Research Article

The Downside of Greater Lexical Influences: Selectively Poorer Speech Perception in Noise

Boji P. W. Lam,^a Zilong Xie,^a Rachel Tessmer,^a and Bharath Chandrasekaran^{a,b,c,d,e}

Purpose: Although lexical information influences phoneme perception, the extent to which reliance on lexical information enhances speech processing in challenging listening environments is unclear. We examined the extent to which individual differences in lexical influences on phonemic processing impact speech processing in maskers containing varying degrees of linguistic information (2-talker babble or pink noise).

Method: Twenty-nine monolingual English speakers were instructed to ignore the lexical status of spoken syllables (e.g., *gift* vs. *kift*) and to only categorize the initial phonemes (/g/ vs. /k/). The same participants then performed speech recognition tasks in the presence of 2-talker babble or pink noise in audio-only and audiovisual conditions.

Results: Individuals who demonstrated greater lexical influences on phonemic processing experienced greater speech processing difficulties in 2-talker babble than in pink noise. These selective difficulties were present across audio-only and audiovisual conditions.

Conclusion: Individuals with greater reliance on lexical processes during speech perception exhibit impaired speech recognition in listening conditions in which competing talkers introduce audible linguistic interferences. Future studies should examine the locus of lexical influences/interferences on phonemic processing and speech-in-speech processing.

Daily communication seldom takes place in the absence of background noise. For instance, listeners are often required to attend selectively to a talker when there are competing talkers in the background. This well-known “cocktail-party” phenomenon (Cherry, 1953) presents a formidable computational challenge to the brain on a daily basis (Mesgarani & Chang, 2012; Xie, Maddox, Knopik, McGeary, & Chandrasekaran, 2015). Previous studies examining the dynamics of speech-in-speech processing have focused on masker properties, such as similarities in voice features between target and masking talkers (e.g.,

Brouwer, Van Engen, Calandruccio, & Bradlow, 2012; Brungart, Simpson, Ericson, & Scott, 2001), the impact of increasing the number of masking talkers (e.g., Freyman, Balakrishnan, & Helfer, 2004; Simpson & Cooke, 2005), and the spatial separation of target from masker (e.g., Freyman, Balakrishnan, & Helfer, 2001; Li, Daneman, Qi, & Schneider, 2004). Recent studies have documented significant individual differences in speech-in-speech processing ability that relate to listener characteristics, such as working memory or neurocognitive profiles, for cocktail-party-like scenarios (e.g., Xie et al., 2015; Zekveld, Rudner, Johnsrude, & Rönnberg, 2013). These studies have collectively indicated that individual differences in cognitive processing may relate to speech processing ability in conditions that require listeners to ignore irrelevant linguistic information in the background (e.g., Anderson, White-Schwoch, Parbery-Clark, & Kraus, 2013; Chandrasekaran, Van Engen, Xie, Beevers, & Maddox, 2015; Xie et al., 2015; Zekveld, Rudner, Johnsrude, & Rönnberg, 2013). Despite the fact that heightened lexical interferences from competing talkers characterize speech-in-speech listening conditions, the extent to which individual differences in susceptibility to lexical influences impact speech-in-speech processing has not been systematically examined, and will be a focus of the current study.

^aDepartment of Communication Sciences & Disorders, Moody College of Communication, The University of Texas at Austin

^bDepartment of Psychology, College of Liberal Arts, The University of Texas at Austin

^cInstitute for Mental Health Research, College of Liberal Arts, The University of Texas at Austin

^dDepartment of Linguistics, College of Liberal Arts, The University of Texas at Austin

^eInstitute for Neuroscience, The University of Texas at Austin

Correspondence to Boji P. W. Lam: bojilam0705@gmail.com

Editor: Nancy Tye-Murray

Associate Editor: Mitchell Sommers

Received April 2, 2016

Revision received September 15, 2016

Accepted December 15, 2016

https://doi.org/10.1044/2017_JSLHR-H-16-0133

Disclosure: The authors have declared that no competing interests existed at the time of publication.

