IJIELTR (2014) ISSN: 2156-5716
Volume 3, Number 2 pp.

COMPETITIVE LEXICAL ACTIVATION DURING ESL SPOKEN WORD RECOGNITION

Phillip Hamrick*
Kent State University, the USA

Nick B. Pandža
University of Maryland, the USA

ABSTRACT

Recognizing a spoken word involves the activation of multiple candidate words in memory and subsequent competition between those candidate words. This process of word candidate activation and competition are increased and protracted in second language learners, relative to native speakers (Weber & Cutler, 2004). While previous work has established that accurate second language word recognition is influenced by word frequency and neighborhood density, little work has focused on the frequencies and neighborhood densities of the lexical candidates that are activated and compete before accurate word recognition occurs. The present study investigated whether frequency and neighborhood density contribute to the sustained activation and competition of second language words by focusing on word recognition by Arabic learners of ESL using a variant of the gating task. Participants listened to fragments of 24 monosyllabic English words presented repeatedly, pseudo-randomized, with the amount of the word presented increasing in 70 millisecond increments. After hearing each fragment, participants indicated the word they believed they were hearing. The results revealed that, prior to successful word recognition, participants consistently produced words that were higher in frequency and from denser phonological neighborhoods than the target words they were hearing, suggesting that activation levels of lexical competitors are driven at least in part by frequency and lexical neighborhood density. Pedagogical implications are discussed.

KEYWORDS: Frequency, listening, neighborhood density, spoken word recognition,

INTRODUCTION

A key ingredient in effective listening is successful spoken word recognition. Native speakers rely on a perceptual system that has been fine-tuned to the phonetic details of their

*E-mail: philliphamrick@gmail.com

mother tongue. However, this fine-tuning process ultimately results in a speech processing system that is warped by experience (Kuhl, 2000) and maturation (Abrahamsson and Hyltenstam, 2009). As a result, spoken word recognition in a second language (L2) comes to rely on a speech processing system maximally adapted to the first language (L1) and fit for processing only some aspects of the L2. The ubiquitous nature of the resulting difficulties in L2 listening have spawned enormous theoretical (for overviews, see Cutler, 2012; Kuhl, 2000; Field, 2008; Flege, 2002) and empirical (for an overview, see Weber and Broersma, 2012) literatures. However, the perceptual processing of phonetic information in speech is just the first step in the process of effective listening.

A second step entails the activation and subsequent competition of candidate words in the mental lexicon. As with native speakers, L2 learners use phonetic information to activate lexical candidates and these candidates then compete with one another for recognition. However, recent research suggests that the processes of activation and competition are increased and protracted relative to native speakers (Broersma, 2012; Weber and Cutler, 2004). In short, more words become activated and then compete for longer to be recognized (e.g., Cutler and Broersma, 2005; Marian and Spivey, 2003; Schulpen, Dijkstra, Schriefers, and Hasper, 2003). The increase in lexical candidate activation and competition duration stems from a variety of factors, including (a) having lexical entries in more than one language, (b) spurious activation of extra lexical candidates due to inaccurate phonemic processing (e.g., Broersma and Cutler, 2008; Cutler, 2005; Cutler and Otake, 2004; Sebastian-Galles, Echeverria, and Bosch, 2005), and (c) reduced ability to inhibit incorrect competitors (e.g., Ruschemeyer, Nojack, and Limback, 2008; Weber, 2012; Weber and Cutler, 2004). Consequently, L1 and L2 word recognition differ in that activation and competition in L2 listeners likely demands more processing resources and results in slower, less accurate word recognition.

The increase in activation for L2 learners is due in part to the activation of more words from both the L1 and the L2. Since learners presumably have L2 words stored parasitically with their L1 counterparts (Sunderman and Kroll, 2006) it follows that L2 listening involves the activation of both L2 and L1 words. This finding has been reported widely in the literature. For example, in a set of cross-modal priming experiments Schulpen, Dijkstra, Schriefers, and Hasper (2003) found that when Dutch listeners heard English words (e.g., 'leaf'), they also activated their cross-lingual homophones from Dutch (e.g., *lief* – 'sweet'). This effect is not limited to whole-word cognates, but also occurs even when the L1 and L2 words only partially overlap in their onsets (Marian and Spivey, 2003; Weber and Cutler, 2004). However, cognate-based activation of L1 words may not cause too many problems for L2 learners. Instead, incongruency between L2 context and the meanings of L1 words may allow learners to rapidly discard irrelevant L1 meanings (Weber and Broersma, 2012). L1 interference effects aside, the activation of L2 words comes with its own problems for the learner. Cutler (2005) proposes several ways in which poor L2 phonemic processing can create lexical activation problems. First, L2 learners are confronted with pseudohomophones from minimal pairs in the L2. For example, Japanese listeners have difficulty discriminating between /ɹ/ and /l/ in English[1] because they are outlier allophones of a single underlying phonemic category in Japanese. Consequently, words like ‘right’ and ‘light’ activate one another for Japanese learners of English. A second source of phonemic confusion in L2 lexical activation stems from onset matching. For example, when native speakers of English are presented with the first syllable of ‘panda,’ they activate English words that have matching onsets, including ‘panel’ and ‘panic’ (Zwitserlood, 1989, as cited in Weber and Broersma, 2012). This finding forms one of the central tenets of Cohort Theory (Marslen-Wilson and Tyler, 1980). In Cohort Theory, the incremental processing of spoken words leads to the activation of all words that match the current input. For example, hearing a word-initial [p] activates all words beginning with the phoneme /p/ (e.g., ‘panel,’ ‘panic,’ ‘person’), but hearing the second allophone [æ] eliminates members of the cohort that do not match the input (e.g., ‘person’). However, for L2 learners with different vowel distinctions, the number of lexical items activated after hearing the first syllable of ‘panda’ grows. In addition to ‘panel’ and ‘panic,’ Dutch learners of English activate ‘pencil’ and ’penny’ (Weber and Cutler, 2004). Thus, if L2 learners have difficulty distinguishing phonemic contrasts, then they need not even hear a whole word in order to erroneously activate numerous extra competitors. Moreover, Cutler’s (2005) analysis of English lexical statistics suggests that word onset confusion may occur frequently.

