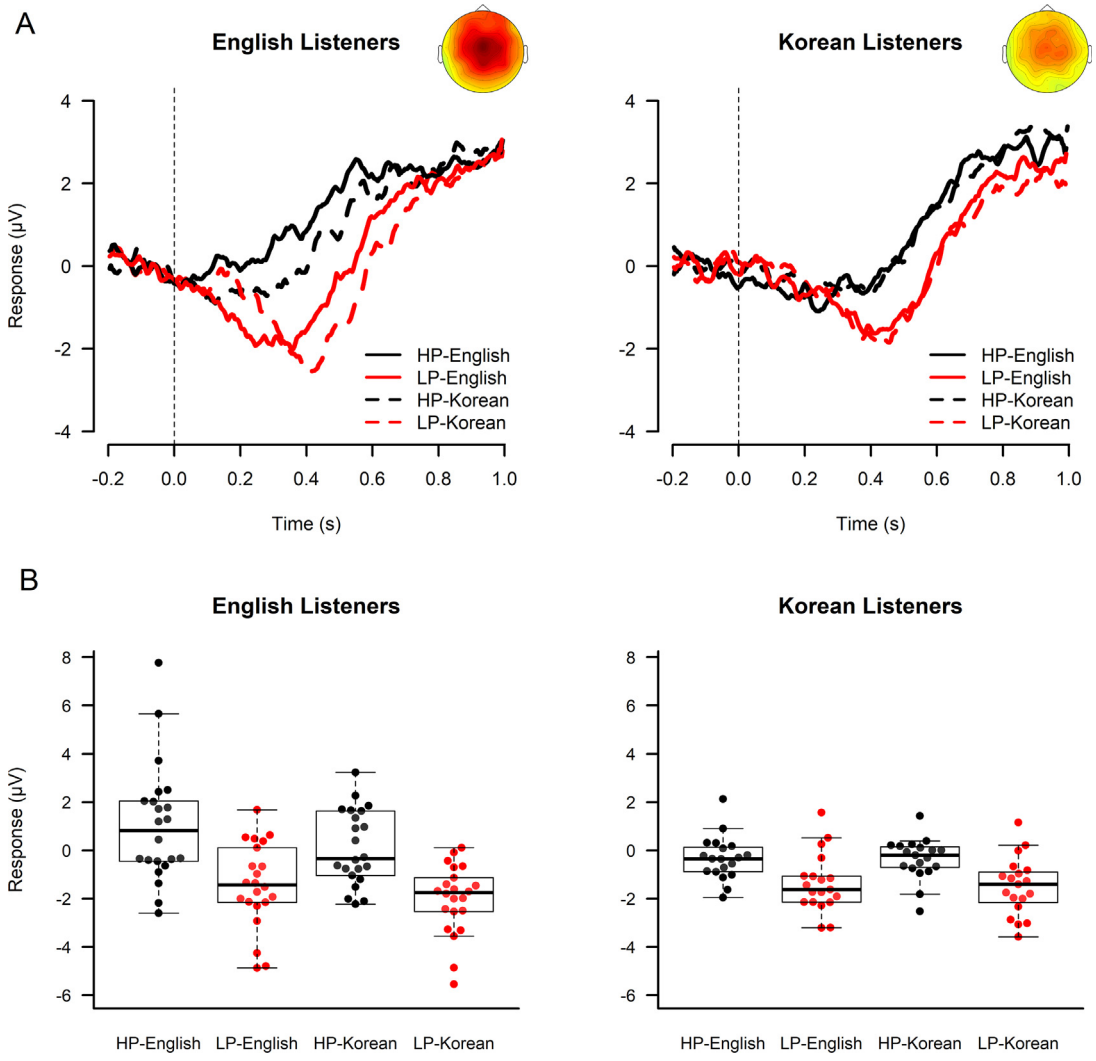

Fig. 3. Results of the N400 analysis for English (L1) and Korean (L2) listeners. (A) Grand average ERPs (N400) for sentence-final words by sentence type (HP: high cloze probability sentences, LP: low cloze probability sentences) and speaker accent (English and Korean) are plotted, with sensor-space topographies of the mean N400 differences between HP and LP sentences. (B) Combined boxplot and beeswarm plots of individual N400 values. English listeners had greater N400 differences based on the predictability of the sentences (e.g., smaller N400 magnitude for HP sentences), and had larger overall N400 amplitudes for the Korean accent. Korean listeners had smaller differences based on the predictability of the sentences, and similar overall N400 amplitudes for both accents.

The N400 results were a closer fit with previous papers; L1 listeners have sometimes been found to have more lexical processing when listening to an L2 accent, and L2 listeners have sometimes been found to have more similar lexical processing for high- and low-predictability sentences (Hahne, 2001; Romero-Rivas et al., 2015; c.f., Goslin, et al., 2012; Hahne & Friederici, 2001). Even though L2 speech errors are often thought of in terms of isolated phonological substitutions (e.g., the word *bit* being pronounced as *beat*), there are typically broader and more continuous degrees of mismatches between speakers and listeners, which may cause a larger number of word candidates to be activated and thereby increase lexical competition (e.g., Weber & Cutler, 2004). It is also possible that this activation of lexical candidates is mediated by accent adaptation processes at earlier perceptual levels (e.g., Goslin, et al., 2012). However, the magnitude of the N400 also can collapse in stimulus conditions where lexical processing is consistently less successful (e.g., Obleser & Kotz, 2011; Obleser et al., 2007) or when the listener has less expectation that the speech is meaningful (Bonte, Parviainen, Hytönen, & Salmelin, 2006). In the present study, it appears that our particular Korean accent was strong enough to require additional lexical processing by L1 listeners (e.g., larger numbers of activated candidates), but not so strong that the speech began to be meaningless. Likewise, our L2 listeners were able to compensate for their recognition difficulties by deploying more lexical processing than L1 listeners, at least for high-predictability sentences, and these listeners were proficient enough so they were not overwhelmed by failures of lexical selection.