Lexical Influences on Speech Processing: The Ganong Effect

Lexical influences on speech processing are pervasive and widely discussed (Ganong, 1980; for a more recent review, see Samuel, 2011). A well-studied example of this phenomenon, the Ganong effect (Ganong, 1980), demonstrates the impact of lexical knowledge on phoneme identification. In a typical Ganong paradigm, listeners are instructed to identify the initial phoneme presented in a syllabic context. The phoneme varies on an acoustical continuum, with the ends of the continuum either forming a word or a nonword. For example, in the original experiment listeners were presented with two word contexts of /*gift*/ (“gift”) and /*kɪs*/ (“kiss”) and instructed to identify the initial phoneme. The voice onset time (VOT) of the initial phoneme /*g*/ or /*k*/ was manipulated in small increments into a VOT continuum. Listeners reported a bias for /*g*/ responses for stimuli embedded in a *gift*–*kift* continuum (real word vs. nonword) and a bias for /*k*/ responses embedded in a *giss*–*kiss* continuum (nonword vs. real word). Even though listeners were specifically instructed to focus on the initial phoneme, they often responded to an identical token differently as a function of the word context (i.e., *-ift* vs. *-iss*; Ganong, 1980). In addition, longer words (e.g., “distinguish,” /*distɪŋɡwɪʃ*/), for which fewer alternate lexical candidates are available, elicit a stronger Ganong effect on final phoneme identification (e.g., /*s*/ vs. /*f*/) than shorter words (e.g., “fish,” /*fɪʃ*/; Pitt & Samuel, 2006). This finding indicates that lexical words that are more strongly activated would be more difficult to suppress and thus induce greater susceptibility to lexical influences. On the basis of these findings, the Ganong effect is hypothesized to index difficulty in ignoring the lexical influence of the carrier syllable as information that is irrelevant to the task at hand (i.e., phonemic identification). The amount of contextual bias an individual demonstrates in labeling the target phoneme is used to index the extent of lexical influence on phonetic processing (Pitt & Samuel, 2006). Therefore, the Ganong effect has been well studied as a measure of the interplay between lexical activation and phonetic processing (Mattys & Wiget, 2011). The current study aims to examine the extent to which susceptibility to lexical influences, as indexed by the Ganong effect, is associated with speech processing abilities when irrelevant linguistic information is present due to interfering competing talkers.

Susceptibility to Lexical Influences May Relate to Speech-In-Speech Processing Challenges

Several lines of evidence indicate that the Ganong effect reflects the inability to suppress lexical activation. For example, older adults demonstrate a more robust Ganong effect relative to younger adults (Mattys & Scharenborg, 2014), and they are also more prone to contextual biases and “false hearing” (Rogers, Jacoby, & Sommers, 2012). Comparing across studies on children with specific language impairment (SLI) also reveals similar associations.

Children with SLI are less effective in ignoring lexical competitors (McMurray, Munson, & Tomblin, 2014; McMurray, Samelson, Lee, & Tomblin, 2010), and more likely to vacillate on their responses when exposed to word fragments with multiple alternative continuations (Mainela-Arnold, Evans, & Coady, 2008). Children with SLI also display greater lexical influences on speech perception than age-matched controls (Schwartz, Scheffler, & Lopez, 2013). Taken together, these findings suggest that greater susceptibility to lexical influence on speech perception is associated with poorer ability to suppress lexical activation and competition.

The association between susceptibility to lexical influence and ability to suppress lexical competition may be highly relevant to spoken word recognition, especially in nonideal listening conditions. Spoken word recognition involves competition among multiple lexical candidates (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Luce & Pisoni, 1998), which are activated in parallel, that agree with the partial input of the speech stream (Marslen-Wilson & Zwitserlood, 1989). Accurate word recognition demands suppression of these alternate candidates (Kapnoula & McMurray, 2016). Activation of lexical information embedded in the speech maskers in speech-in-speech conditions poses additional challenges to the speech processing system. For instance, greater similarities between the target and masker speech (e.g., different vs. same language; Brouwer et al., 2012) and/or more experience with the masker speech (Van Engen, 2012; Van Engen & Bradlow, 2007) are associated with greater interference on speech recognition accuracy. In particular, Van Engen and Bradlow (2007) studied native English speakers and found that two-talker English speech maskers resulted in greater interference than Mandarin maskers. Extending this study, Van Engen (2010) studied native Mandarin speakers who learned English as a second language and demonstrated a smaller release from their native language (Mandarin) than their less proficient second language (English). These findings suggest that word recognition performance in speech-in-speech conditions are associated with the amount of activation of lexical competitors and elevated difficulties in suppression.

To summarize, both the Ganong paradigm and speech-in-speech processing require listeners to suppress interfering linguistic information for good performance. The Ganong paradigm requires participants to suppress lexical information from word contexts (e.g., *-ift* and *-iss*), but speech-in-speech processing requires suppression of lexical competitors from speech maskers. If this association is valid, greater susceptibility to lexical influence should also predict greater difficulties in speech-in-speech processing. Interfering speakers would heighten lexical competition and exert significant demands on suppressing irrelevant lexical information to enhance task performance, which would be challenging for listeners who are susceptible to lexical influence. The current study aims to examine the extent to which speech processing abilities in conditions with interfering lexical information is associated with susceptibility to lexical influences, which indexes the ability to suppress lexical activation (Mattys & Wiget, 2011; Pitt & Samuel, 2006).