Finally, learners are prone to activating L2 words that are phonologically similar to other words and non-words that are embedded in real words or across real word boundaries. In native speakers, for example, hearing the word ‘strange’ may temporarily activate words like ‘stray’ and ‘range.’ Unlike native speakers, the activation of embedded words is made worse by the spurious activation of embedded near-words. For example, an L2 learner who has difficulty distinguishing the English vowels /ɛ/ and /æ/ might activate the word ‘deaf’ when hearing the near-word ‘daf’ embedded in the real word ‘daffodil.’ Again, Cutler’s (2005) analysis indicates that activation due to word and near-word embedding is potentially very frequent in English.

To make matters worse for the L2 learner, once words are activated, the subsequent competition between them is prolonged. For example, Broersma and Cutler (2011) found that native speaking English participants temporarily activated an embedded word (e.g., ‘deaf’) while listening to a longer word (e.g., ‘definite’), but by the time they had heard a complete word, the embedded word ‘deaf’ was no longer active. In contrast, Dutch learners of English activated ‘deaf’ on the basis of the similar sounding near-word ‘daf’ in the word ‘daffodil.’ Crucially, even after the Dutch-speaking English learners had heard the entire word ‘daffodil’ the word ‘deaf’ was still activated.

Thus, L2 learners are confronted with increased and prolonged competition between lexical competitors during spoken word recognition. Competition is increased in part due to the activation of a larger number of lexical candidates and subsequent difficulties in inhibiting incorrect candidates (Broersma, 2012; Broersma and Cutler, 2011). However, these findings leave open several questions. Indeed, little is known about the specific factors that make lexical candidates become and stay activated in the face of increasing L2 phonetic information during listening. The present study focused on two factors that have been shown previously to influence successful L2 word recognition: frequency and neighborhood density[2] (e.g., Imai, Walley, and Flege, 2005). In particular, we investigated the effects of frequency and neighborhood density on L2 lexical activation and competition. To investigate this issue, the present study employed the gating paradigm (e.g., Grosjean, 1980) in an English as a second language (ESL) setting. The gating paradigm was chosen for two important reasons. First, it requires participants to generate lexical competitors. This means that L2 learners themselves produce lexical competitors, rather than having possible lexical competitors be produced by an experimenter. Second, it allows for the analysis of the characteristics of the lexical competitors that L2 learners produce (i.e., in terms of their frequencies and neighborhood densities).

The Present Study

In the present study, we employed the gating task for reasons similar to that of Wayland, Wingfield, and Goodglass (1989). In particular, we wanted to obtain more fine-grained data about the activated lexical competitors prior to accurate word recognition. Our study was novel in that it examined both frequency and neighborhood density characteristics of lexical competitors in L2 learners, which, to our knowledge, have not previously been examined. The choice of these characteristics was inspired by the fact that they are already known to influence the speed and accuracy of spoken word recognition in both the L1 (e.g., Marslen-Wilson, 1987) and the L2 (e.g., Imai et al., 2005).

As in previous gating studies, we were interested in how the properties of the target words themselves would influence word recognition. However, the focal point of our analysis was on the properties of the lexical competitors produced by L2 learners before they correctly recognized the words they were hearing. In particular, the following hypotheses were formed:

1) Following prior research demonstrating diminished word recognition abilities in adult L2 learners, we predict that participants will recognize words less often and at later gates than is often reported for native speakers.
2) Following Tyler (1984) and Wayland et al. (1995), prior to accurate word recognition participants will produce words that are higher in frequency than the target words they are hearing; however, in light of evidence that competition persists for longer in the L2 (e.g., Broersma, 2012; Broersma & Cutler, 2011), we predict that participants will continue to produce words that are high in frequency across all gates (cf. Wayland et al., 1995).
3) We predict that participants will produce words that are high in neighborhood density (neighborhood density > 20) across all gates.

METHOD

Participants

Twenty-two native speakers of Arabic learning English as a second language at a large midwestern university participated in the study across four sessions. Participants came from three intact listening courses in an ESL skills-based academic English program. All participants were placed within high-intermediate and advanced level courses. TOEFL scores were not available for individual students; however, the mean TOEFL scores for these levels are 450-500. All participants had spent between two months and two years in an English-speaking country. One participant was bilingual from childhood. In order to rule out knowledge of phonemic contrasts apart from Arabic and English, this participant was excluded from analysis.

Gating paradigm

In the gating task, participants are played words fragments, usually between 50 and 100 milliseconds in duration (called gates). After each gate of a word is played, participants are instructed to write down the word they believe to be hearing. This process continues with longer stretches of each word being played as the experiment goes on.

The gating task has been used extensively in the L1 spoken word recognition literature to investigate the amount of phonetic information required for people to identify a word (for an overview, see Grosjean, 1996). For shorter words in the L1, word recognition times in the gating task have been reported in the 200-350 millisecond range or less (e.g., Grosjean, 1980; Marslen-Wilson and Warren, 1994; Metsala, 1997; Wayland, Wingfield and Goodglass, 1989). A few studies have even investigated the words generated prior to word recognition. For example, Tyler (1984) and Wayland et al. (1989) investigated the words participants produced prior to recognition, which presumably represent a fraction of all of the words activated during word recognition. These words form a cohort (Marslen-Wilson and Tyler, 1980). Both studies reported that participants had an initial bias toward producing high frequency words at short gates, but Wayland et al. (1989) found that this bias rapidly diminished as gate length increased. In that study, participants produced words of a higher frequency than the targets 38% of the time at the 100 millisecond gate while only doing so 5% of the time after 300 milliseconds of the target word had been heard. In short, for those native speakers, the early cohort was dominated by high frequency words but only until enough phonetic information had been heard to reduce the cohort to lower frequency words that were more similar to the target word.