It thus appears that listening effort can have diverse effects on the auditory and lexical processing of speech, and that these effects differ for L1 and L2 listeners. It is possible that the listener strategies used here (e.g., focusing more on the speech signal when listening is difficult) may not necessarily have been effective for improving intelligibility, but previous research has shown that higher-level attention or additional listening effort can indeed aid speech comprehension (e.g., O'Sullivan et al., 2015; Erb & Obleser, 2013; Peelle, Troiani, Grossman, & Wingfield, 2011). That is, we cannot know how well listeners would have understood the speech had this enhanced processing not occurred, but the present study demonstrates that speech processing does not occur in a purely involuntary, bottom-up fashion, but is modulated by a complex combination of auditory-phonetic processing, linguistic knowledge, and attention.

Acknowledgments

We thank G. Borghini and K. McCarthy for comments on the manuscript. This study was supported by the Economic and Social Research Council of the UK and the Kwanjeong Educational Foundation of South Korea.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.cognition.2018.06.001.

References

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America, 98*(23), 13367–13372. http://dx.doi.org/10.1073/pnas.201400998.

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), 454–464. http://dx.doi.org/10.1111/j.1469-8986.2006.00448.x.

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America, 114*(3), 1600–1610. http://dx.doi.org/10.1121/1.1603234.

Boersma, Paul & Weenink, David (2016). Praat: Doing phonetics by computer [Computer program]. Version 5.3.69, retrieved 28 March 2014 from < http://www.praat.org/ > .

Bonte, M., Parviainen, T., Hytönen, K., & Salmelin, R. (2006). Time course of top-down and bottom-up influences on syllable processing in the auditory cortex. *Cerebral Cortex, 16*(1), 115–123. http://dx.doi.org/10.1093/cercor/bhi091.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers, The Journal of the Acoustical Society of America, 109, 1101–1109. org/10.1121/1.1345696.