Examining the Association Between Susceptibility to Lexical Influences and Speech Processing Abilities in Maskers With or Without Lexical Interferences

In the current study, each participant underwent the classic Ganong task (Ganong 1980; Mattys & Wiget, 2011) and a speech-in-noise processing task. The Ganong task required participants to identify the initial phoneme (/g/ or /k/) presented in a word context (-iss or -ift). In the speech-in-noise task, participants were instructed to listen to spoken sentences mixed with either competing talkers or pink noise maskers and type what they heard. Competing talkers and pink noise maskers exert distinct effects on perceptual and cognitive systems (e.g., Zekveld et al., 2013). In particular, competing talkers create a significant amount of lexical interferences but pink noise maskers do not. For the competing-talker condition, we employed a two-talker masker because prior work has shown that the amount of audible lexical interferences decreases as the number of talkers in the speech masker increases (Van Engen & Bradlow, 2007).

To improve the external validity of this study, we included an audiovisual (AV) condition to emulate another common listening environment. Daily communication frequently involves face-to-face interaction with visual cues present. Visual cues have been shown to aid speech processing in the presence of background noise (Grant & Seitz, 2000; Sumbly & Pollack, 1954; Xie, Yi, & Chandrasekaran, 2014). Helfer and Freyman (2005) demonstrated that visual cues were particularly useful in speech maskers, arguably because visual cues assist listeners in extracting and attending to the target speech and also providing phonetic supplementation. Because cocktail-party-like scenarios are ubiquitous in everyday life, the frequent encounters of speech-in-speech processing may allow typical listeners to utilize visual cues effectively to compensate for signal degradation. In short, given the usefulness of visual cues in the competing-talker condition, this study included the AV conditions to examine the extent to which the provision of visual cues moderates the association between susceptibility to lexical influences and speech processing in different masker types.

We made the following predictions: As the Ganong effect is related to difficulties suppressing irrelevant lexical information during speech perception (Mattys, Seymour, Attwood, & Munafò, 2013; Mattys & Scharenborg, 2014; Mattys & Wiget, 2011), we posit that greater Ganong effect relates to poorer spoken word recognition in the competing-talker condition. In contrast, we predict that a greater Ganong effect will exhibit a weaker association with spoken word recognition in the pink noise condition because there is no audible linguistic information in this masker type.

Method

Participants

Twenty-nine monolingual English-speaking young adults, aged 18–35 years, (20 women, 9 men) were

recruited from The University of Texas at Austin. All participants provided written informed consent and received monetary compensation for their participation. Participants reported no previous history of language or hearing problems, and reported normal or corrected-to-normal vision. All participants underwent an audiological test to ensure thresholds ≤ 25 dB hearing level (HL) at 1000 Hz, 2000 Hz, and 4000 Hz for each ear. All materials and procedures were approved by the Institutional Review Board at the University of Texas at Austin. The number of participants was preset at between 25 and 30, which approximates a previous study utilizing the Ganong paradigm (Mattys et al., 2013).

Phoneme Categorization Task

Stimuli

Preparation of the stimuli followed the procedures described in Mattys and Wiget (2011) and contained three eight-step continua: *gift–kift* (real word vs. nonword), *giss–kiss* (nonword vs. real word), and *gi–ki* (nonword vs. nonword). The *gift–kift* and *giss–kiss* continua were included to estimate the magnitude of lexical influence, but *gi–ki* was included to assess lexical-influence-free phoneme categorization. Phoneme categorization performance in the *gi–ki* condition provides information about whether differences in the amount of lexical influence on phoneme categorization is due to differences in strength of lexical influence or due to differences in acoustic-phonetic processing of /g/ and /k/ (e.g., Mattys & Wiget, 2011; Mattys et al., 2013; Mattys & Scharenborg, 2014). Multiple productions of the words *gift* and *kiss* were recorded by a female native speaker of American English in a sound-attenuated booth with a sample rate of 44 kHz and sample resolution of 16 bit. The clearest tokens were selected, which were then split into initial consonant /g/–/k/, vowel /ɪ/, and coda /ft/ and /s/. We chose a vowel that originated from one of the *kiss* tokens because this vowel exhibited relatively neutral coarticulation. We created an eight-step continuum of the initial consonants /g/–/k/ out of /k/ by editing out the aspiration using Praat (Boersma & Weenink, 2010). The final continuum had the following VOT values: 15 ms, 23 ms, 28 ms, 33 ms, 38 ms, 43 ms, 48 ms, and 56 ms. The difference in VOT increments was a strategic decision; the end points are further away to ensure clear end points. The continuum in the middle is denser to capture the VOT area of uncertainty. The eight-step /g/–/k/ continuum was then recombined with the vowel /ɪ/, and coda of either /ft/ or /s/, making a total of 24 syllables. The average duration of the three continua of *gift–kift*, *kiss–giss*, and *gi–ki* were 521 ms, 468 ms, and 193 ms, respectively. The procedures and properties of the stimuli were consistent with those used in previous studies (Mattys et al., 2013; Mattys & Wiget, 2011).