Materials

Stimuli

Target and filler words conformed to several constraints. For example, all words conformed to a CVC(C) structure. The target items were controlled for initial phoneme, frequency, word length (as measured by number of gates), and were divided into high and low neighborhood density groups based on Coltheart's N-statistic (Coltheart, Davelaar, Jonasson and Besner, 1977). Phonological neighborhood density is a similarity metric, representing the number of words with a phonological relationship that, in the case of Coltheart's N-statistic, only differ from the target by replacing only one letter of a that target, keeping the order constant.

Lexical statistics for the target words were extracted from the Washington University Hearing and Speech Lab Neighborhood Database (based on the Hoosier Mental Lexicon Database; Nusbaum, Pisoni and Davis, 1984). In order to assess[3] contributions from phonological knowledge, target words were selected to begin with either /b/ or /p/. Both consonants exist as phonemes in English, but they are allophones of a single phoneme, /b/, in Arabic. In Arabic, [p] occurs before voiceless consonants, and [b] elsewhere. Word frequency and length was not significantly different across word-initial phoneme ($F(1, 11) = .51$, $p = .49$; $F(1, 11) = .92$, $p = .36$) or neighborhood density ($F(1, 11) = .004$, $p = .95$; $F(1, 11) = 3.05$, $p = .11$). There was a significant difference in neighborhood density between high- and low-density words, $F(1, 11) = 21.31$, $p = .001$. The target stimuli and their lexical statistics are presented in Table 1 (see Appendix 1).

Given that the consonants /b/ and /p/ are confusable to Arabic learners of English and that Modern Standard Arabic has three realized vowel categories (high front, high back, and low front, each with a short/long distinction), minimal pairs were avoided and vowels differ as much as possible given the limitations of the constraints described above. Differentiating the vowels as much as possible in the targets presented within each session reduced the likelihood of participants believing they were hearing the same word when in fact they were hearing different words. However, it is worth noting that we could not completely remove all minimal pairs from the stimuli, primarily because we aimed to keep all of the above constraints on frequency, density, and duration satisfied. Consequently, the high density /p/-initial words are more similar to one another than we would have otherwise preferred. Likewise, the minimal pairs patch/pitch, bell/bull, and pat/patch remained as stimuli. Crucially, both members of a minimal pair never appeared within the same session for a participant.

Stimulus Creation

All stimuli were recorded by a male native speaker of American English in a soundproofed laboratory using a Blue Yeti USB microphone recording into Audacity 2.0.4. Carrier sentences did not yield natural-sounding words for single-word presentation, so all words were recorded in isolation. During recording, each word was repeated three times in order to get the best quality production for each word. Praat and Audacity were used to clean the files of background noise and aberrant sounds that resulted from clipping the audio files.

To prepare the words for presentation in the gating paradigm, all words were sliced into gate-length units by extracting increasingly longer fragments of the words. All plosive-initial words included 50 milliseconds of closure before the burst. In order to allow for the delivery of enough phonetic information to discriminate the initial consonant (e.g., silence before the burst, pre-voicing, voice onset time, etc.), we opted for an initial gate duration of 140 milliseconds. Subsequent gates consisted of successively longer fragments of the word, with each gate increasing the amount of the word played by 70 milliseconds, e.g., 210, 280, 350 milliseconds and so on. The first and last 10 milliseconds of each gate were ramped up and down in volume to avoid any sound artifacts coming from the abrupt starting or stopping of high amplitude sound.

Procedure

To reduce the effects of fatigue, the experiment was divided into four sessions, each taking place during normal class time. Each intact class used a different order of sessions. All sessions within a class level transpired within a single week. In all classes, the gating task was proctored by the first author and three ESL teachers. All students were told that they would hear a mixture of words and word fragments and that their task was to write down the first word that came to mind after hearing each word or word fragment.[4] Before the first session, participants completed a practice session to orient themselves to the task. The practice session consisted of the presentation of two filler words that were not reused again during the study.

In each session, participants listened to fragments of the target and filler words. Stimuli were delivered via a timed PowerPoint presentation. The PowerPoint slides displayed only a black screen with the trial number printed in white Calibri font size 96. The number indicated which answer sheet item they should be filling out for that word fragment so they would not lose their place. Each trial during the gating task began with a bell tone followed by two seconds of silence and then the gate (140, 210, 280, 350, 420, 470, 560, 630, 700, or 770 milliseconds in duration), followed by 10 seconds of silence for the participants to write down the word they thought they were hearing and a confidence rating.[5] Although it has been suggested that five to eight seconds is sufficient time for participants to respond and indicate their confidence (Jiang, 2012), our pilot research suggested that six seconds was too short for non-native speakers to do both for the majority of trials.

Session	/b/-initial targets	/p/-initial targets	Fillers
Practice	N/A	N/A	mark (7), sort (7)
1	bell (6), bid (5)	pat (6)	hard (7), yard (7), greet (6)
2	boss (9)	patch (8), push (8)	farm (8), host (8), swan (8)
3	bull (6), boot (4)	pig (5)	fork (7), lord (7), star (7)
4	bath (8)	pause (10), pitch (7)	feast(9), form (8), horn (8)

Table 2. Distribution of Stimuli over the Testing Sessions in the Gating Task
The length of each word in gates is reported in parentheses.

Each session took approximately 10 minutes. To avoid response perseveration (Walley, Michela and Wood, 1995), stimulus presentation order was pseudo-randomized by gate length so that the first gate of 140ms was heard for all words, followed by the second gate of 210ms for all words, and so on, so as to eliminate a target from immediately preceding or following another target. The order of presentation of fragments was also different at each gate length so that participants did not hear the word fragments in the same order. Finally, none of the three targets in each session contained the same vowel.