Calandruccio, L., & Smiljanic, R. (2012). New sentence recognition materials developed using a basic non-native English Lexicon. *Journal of Speech, Language, and Hearing Research, 55*(5), 1342–1355. http://dx.doi.org/10.1044/1092-4388(2012/11-0260).

Carey, D., Mercure, E., Pizzioli, F., & Aydelott, J. (2014). Auditory semantic processing in dichotic listening: Effects of competing speech, ear of presentation, and sentential bias on N400s to spoken words in context. *Neuropsychologia, 65*, 102–112. http://dx.doi.org/10.1016/j.neuropsychologia.2014.10.016.

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America, 123*(1), 414–427. http://dx.doi.org/10.1121/1.2804952.

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience, 10*(604), 1–14. http://dx.doi.org/10.3389/fnhum.2016.00604.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. http://dx.doi.org/10.1016/j.jneumeth.2003.10.009.

Ding, N., Chatterjee, M., & Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage, 88*, 41–46. http://dx.doi.org/10.1016/j.neuroimage.2013.10.054.

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences of the United States of America, 109*(29), 11854–11859. http://dx.doi.org/10.1073/pnas.1205381109.

Ding, N., & Simon, J. J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience, 8*(May), 311. http://dx.doi.org/10.3389/fnhum.2014.00311.

Erb, J., & Obleser, J. (2013). Upregulation of cognitive control networks in older adults' speech comprehension. *Frontiers in Systems Neuroscience, 7*(December), 116. http://dx.doi.org/10.3389/fnsys.2013.00116.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491–505. http://dx.doi.org/10.1111/j.1469-8986.2007.00531.x.

Flege, J. (1992). Speech learning in a second language. In C. Ferguson, L. Menn & C. Stoel-Gammon, Phonological development: Models, research, and application (1st ed., pp. 565-604). York Press.

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

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLoS Biology, 11*(12), 1–14. http://dx.doi.org/10.1371/journal.pbio.1001752.

Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 363*(1493), 1055–1069. http://dx.doi.org/10.1098/rstb.2007.2159.

Hahne, A. (2001). What's different in second-language processing ? Evidence from event-related brain potential. *Journal of Psycholinguistic Research, 30*(3), 251–266.

Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. Bilingualism: Language and Cognition, 4(2), 123–141. 10.1017/S1366728901000232.

Hanulíková, A., van Alphen, P. M., van Goch, 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. http://dx.doi.org/10.1162/jocn_a_00103.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J. D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage, 87*, 96–110. http://dx.doi.org/10.1016/j.neuroimage.2013.10.067.

Howard, M. F., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology, 104*, 2500–2511. http://dx.doi.org/10.1152/jn.00251.2010.

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), 896–907. http://dx.doi.org/10.1121/1.1823291.

Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*(1), B47–B57. http://dx.doi.org/10.1016/s0010-0277(02)00198-1.

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 30*(2), 620–628. http://dx.doi.org/10.1523/JNEUROSCI.3631-09.2010.

Kong, Y.-Y., Somarowthu, A., & Ding, N. (2015). Effects of spectral degradation on attentional modulation of cortical auditory responses to continuous speech. *Journal of the Association for Research in Otolaryngology, 16*(6), 783–796. http://dx.doi.org/10.1007/s10162-015-0540-x.

Kutas, M., & Federmeier, K. D. (2000). Electropsysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science, 12*(12), 463–470. http://dx.doi.org/10.1016/S1364-6613(00)01560-6.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. Nature Reviews Neuroscience, 9(12), 920–933. 10.1038/Nrn2532.

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication, 52*(11–12), 864–886. http://dx.doi.org/10.1016/j.specom.2010.08.014.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*, 1–14. http://dx.doi.org/10.3389/fnhum.2014.00213.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54*(6), 1001–1010. http://dx.doi.org/10.1016/j.neuron.2007.06.004.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190. http://dx.doi.org/10.1016/j.jneumeth.2007.03.024.