Procedure

During testing, the stimuli were binaurally presented to participants over Sennheiser HD-280 Pro headphones (Sennheiser, Wedemark, Germany) at a comfortable, fixed

listening level (approximately 70 dB SPL). Each of the 24 syllables that form a continuum of VOT in the context of *-i*, *-ift*, and *-iss* was presented randomly five times, resulting in a total of 120 trials. Before the experiment, participants were explicitly instructed to focus on the initial consonant and to ignore the meaning of the syllables. On each trial, after stimulus presentation, participants were asked to decide whether the first sound of the syllable was /g/ or /k/. Following the procedures used in Mattys and Wiget (2011) and Mattys et al. (2013), the magnitude of susceptibility to lexical influence was calculated as the average percentage change in identification in the *giss-kiss* continuum from the counterparts in the *gift-kift* continuum across the eight steps. To illustrate, the left panel of Figure 1A denotes an individual who exhibits high susceptibility to lexical influences. As shown in this figure, the proportion of /g/ responses by the participant at 15 ms VOT is 100% in *-ift* context, but only 60% in *-iss* context. As a result, the percentage difference of /g/ response is 40% at 15 ms VOT. In contrast, the right panel of Figure 1A shows an individual who exhibits low susceptibility to lexical influences. This individual demonstrates a smaller percentage difference of /g/ response at 15 ms VOT. The procedure and result are the same if the /k/ response is used for measurement instead of /g/. In this case, the percentage of /k/ responses at 15 ms VOT for the individual displayed on the left panel of Figure 1A is 0% in *-ift* context and 40% in *-iss* context, which gives rise to a percentage difference of /k/ response with an equal amount of 40%. The magnitude of susceptibility to lexical influence (i.e., Ganong effect) was calculated as the average percent differences of /g/ response between *giss-kiss* continuum and *gift-kift* continuum across the 8 VOTs.

Speech Perception in Noise Task

Target Sentences

The target stimuli were taken from the Basic English Lexicon (BEL; Calandruccio & Smiljanic, 2012). The BEL sentences were created based on conversational speech samples from 100 nonnative English speakers (Rimikis, Smiljanic, & Calandruccio, 2013). The BEL consists of 500 sentences divided into 20 test lists of 25 sentences each. Sentences were sorted into test lists to equate for the distribution of syntactic structures, vocabulary (i.e., rate of occurrence of a keyword in the master lexicon), number of syllables, and high-frequency speech evenly across lists. Each BEL sentence (e.g., The GIRL LOVED the SWEET COFFEE) is five to seven words long and contains four keywords (uppercase words) for intelligibility scoring. Out of the total 2,000 keywords, 939 keywords are unique and not repeated across any other sentence.

The 80 BEL sentences selected for the current study were identical to the sentence stimuli selected by Xie et al. (2014). The sentences were video-recorded by one male native American English speaker. For each target sentence, the audio was detached from the video and RMS amplitude was equalized to 50 dB SPL using Praat (Boersma & Weenink, 2010).

Masks

Two types of maskers were used in this task. The first type of masker consisted of the voices of two male American English speakers reciting sentences unrelated to any of the topics used in the target sentences. These two speakers were different from the speaker who produced the target sentences.

Figure 1. (A) Percentage of /g/ responses for the *gift-kift* (solid black line with circles), *gi-ki* (light gray dotted line with triangles), and *giss-kiss* (dark gray solid line with squares) continua from an example participant with high susceptibility of lexical influence (left panel) and an example participant with low susceptibility of lexical influence (right panel). The magnitude of lexical influence is calculated as the average proportion differences of /g/ response between the *gift-kift* continuum and *giss-kiss* continuum across the eight steps of voice onset time (VOT). (B) Performance on the speech perception in noise task of the two example participants displayed in Figure 1A.

The sentences were simple, semantically meaningful sentences from Bradlow and Alexander (2007). The second type of masker was steady-state pink noise, which was created using the noise generator option in Audacity software (Version 1.2.6; Audacity Developer Team, 2008). These two types of maskers were identical to those described in Xie et al. (2015, experiment 1).