Data Analysis and Coding

All of the non-English word responses that participants produced were removed from the analysis, including misspelled near-words (e.g., 'bause' instead of 'pause'), non-words (e.g., 'paazze'), single letters, or no responses. We followed this procedure to avoid over- or under-interpreting the knowledge and word recognition abilities of the participants. However, given that some words produced were likely not ones the participants know (e.g., 'pall') and given that Arabic ESL learners have well-documented difficulties with English spelling (e.g., Saigh and Schmitt, 2011), this likely leads to a conservative interpretation of the word recognition process. The discarding of all incorrectly spelled responses resulted in the loss of 23.66% of the original 1712 data observations, leaving 1307 responses for analysis.

RESULTS AND DISCUSSIONS

Hypothesis 1

Following prior research that has demonstrated diminished word recognition abilities in adult L2 learners, we predicted that participants would recognize words less often and at later gates than is often reported for native speakers. To address this hypothesis, we first determined the average number of gates it took for participants to produce the correct answer (the word isolation point, Grosjean, 1980). Participants responded incorrectly on the majority of trials across gates (N = 1115, 85.3%), so the accuracy analyses focused only on 14.7% (N = 192) of the data (i.e., participants' correct responses). For this small proportion of words that participants recognized correctly, the mean isolation point was between gates 5 and 6 (420-490 ms). To compare ESL learners with native-speakers, we took a conservative estimate of the amount of time it took adult native speakers to recognize short words (350 ms) and compared it to ESL learners' mean isolation points. A one-sample t-test revealed that the isolation points for ESL learners (M = 440.38, SD = 107.60) were significantly later than our adult native speaker estimate (M = 350), $t(78) = 7.47$, $p < .001$. Participants required substantially more phonetic information before correctly identifying the word they were hearing, much longer than a conservative estimation for native speakers in shorter words (cf. Grosjean, 1980; Metsala, 1997; Wayland et al., 1989).

These findings confirm our first hypothesis. The late isolation points for our ESL learners combined with their low accuracy demonstrate that L2 learners are slower and make more errors in recognizing words than native speakers, consistent with previous research (for an overview, see Cutler, 2012). When word fragments are presented in the gating task, native speakers have been reported to identify similar short words correctly within 200-350 milliseconds of the onset (e.g., Grosjean, 1980; Metsala, 1997; Wayland et al., 1989). In comparison, ESL learners in our study recognized words far less often (14.7% of total gates), and when they did recognize the word it was not until 420-490 milliseconds after the onset.

Hypothesis 2

We also predicted that, prior to recognition, participants would produce words higher in frequency than target words and subsequently continue to produce words that were higher in frequency than targets across all gates. For these analyses, we only analyzed the words participants produced prior to recognition (henceforth, P2R). Because the frequencies of the P2R and target words were not normally-distributed, Ds > .20, ps < .05, they were first log-transformed to achieve normality. Then, the mean log-transformed frequencies of P2R and target words were submitted to an independent samples t-test. The mean frequency of P2R words (M = 1.82, SD = .18) was significantly larger than the mean frequency of target words (M = 1.28, SD = .16), $t(20) = 13.71$, $p < .001$. This result indicates that, overall, participants preferentially produced words of a higher frequency than the target words they were hearing, consistent with our second hypothesis (Figure 1).[6]

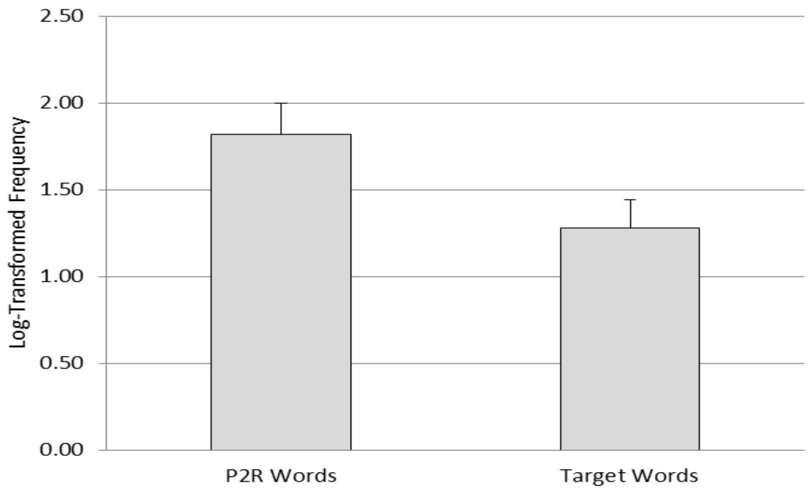

Figure 1. Mean Log-transformed Frequencies of P2R and Target Words overall

Although the frequencies of P2R words were higher than the frequencies of target words overall, it is possible that preferences for high frequency responses in the gating task were limited to just the first few gates (Wayland et al., 1989). To address the other half of our second hypothesis, that learners would produce high frequency words across gates, a 2 X 9 ANOVA was conducted on the log-transformed frequencies of P2R words with Gate as the within-subjects variable.[7] The ANOVA revealed no significant effect of Gate, $F(8, 112) = 1.11$, $p = .36$. This result suggests that participants were producing P2R words with consistently high frequencies. A visual inspection of the frequencies of P2R and target words across gates in Figure 2 suggests that the frequencies of P2R words are consistently higher than the frequencies of target words.

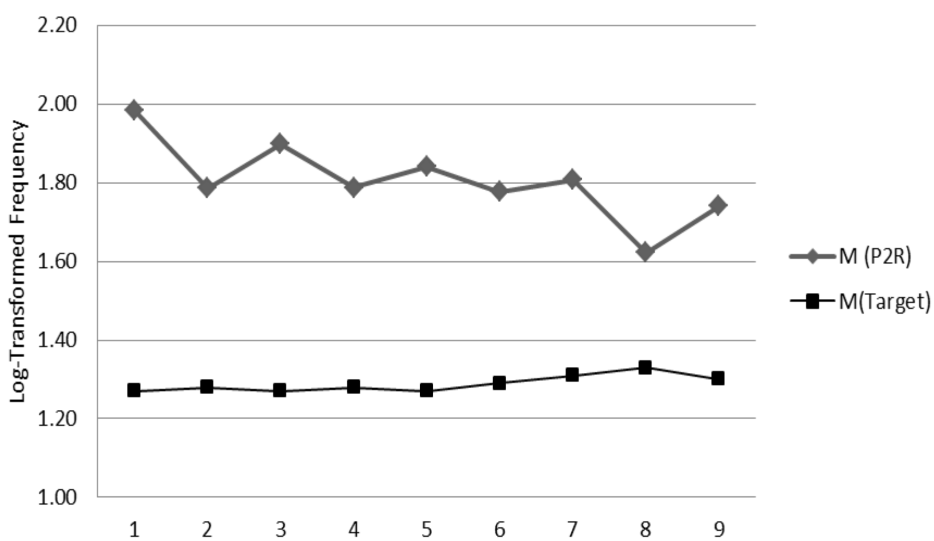

Figure 2. Mean Log-transformed Frequencies of P2R Words and Target Words across each Gate

