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

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