Mixing Targets and Maskers

Each target sentence audio clip was mixed with five corresponding two-talker babble clips and five corresponding pink noise clips to create stimuli of the same target sentence for each masker type with the following five SNRs: -4, -8, -12, -16, and -20 dB. Each final stimulus was composed as follows: 500 ms noise before the onset of the target sentence, the target and noise together, and 500 ms noise after the offset of the target sentence. The mixed audio clips served as the stimuli for the audio-only (AO) condition. The mixed audio clips were then reattached to the corresponding videos to create the stimuli for the AV condition using Final Cut Pro software (Apple, Cupertino, CA). A freeze frame of the speaker was captured and displayed during the 500 ms noise present before the onset and after the offset of the target sentence. In total, there were 400 AO and 400 corresponding AV stimuli mixed with two-talker babble (80 sentences $\times$ 5 SNRs), and 400 AO and 400 corresponding AV stimuli mixed with pink noise (80 sentences $\times$ 5 SNRs).

Procedure

Each participant completed 20 experimental conditions: 2 (type of noise: two-talker babble or pink noise) $\times$ 5 (SNR: -4, -8, -12, -16, or -20 dB) $\times$ 2 (modality: AO or AV). In each condition, the participants listened to four unique target sentences binaurally presented through Sennheiser HD-280 Pro headphones. These target sentences were randomly chosen from the full set of 80 target sentences, and were never repeated across experimental conditions within each participant. Hence, the target sentences in a given condition were randomized across participants. The target sentences across the 20 conditions were mixed and presented in random order to the participants. As in a previous study (Xie et al., 2014), participants were asked to type out the target sentence after each stimulus presentation. If they were unable to understand the entire sentence, they were asked to report any intelligible words and make their best guess. If they were not able to make out any words, they were instructed to type "X." Participants had unlimited time to respond. Responses were scored as accurately typed keywords (four per sentence). Keywords with added or omitted morphemes were scored as incorrect.

Additional Behavioral Measures

From the same participants, we measured lexical retrieval speed with a rapid naming task (Woodcock, McGrew, & Mather, 2001) because faster lexical access may strengthen the lexical status of the phonetic stimuli and aggravate difficulty suppressing irrelevant lexical

information (Stewart & Ota, 2008). Participants named 120 simple pictures as rapidly and accurately as possible in 2 minutes. The dependent variable was completion time. Second, we measured participants' ability to segregate syllables with a 20-item phoneme elision task in the Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999) because failure to process speech sounds analytically may aggravate difficulty focusing on the initial phoneme. Last, on a control condition, we measured participants' ability to process acoustic-phonetic features of speech sounds within the /g1-/ /k1/ continuum, a context in which relatively few lexical processes are at play. Previous studies found that, although lexical influence as measured by the *gift-kift* and *giss-kiss* continua was found to be elevated by factors that negatively impact one's ability to ignore irrelevant information, relatively little effect was found on the *gi-ki* continuum (Mattys & Wiget, 2011). Also, the -i context is typically of little relevance to measuring lexical influence (Mattys & Scharenborg, 2014). The limited association between response patterns in the -i context and lexical influence relative to -ift and -iss continua allows the -i context to index perceptual processing of phonemes. The performance in the *gi-ki* categorization was calculated as the proportion of /g/ responses at each of the eight-step *gi-ki* continuum and the total proportion of /g/ responses averaged across the eight-step continuum.

Results

Susceptibility to Lexical Influences

The average percentage change in identification in the *giss-kiss* continuum from the counterparts in the *gift-kift* continuum across the eight steps was used to measure the magnitude of susceptibility of lexical influences (Mattys & Wiget, 2011; Mattys et al., 2013). The magnitude of susceptibility of lexical influences for all 29 participants ranged from 15% to 100%. The distribution of magnitudes of susceptibility of lexical influences is shown in Figure 2.

Relationship Between Susceptibility to Lexical Influence and Speech Recognition in Noise Across All SNRs

Here we analyzed the full set of speech recognition data to determine the extent to which susceptibility to lexical influence predicted the proportion of correct keywords under any of the test conditions. Before inferential statistics, we first describe the overall findings. Regardless of the stimulus presentation modality, the group with large susceptibility to lexical influences was less accurate in recognizing keywords embedded in two-talker babble, relative to the group with small susceptibility to lexical influences (mean proportion of correct keywords collapsed across five SNRs: $M = .392$ [95% CI = .311 to .474] vs. $M = .532$ [95% CI = .448 to .615] for the AO condition; $M = .602$ [95% CI = .53 to .674] vs. $M = .742$ [95% CI = .678 to .806] for the AV condition).