Overall, our second hypothesis was confirmed. Contrary to previous findings with native speaking adults (e.g., Wayland et al., 1989), the production of high frequency competitors prior to correct word recognition was not sharply reduced after the first few gates. Previous research indicated that native speakers' preferences for high frequency words diminish rapidly after 100-250 ms of the target word has been heard (Tyler, 1984; Wayland et al., 1989). That finding is consistent with Cohort Theory and its later instantiations (Gaskell and Marslen-Wilson, 1997; Marslen-Wilson and Tyler, 1980), in which initially high lexical activation for high frequency words is followed by effective inhibition of incorrect competitors, regardless of the frequency, as more phonetic information is processed from the input. This pattern in native speakers did not hold in our ESL learner population. Instead, prior to recognition, participants produced words from their activated cohorts that were higher frequency than the target words they were hearing at all gate lengths. This suggests that high frequency may cause lexical competitors to remain competitive within an L2 cohort even after most or all of a word has been heard. If this is true, then the findings by Broersma (2012) and Broersma and Cutler (2011) that lexical activation decays more slowly in the L2, drawing out the competition process, should not be limited to activation of low frequency English words (e.g., activating 'deaf' when 'daffodil' is heard). Indeed, one might predict that the slow decline in activation of lower-frequency lexical competitors like 'deaf' may still be faster than the reduction in activation of higher-frequency lexical competitors like 'big.' However, more research is needed to address this prediction.

Hypothesis 3

We also predicted that participants would produce words that are high in neighborhood density (which we defined a priori to be greater than 20) across all gates. In other words, we predicted that P2R words would be consistently drawn from dense phonological neighborhoods. We separately analyzed P2R words that were produced in response to high density targets (P2R High words) and P2R words that were produced in response to low density targets (P2R Low words). These analyses only examined P2R responses from the first seven gates because there were too few incorrect responses (fewer than 3) to high density targets in gates 8 through 10.

The mean neighborhood density for P2R Low words (M = 24.03, SD = 1.91) was significantly larger than the neighborhood density for target words from a low density neighborhood (M = 12.33, SD = .66), $t(20)$ = 27.96, p < .001. The mean neighborhood density for P2R High words (M = 26.39, SD = 2.75) was not significantly larger than that of target words from a high density neighborhood (M = 25.81, SD = 2.43), $t(20)$ = .97, p = .34, CI 95% [-.66, 1.83]. Finally, a paired-samples t-test revealed that P2R Low words were significantly lower in density than P2R High words, $t(20)$ = 4.13, p = .001. Overall, these results suggest that participants generally produced P2R words from high density neighborhoods (by our own a priori definition of high density as being greater than 20), regardless of whether they were hearing target words from a low or high density neighborhood. They also produced higher-density words when hearing high density targets than when hearing low density targets.

We also wanted to determine whether participants produced words of differing densities across gates. A 2 X 7 repeated measures ANOVA with target word neighborhood density (High, Low) and Gate (1-7) as within-subjects variables was conducted on the neighborhood densities of P2R words. The results revealed no significant effect of Target word neighborhood density, $F(1, 7) = 1.96$, $p = .20$, a significant effect of Gate, $F(6, 42) = 2.73$, $p = .02$, $\eta_p^2 = .28$, and a significant Target*Gate interaction (Greenhouse-Geisser corrected), $F(2, 18.57) = 11.15$, $p < .001$, $\eta_p^2 = .61$. Considering these results in light of Figure 3, it seems that the amount of phonetic information participants had available at different gates influenced the neighborhood density of the words they produced. Finally, inspection of Figure 4 also reveals that participants produced P2R High words that were actually higher in density than the target words on four out of seven gates. Although the P2R words participants produced in response to low density target words in gate seven was below the hypothesized threshold of high density (18.26 as opposed to greater than 20), the overall pattern of results supports our third hypothesis: prior to recognition, participants produced words that were generally from high density phonological neighborhoods.

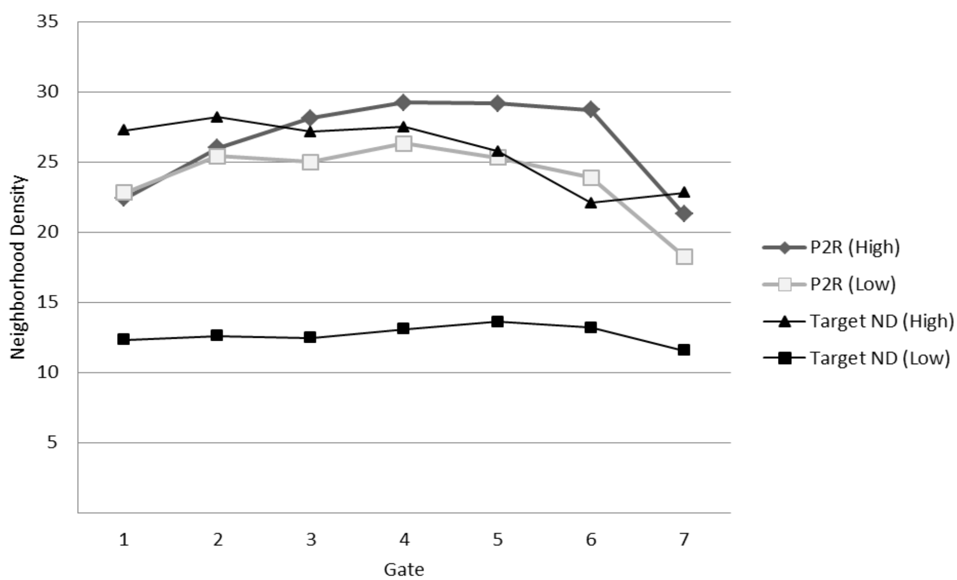

Figure 3. Mean Neighborhood Densities for P2R Words and Target words.