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology, 59*(3), 203–243. http://dx.doi.org/10.1016/j.cogpsych.2009.04.001.

Mattys, S. L., & Palmer, S. D. (2015). Divided attention disrupts perceptual encoding during speech recognition. *The Journal of the Acoustical Society of America, 137*(3), 1464–1472. http://dx.doi.org/10.1121/1.4913507.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. *Journal of Experimental Psychology: General, 134*(4), 477–500. http://dx.doi.org/10.1037/0096-3445.134.4.477.

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group "white paper". *International Journal of Audiology, 53*(7), 433–440. http://dx.doi.org/10.3109/14992027.2014.890296.

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. http://dx.doi.org/10.1121/1.3664087.

Mirman, D., McClelland, J. L., Holt, L. L., & Magnuson, J. S. (2008). Effects of Attention on the Strength of Lexical Influences on Speech Perception: Behavioral Experiments and Computational Mechanisms. *Cognitive Science, 32*, 398–417.

Molloy, K., Griffiths, T. D., Chait, M., & Lavie, N. (2015). Behavioral/cognitive inattentional deafness: Visual load leads to time-specific suppression of auditory evoked responses. *The Journal of Neuroscience, 35*(49), 16046–16054. http://dx.doi.org/10.1523/JNEUROSCI.2931-15.2015.

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., ... Lalor, E. C. (2015). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral Cortex, 25*(7), 1697–1706. http://dx.doi.org/10.1093/cercor/bht355.

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage, 55*(2), 713–723. http://dx.doi.org/10.1016/j.neuroimage.2010.12.020.

Obleser, J., Wise, R. J. S., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience, 27*(9), 2283–2289. http://dx.doi.org/10.1523/JNEUROSCI.

4663-06.2007.
Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 1–9. http://dx.doi.org/10.1155/2011/156869.
Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3*, 1–17. http://dx.doi.org/10.3389/fpsyg.2012.00320.
Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex, 23*(6), 1378–1387. http://dx.doi.org/10.1093/cercor/bhs118.
Peelle, J. E., Troiani, V., Grossman, M., & Wingfield, A. (2011). Hearing loss in older adults affects neural systems supporting speech comprehension. *The Journal of Neuroscience, 31*(35), 12638–12643. http://dx.doi.org/10.1523/JNEUROSCI.2559-11.2011.
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), 1653–1662. http://dx.doi.org/10.1121/1.3613698.
Rimmele, J. M., Zion Golumbic, E., Schröger, E., & Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex, 68*, 144–154. http://dx.doi.org/10.1016/j.cortex.2014.12.014.
Romero-Rivas, C., Martin, C. D., & Costa, A. (2015). Processing changes when listening to foreign-accented speech. *Frontiers in Human Neuroscience, 9*, 167. http://dx.doi.org/10.3389/fnhum.2015.00167.
Stowe, L. A., & Sabourin, L. (2005). Imaging the processing of a second language: Effects of maturation and prociency on the neural processes involved. International Review of Applied Linguistics in Language Teaching (IRAL), 43(4), 329–353. 10.1515/iral.2005.43.4.329.
Stringer, L. M. (2015). *Accent intelligibility across native and non-native accent pairings : Investigating links with electrophysiological measures of word recognition (Unpublished MPhil dissertation)*. London: University College London.
Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition, 18*(4), 380–393. http://dx.doi.org/10.3758/BF03197127.
Van Wijngaarden, S. J., Steeneken, H. J. M., Houtgast, T., van Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native talkers. *The Journal of the Acoustical Society of America, 112*(6), 3004–3013. http://dx.doi.org/10.1121/1.1512289.
Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50*(1), 1–25. http://dx.doi.org/10.1016/S0749-596X(03)00105-0.
Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., ... Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron, 77*(5), 980–991. http://dx.doi.org/10.1016/j.neuron.2012.12.037.