Figure 2. Distribution of the magnitude of susceptibility to lexical influences. The vertical dashed line indicates the median value. The bin size is 5%.

However, the difference in keyword recognition accuracy between the group with large susceptibility to lexical influence and the group with small susceptibility to lexical influence was much smaller for pink noise (mean proportion of correct keywords collapsed across five SNRs: $M = .411$ [95% CI = .314 to .509] vs. $M = .393$ [95% CI = .288 to .498] for the AO condition; $M = .523$ [95% CI = .432 to .615] vs. $M = .572$ [95% CI = .477 to .668] for the AV condition). The selective speech recognition difficulty observed in the two-talker babble condition for individuals with large susceptibility to lexical influence is also illustrated in Figure 1B. Also, regardless of susceptibility to lexical influences, type of noise, and stimulus presentation modality, the proportion of correct keywords increased as SNR became higher. Regardless of susceptibility to lexical influences and type of noise, the proportion of correct keywords was higher in the AV condition than in the AO condition, especially for SNRs lower than -4 dB.

Statistical analyses were performed using the lme4 package 1.1–7 in R version 3.2.0 (Bates, Maechler, & Bolker, 2012). The magnitude of susceptibility to lexical influence was included in a generalized linear mixed-effects logistic regression model to test the effect on speech intelligibility. In this model, keyword identification (correct or incorrect) was the dichotomous dependent variable. SNR, type of noise, modality, magnitude of susceptibility to lexical influence, and their interactions were included as fixed effects. Type of noise and modality were treated as categorical variables. Original SNR and the magnitude of susceptibility to lexical influence were mean centered. The mean-centered SNR and lexical influence were treated as continuous variables. In the model, the reference levels were AO and pink noise.

Following Barr, Levy, Scheepers, and Tily (2013), we included both random intercepts and random slopes.

For the random slopes, we first tried the maximally appropriate structures. However, the model failed to converge after 20,000 iterations. To achieve convergence, we progressively dropped the random slopes that accounted for least variance in the model. The final random-effects structure that converged was construed as this: (1 | type of noise | subject) + (1 | sentence). Although this random-effects structure was not maximal, it did include the random slope for the variable (i.e., type of noise) that was our major focus in the current study.

We examined main effects of SNR, type of noise, modality, and magnitude of susceptibility to lexical influence via model comparison. To do so, we compared the base model (which included the random-effects structure) to the same model with the addition of SNR, type of noise, modality, or magnitude of susceptibility to lexical influence. We tested each interaction among the independent variables by comparing a model with such interaction and the lower level effects to a model with only the lower level effects. Model comparisons were achieved using the likelihood ratio (Baayen, Davidson, & Bates, 2008).

Of our primary interest, there was a significant main effect of magnitude of susceptibility to lexical influence, $\beta = -.006$, $SE = .003$, $\chi^2(1) = 3.978$, $p = .046$, indicating that the probability of correct keyword identification was higher for individuals with smaller susceptibility to lexical influence. It is important to note that there was a significant interaction between type of noise and magnitude of susceptibility to lexical influence, $\beta = -.012$, $SE = .005$, $\chi^2(1) = 5.023$, $p = .025$, indicating that the effect of susceptibility to lexical influence on speech intelligibility was stronger in two-talker babble conditions than in pink noise conditions. This was confirmed by a follow-up analysis showing that the estimate for the effect of susceptibility to lexical influence was larger in two-talker babble conditions, $\beta = -.018$, $SE = .006$, $\chi^2(1) = 8.339$, $p = .004$, than in pink noise conditions, $\beta = -.0058$, $SE = .0026$, $\chi^2(1) = 4.604$, $p = .032$. It is notable that the three-way interaction among type of noise, modality, and magnitude of susceptibility to lexical influence was not significant, $\beta = .002$, $SE = .004$, $\chi^2(1) = 0.359$, $p = .549$. This indicates that the interaction effect between type of noise and magnitude of susceptibility to lexical influence was not affected by the presentation mode (AV or AO) of the stimuli.

There was a significant main effect for SNR, $\beta = .246$, $SE = .006$, $\chi^2(1) = 2684.1$, $p < .001$, indicating that the probability of correct keyword identification increases as SNR increases. SNR also significantly interacted with type of noise, $\beta = -.111$, $SE = .011$, $\chi^2(1) = 101.72$, $p < .001$, and modality, $\beta = -.073$, $SE = .011$, $\chi^2(1) = 43.249$, $p < .001$, indicating that the intelligibility benefit from SNR increment was less for two-talker babble (relative to pink noise), and for AV conditions (relative to AO conditions).