P2R (High) refers to words produced in response to high density Targets, while P2R (Low) refers to words produced in response to low density Targets. Our third hypothesis was mostly supported. Participants preferred to produce words from higher density neighborhoods, and these neighborhoods were often, but not always, denser than that of the target words. This result is consistent with the Neighborhood Activation Model (NAM; Luce and Pisoni, 1998), which posits that words are recognized based on their similarity to phonological information stored in the lexicon. That is, when listening to words, nonwords, or fragments of words, adults activate a number of lexical candidates based on their phonological similarity to stored words and sublexical units. Having a denser phonological neighborhood implies the activation of more lexical or sublexical candidates, increasing activation and slowing word recognition. Our results show that prior to recognition ESL learners activated dense neighborhoods and were indeed slower to recognize words. Moreover, the results are consistent with overwhelming evidence that L2 word recognition involves substantial competition from words that are phonologically similar but that may not be activated in native speakers. For example, as late as gate seven[8] (560 milliseconds) participants were still producing competitors like 'past' when hearing target words like 'bath.' By comparison, the difference in the first two phonemes of these words and in co-articulatory cues (Marslen-Wilson and Warren, 1994) would be more than enough to prevent activation of 'past' when hearing 'bath' in native speakers of English. In short, it is likely that when participants heard fragments of the target word, they activated phonologically similar words and neighbors that native speakers probably would not activate. If the activated cohort had a higher overall neighborhood density than the target word, or if specific words in the activated cohort were very high in frequency, then participants would be more likely to produce the incorrect word, slowing the recognition process (Luce and Pisoni, 1998).

Taken together, the results are consistent with the idea that the sustained activation of L2 words that are high frequency and that come from dense phonological neighborhoods contributes to more lexical activation and prolonged competition effects during spoken word recognition (Broersma and Cutler, 2011).

PEDAGOGICAL IMPLICATIONS

The results of this and prior research suggest that L2 learners face substantially more lexical activation and competition than native speakers. Although, in principle, the mechanisms of activation and competition appear to be qualitatively the same in native and non-native speakers, they are (at least) quantitatively different. In terms of English language teaching, this means that ESL/EFL instructors need to be aware that their learners are likely to rely heavily on top-down and bottom-up processing during listening. Top-down processing—the use of prior knowledge—will result in L2 learners relying on lexical information (i.e., frequency and neighborhood density) during listening and world knowledge, while bottom-up processing involves L2 learners parsing the auditory input one unit (e.g., phoneme, syllable) at a time.

Now, since L2 learners' bottom-up phonetic parsing of the input is likely to be insufficient (Kuhl, 2000), then they may become over-reliant on top-down lexical information. There are at least two orthogonal teaching strategies that teachers may use to moderate learner over-reliance on lexical information. First, if teachers want learners to modify how they use lexical information in listening, then they could use pre-listening tasks in order to help learners activate other prior knowledge beyond lexical frequency and neighborhood density. For example, L2 learners may benefit from pre-listening activities that ask them to predict what comes next in a listening activity based on word lists, pictures, or other supplementary information (Brown, 2011, p. 31). This should enable L2 learners to use other top-down knowledge to give them additional listening cues beyond frequency and phonological similarity.

Second, teachers may want to use explicit instruction and extended practice on L2 phonology and phonotactics to promote better bottom-up processing of the listening material to begin with. This may reduce some of the lexical competition effects caused by ineffective phonological parsing. Evidence for this conclusion comes from a study by Al-jasser (2008), who demonstrated that explicit instruction on English phonotactics may aid learners in more efficient phonological parsing. Teachers can combine such explicit instruction with tasks that encourage practice at bottom-up processing, such as dictations of reduced forms and minimal pair discrimination (Brown, 2011, p. 49). Taken together, this mix of explicit instruction, practice, and top-down/bottom-up processing should give the L2 learner more robust listening abilities. Ideally, these pedagogical strategies would reduce L2 learners' (over-)dependence on lexical frequency and phonological similarity information like that seen in the present study.

The development of a more robust, rich vocabulary also cannot be underestimated. Given that there is substantial evidence that phonetic information stored in the lexicon becomes tuned with experience and a larger vocabulary (Perceptual Assimilation Model: e.g., Best and McRoberts, 2003; Bundgaard-Nielsen, Best, and Tyler, 2011; Lexical Restructuring Hypothesis: Storkel, 2002), it follows that students with larger L2 vocabularies may be able to more effectively process the sound input and reduce the number of spurious lexical competitors activated during listening. Likewise, this suggests that the current resurgence in the popularity of teaching minimal pairs may be useful for aiding bottom-up listening (Brown, 2011). Moreover, given that L2 learners also mishear near-words, it may be useful for teachers to do practice activities with both minimal pairs and near-minimal pairs to reduce spurious lexical activation.

Finally, instructors should consider the present results in light of other research suggesting that the simultaneous presentation of spoken and written forms of words provides L2 learners with a way of creating separate lexical representations for words that are phonologically similar (e.g., Cutler and Weber, 2007; Escudero, Hayes-Harb and Mitterer, 2008). If L2 learners are provided with a reliable mapping between orthographic and phonological forms, they may be more likely to develop accurate lexical representations, which can subsequently aid in spoken word recognition. However, if teachers undertake such a task they may want to spend more time on those words that are likely to produce spurious competitors during listening. For Arabic-speaking ESL/EFL learners, this means focusing on words with /b/ and /p/ and vowel contrasts that are difficulty for Arabic speakers.

LIMITATIONS

The present study has several limitations that must be addressed to obtain a better understanding of the factors that drive L2 lexical competition. First, the gating task provided participants with words and word fragments without any sentential or discourse context, both of which may influence recognition in the gating task (Tyler, 1984). Consequently, their high error rates may have been an artifact of presenting the words in isolation. Second, the gating task may not truly reflect on-line word processing (although see Tyler and Wessels, 1985, for evidence that the gating task produces similar results to on-line tasks). If the gating task is not truly an on-line measure, then the reported effects may derive from strategic, decision-making heuristics, rather than reflect the activation levels of words in the mind. It is possible that participants entertained several lexical competitors before producing an answer. Despite being told to write the first word that came to mind, participants may have vacillated between competitors. If so, they may have relied on a feeling of correctness, familiarity, and/or plausibility of being correct to make their decision. For example, a participant might have activated and contemplated both ‘back’ and ‘batch,’ but chose ‘back’ due to higher levels of familiarity, not knowing the correct spelling of ‘batch,’ etc.

Another limitation of the study is that the high frequency and high density characteristics of the lexical competitors produced during the gating task do not necessarily apply to all lexical competitors activated—but not produced—during spoken word recognition. For example, some lexical candidates should be activated purely based on their ability to match the fine-grained details in the phonetic input (Marslen-Wilson and Warren, 1994), regardless of their frequency or neighborhood density. Consistent with previous research, participants did not exclusively produce high frequency words prior to recognition (cf. Wayland et al., 1989); however, such low frequency words were rare. Consequently, we can only claim that these high frequency, high density characteristics apply to the lexical competitors that were produced in the gating task prior to correct word recognition.

One possible solution to the above limitations is to use more robust on-line tasks (e.g., lexical decision tasks, eye-tracking) to measure incremental aspects of spoken word recognition. While these studies are unable to produce the corpus of word recognition data that a gating task can produce, they are less prone to contamination from conscious reasoning or decision making. A wealth of studies on L2 spoken word recognition have already been conducted using these measures (for an overview, see Weber and Broersma, 2012); however, to our knowledge none have used words that participants themselves generate as competitors for the stimuli of the on-line experiments. The advantage of the gating task is that participants provide the words that could be used in subsequent reaction time or eye-tracking studies. We advocate such a method, as it might disentangle which on-line aspects of word recognition the gating task might truly capture.

Finally, this study was novel in that it conducted the gating task on intact ESL classes. While this gave the study some ecological validity, it also introduced possible problems that could be avoided in a more controlled laboratory environment. For example, by conducting the gating task in a lab, individual participants could be continuously monitored for the production of correctly spelled words. Likewise, the experimenter could crosscheck participants' written responses with their spoken responses to make sure the written word was in fact the word the participant intended.

CONCLUSIONS

This study uniquely targeted the words L2 learners activate prior to accurate word recognition by having the participants produce the words in a gating task. As with native speakers, frequency and neighborhood density appeared to contribute to activation and competition in the L2. At the same time, the results indicated that lexical activation and competition is greater and protracted in L2 spoken word recognition, consistent with prior research (e.g., Broersma and Cutler, 2011). This suggests a quantitative, but not necessarily qualitative, distinction between L2 word recognition and L1 word recognition. Adult L1 speakers also activate incorrect lexical candidates during spoken word recognition and factors such as frequency and phonological similarity influence lexical competition. However, adult L1 speakers rapidly reduce the number of lexical competitors regardless of their frequency (Wayland et al., 1989), whereas our ESL learners did not. Therefore, although the processes of lexical competition and activation are similar in the L1 and L2, the process of reducing the activation and number of lexical competitors in the L2 is slower and less efficient. This likely contributes to inefficient L2 listening. More research is needed into the development of teaching methods and materials that help L2 learners avoid some of the negative effects of such enhanced and protracted lexical competition during listening.

NOTES

1. In the Midwestern United States where this study was conducted, the primary pronunciation of the rhotic consonant is typically /ɹ/ or the retroflex /ɻ/. However, we acknowledge a variety of pronunciation possibilities of this consonant in English.
2. Frequency refers to how often a word occurs in actual language use. Typically, frequency is estimated from corpora, and so may not reflect how often a given speaker will have heard or used the word. For our purposes, neighborhood density refers to the number of words that differ from a given word by changing a single phoneme within a word, for example 'bat' and 'cab' are neighbors of 'cat.'
3. The results of these analyses are the subject of a separate paper (Pandža and Hamrick, 2014).
4. All participants also completed a modified version of the Language History Questionnaire (Li, Sepanski, & Zhao, 2006), the results of which are not reported here.
5. Confidence rating analyses are not reported here; however, it is worth noting that confidence was positively correlated with gate number.
6. It should be noted that initial exploratory analyses showed no significant differences in the frequencies of P2R words produced in response to high vs. low density words or /b/ vs. /p/ initial words, ps > .05.
7. Gate 10 was not included in these analyses because the majority of the participants did not know the single target word occurring at that gate.
8. It is possible, though unlikely, that the participants did not know words like "bath," since they were from high-intermediate and advanced ESL classes.

ACKNOWLEDGEMENTS

We would like to thank our ESL participants and their teachers: Kelli Ryan, Katie McClaren, Randi Cline, and Mishel Sanches. We would also like to thank the anonymous reviewers for their helpful comments on earlier versions of this manuscript. We declare no conflict of interest.

REFERENCES

Abrahamsson, N., and Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59,* 249-306.

Al-Jasser, F. (2008). The effect of teaching English phonotactics on the lexical segmentation of English as a foreign language. *System*, *36*, 94-106.

Best, C. T., and McRoberts, G. W. (2003). Infant perception of nonnative contrasts that adults assimilate in different ways. *Language and Speech*, *46*, 183-216.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341-345.

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes, 27,* 1205-1224.

Broersma, M., and Cutler, A. (2008). Phantom word recognition in L2. *System*, *36*, 22-34.