In addition, there were significant main effects for type of noise, $\beta = .418$, $SE = .128$, $\chi^2(1) = 9.064$, $p = .003$, and modality, $\beta = .835$, $SE = .047$, $\chi^2(1) = 331.42$, $p < .001$, indicating that the probability of correct keyword identification was higher for two-talker babble conditions (relative

to pink noise), and for AV conditions (relative to AO conditions). There was also a significant interaction between type of noise and modality, $\beta = .494$, $SE = .093$, $\chi^2(1) = 28.176$, $p < .001$, indicating that the type of noise effect became stronger in AV conditions. This was confirmed by a follow-up analysis that revealed higher probability of correct keyword identification for two-talker babble (relative to pink noise) in AV conditions, $\beta = .720$, $SE = .159$, $\chi^2(1) = 15.785$, $p < .001$, but not in AO conditions, $\beta = .198$, $SE = .156$, $\chi^2(1) = 1.544$, $p = .214$.

Further, there were marginally significant three-way interactions among SNR, modality, and magnitude of susceptibility to lexical influence, $\beta = -.001$, $SE = .001$, $\chi^2(1) = 3.637$, $p = .057$, indicating that intelligibility benefit from SNR increment was reduced for the AV condition (relative to AO condition) in individuals with both large and small lexical influence, but the amount of reduction was greater in individuals with large susceptibility to lexical influence. No other interactions reached significance (all $p > .13$).

Relationship Between Susceptibility to Lexical Influence and Speech Recognition in Noise at -12 dB

We used Kendall's tau to calculate the correlation between an individual's susceptibility to lexical influences and their proportion of correct keywords across four conditions: (a) two-talker babble and audio-only condition (2T-AO), (b) two-talker babble and AV condition (2T-AV), (c) pink noise and audio-only condition (P-AO), and (d) pink noise and AV condition (P-AV). Here we focused on the SNR at -12 dB for two reasons: (a) in order to minimize the negative effect on the correlation due to ceiling or floor performance on the speech recognition in noise task, and (b) in alignment with results from Ross, Saint-Amour, Leavitt, Javitt, and Foxe (2007), the effect of visual cues on speech recognition is likely to be largest at -12 dB. There were significantly negative correlations between the susceptibility to lexical influences and the proportion of correct keywords in the 2T-AO condition ($\tau = -.327$, $p = .016$), 2T-AV condition ($\tau = -.368$, $p = .007$), and P-AV condition ($\tau = -.343$, $p = .013$). These results suggest that higher susceptibility to lexical influences is associated with worse performance in speech recognition across these three conditions. There was no significant correlation between susceptibility to lexical influences and proportion of correct keywords in the P-AO condition ($\tau = -.047$, $p = .748$). These correlations are illustrated in Figure 3.

Other Behavioral Measures

A multiple linear regression (using the enter method) was calculated to predict the amount of susceptibility to lexical influence (bias) based on lexical retrieval speed, abilities to process speech sounds analytically, and differences in perceptual processing of phonemes. The regression results indicated that the three behavioral measures explained

negligible variance of the amount of susceptibility to lexical influence (bias), adjusted $R^2 = -.002$, $F(3, 24) = 0.981$, $p = .481$. We applied similar regression analyses to test if the three other measures significantly predicted overall word recognition performance in any of the four noise conditions (i.e., P-AO, P-AV, 2T-AO, or 2T-AV). The results indicated that the three behavioral measures did not significantly predict word recognition performance in any of the four noise conditions: P-AO: adjusted $R^2 = .123$, $F(3, 24) = 2.258$, $p = .107$; P-AV: adjusted $R^2 = .155$, $F(3, 24) = 2.652$, $p = .072$; 2T-AO: adjusted $R^2 = -.093$, $F(3, 24) = 0.231$, $p = .874$; 2T-AV: adjusted $R^2 = -.116$, $F(3, 24) = 0.064$, $p = .978$.

Discussion

Speech processing seldom occurs in ideal listening conditions (Mattys, Davis, Bradlow, & Scott, 2012). The current study examined the extent to which young adults who demonstrated greater susceptibility to lexical influences on phonemic processing also exhibit poorer speech recognition when competing talkers (and, therefore, cognitive-linguistic interference) are present. Using the classic Ganong paradigm to index susceptibility to lexical influences, we demonstrated that listeners with greater susceptibility experienced selectively greater difficulties in competing-talker conditions, wherein good performance requires listeners to ignore irrelevant linguistic information. Furthermore, our findings extended to AV conditions, suggesting that our results also apply to conditions wherein speech processing is enhanced by visual cues (e.g., Helfer & Freyman, 2005). Finally, these selectively greater difficulties are not related to individual differences in lexical access, ability to process speech sounds analytically, or differences in perceptual processing of phonemes.