Broersma, M., and Cutler, A. (2011). Competition dynamincs of second-language listening. *Quarterly Journal of Experimental Psychology*, *64,*74-95.

Brown, S. (2011). *Listening myths: Applying second language research to classroom teaching*. Ann Arbor, MI: University of Michigan Press.

Bijeljac-Babic, R., Biardeau, A., and Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Memory & Cognition, 25,* 447-457.

Brysbaert, M. (1998). Word recognition in bilinguals: evidence against the existence of two separate lexicons. *Psychologica Belgia, 38,* 163– 175.

Bundgaard-Nielsen, R. L., Best, C. T., and Tyler, M. D. (2011). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, *32*, 51-67.

Coltheart, M., Davelaar, E., Jonasson, J.T., and Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). London: Academic Press.

Cutler, A. (2005). *The lexical statistics of word recognition problems caused by L2 phonetic confusion*. Proceedings of the 9th European Conference on Speech Communication and Technology, Lisbon, September 2005.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.

Cutler, A., and Broersma, M. (2005). Phonetic precision in listening. In W. J. Hardcastle & J. M. Beck (Eds.), *A figure of speech: A festschrift for John Laver* (pp. 63-91). Mahwah, NJ: Erlbaum.

Cutler, A., and Otake, T. (2004). *Pseudo-homophony in non-native listening*. Paper presented to the 75th meeting of the Acoustical Society of America, New York.

Cutler, A., and Weber, A. (2007). Listening experience and phonetic-to-lexical mapping in L2. *ICPhS*, 16, 43-48.

de Groot, A. M. B. (2011). *Language and cognition in bilinguals and multilinguals*. New York: Psychology Press.

Escudero, P., Hayes-Harb, R., and Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*, 345-360.

Field, J. (2008). Revising segmentation hypotheses in first and second language listening. *System*, *36,* 35-51.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, and Rohde, A. (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 217-244). Trier: Wissenschaftlicher Verlag.

Francis, W. S. (1999). Analogical transfer of problem solutions within and between languages in Spanish–English bilinguals. *Journal of Memory and Language, 40,* 301–329.

Gaskell, M. G., and Marslen-Wilson, W. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12,* 613-656.

Gollan, T. H., and Kroll, J. F. (2001). Bilingual lexical access. In Rapp, B. (Ed.), *The Handbook of cognitive neuropsychology: What deficits reveal about the human mind.* (pp. 321-345). Philadelphia, PA: Psychology Press.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, *28*, 267-283.

Grosjean, F. (1996). Gating. *Language and Cognitive Processes*, *11*, 597-604.

Imai, S., Walley, A., and Flege, J. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *Journal of the Acoustical Society of America*, *117*, 896-907.

Jiang, N. (2012). *Conducting reaction time research in second language studies*. New York: Routledge.

Kuhl, P. (2000). A new view of language acquisition. *Proceedings of the National Academy of the Sciences, 97,* 11850-11857.

Li, P., Sepanski, S., and Zhao, X. (2006). Language history questionnaire: A Web-based interface for bilingual research. *Behavior Research Methods, 38,* 202-210.

Luce, P., and Pisoni, D. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19,* 1-36.

Marian, V., and Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition, 6*, 97-115.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition, 25,* 71-102.

Marslen-Wilson, W., and Tyler, L. (1980). The temporal structure of spoken language understanding. *Cognition, 8,* 1-71.

Marslen-Wilson, W., and Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, *101*, 653-675.

Metsala, J. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory & Cognition*, *25*, 47-56.

Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (1984). Sizing up the Hoosier mental lexicon. In Research on Spoken Language Processing Report No. 10 (pp. 357-376). Bloomington, IN: Speech Research Laboratory, Indiana University.

Pandža, N. B., and Hamrick, P. (2014). Interactions between frequency and phonological similarity make lexical competitors competitive. Paper presented at the *Georgetown University Roundtable of Languages and Linguistics* (GURT). Washington, DC.

Rüschemeyer, S., Nojack, A., and Limbach, M. (2008). A mouse with a roof? Effects of phonological neighbors on processing of words in sentences in a nonnative language. *Brain and Language, 104*, 113-121.

Siagh, K., and Schmitt, N. (2011). Difficulties with vocabulary word form: The case of Arabic ESL learners. *System*, *40*, 24-36.

Schulpen, B., Dijkstra, A., Schriefers, H., and Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 1155-1178.

Sebastián-Gallés, N., Echeverría, S., and Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*, 240-255.

Storkel, H. L. (2002). Restructuring of similarity neighborhoods in the developing mental lexicon. *Journal of Child Language*, *29*, 251-274.

Sunderman, G., and Kroll, J. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition, 28,* 387-422.

Tyler, L. (1984). The structure of the initial cohort: Evidence from gating. *Perception and Psychophysics*, *36*, 417-427.

Tyler, L., and Wessels, J. (1985). Quantifying contextual contributions to word-recognition processes. *Perception and Psychophysics*, *34*, 409-420.

Walley, A., Michela, V., and Wood, D. (1995). The gating paradigm: Effects of presentation format on spoken word recognition by children and adults. *Perception and Psychophysics*, *57*, 343-351.

Washington University in St. Louis Speech and Hearing Lab Neighborhood Database as a web-based implementation of the 20,000 word Hoosier Mental Lexicon (2013, July; accessed). neighborhoodsearch.wustl.edu/

Wayland, S., Wingfield, A., and Goodglass, H. (1989). Recognition of isolated words: The dynamics of cohort reduction. *Applied Psycholinguistics*, *10*, 475-487.

Weber, A., and Broersma, M. (2012). Spoken word recognition in second language acquisition. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Bognor Regis: Wiley-Blackwell. DOI: 10.1002/9781405198431.

Weber, A., and Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50,* 1-25.

Wickens (2001). *Elementary signal detection theory.* Oxford: Oxford University Press.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32*, 25-64.

Appendix 1. Table 1. Lexical Statistics for the Target Stimuli in the Gating Task