Ganong Effect and Speech-In-Speech Processing is Associated With Lexical Activation and Interferences

Our study extends the well-studied Ganong effect by relating the phenomenon to speech processing ability in noises that resemble day-to-day ecological communicative situations. Our findings indicate an association between lexical influences one experiences within a word (e.g., *-ift*) and sentence recognition in noise.

A putative mechanism underlying the Ganong effect is that lexical representations are activated automatically and may influence detailed phonetic analysis of phonemes (Mattys & Wiget, 2011). Indeed, a multimodal neuroimaging study on the Ganong effect provides support to the view that lexical processes influence lower level phonetic perception (Gow, Segawa, Ahlfor, & Lin, 2008). According to this account, the extent to which phonetic details are fully integrated at a later stage of processing depends on the extent to which lexical influence can be suppressed (Mattys & Wiget, 2011). Whether the locus of interference of speech maskers on target speech processing is at the

Figure 3. Scatterplots illustrate the relationship between the susceptibility to lexical influences and the proportion of correctly identified keywords at auditory-only (upper) and audiovisual (lower) conditions for two-talker babble (left) and pink noise (right) conditions at -12 dB, respectively. The solid dots depict individual participants ($n = 29$). The solid lines show the trend lines fitted by linear regression functions.

sublexical (syllables or phonemes) or lexical (word) level (Boulenger, Hoen, Ferragne, Pellegrino, & Meunier, 2010; Van Engen & Bradlow, 2007) is still an open question. Recent studies postulate that the impact of speech maskers occur at multiple levels via cognitive-linguistic interferences (Freyman et al., 2004). These interferences include increased cognitive load and attention distraction (Cooke, Lecumberri, & Barker, 2008; Mattys et al., 2012), which may be associated with the ability to process speech under unpredictable interruptions and the need to shift constantly between the target and unattended masker (Shinn-Cunningham & Best, 2015). To summarize, although the influence of lexical activation on phonemic processing is more restricted to the processing and integration of phonetic details, speech maskers may likely interfere with word recognition in speech-in-speech condition at multiple processing levels.

Although the Ganong effect and speech-in-speech processing may differ in the locus of lexical interferences, our findings highlight an association between the two processes. Our results indicated that this association may be less related to differences in sensory processing of bottom-up information. First, the Ganong effect exhibited little relationship with perceptual processing of phonemes in the context-free condition (i.e., proportion of /g/ or /k/ responses in a *gi-ki* continuum). Mattys and Scharenborg (2014) have also demonstrated that decline in hearing sensitivity alone could not account for a greater Ganong effect in older adults. In addition, by strategically comparing two

noise conditions that were characterized by the presence or absence of lexical interferences, we demonstrated that susceptibility to lexical influences is more strongly associated with poorer performance in conditions with a great amount of lexical interferences and competition (e.g., Van Engen & Bradlow, 2007; Van Engen, 2012).

Lexical competition and effective suppression is a key component in most current word recognition models (Luce & Pisoni, 1998; Marslen-Wilson, 1987; McClelland and Elman, 1986; Norris, 1994). Successful word recognition is determined by the relative activation levels of the target word as opposed to its neighbors. In addition, the activation level of the neighboring candidates is suppressed if there is an inconsistency with the phonetic input (Sommers, 1996; Sommers & Danielson, 1999). In the context of our findings, the Ganong effect and speech-in-speech processing are likely to be associated with susceptibility to lexical interferences that may result from individual differences in lexical suppression. What remains unclear is whether these individual differences are specific to lexical processing stages or more domain-general cognitive processes. As indicated above, the Ganong effect has been used to index the degree of lexical activation and/or its suppression (e.g., Mattys & Wiget, 2011; Pitt & Samuel, 2006). Relative to individuals who exhibit smaller lexical influences, lexical information may be more strongly activated in individuals who exhibit greater lexical influences. A heightened amount of lexical activation may in turn impact word recognition processes by moderating the relative differences among the target word

and its neighbors. It is important to note that greater lexical activation implies smaller relative differences between the target word and its alternates, giving rise to less effective word recognition and greater demands on ideal listening conditions. Because speech maskers entail even greater activation of lexical candidates that are irrelevant to the target speech, listeners who exhibited greater lexical influences would experience elevated challenges in singling out the target word.

In contrast, greater susceptibility to lexical influences and poorer speech-in-speech processing may also be driven by a generalized suppression ability. Because listeners are always explicitly instructed to ignore certain information in the Ganong paradigms, this methodology has been used to probe one's ability to suppress irrelevant information and attentional control (Mattys & Scharenborg, 2014). Indeed, using this paradigm, susceptibility to lexical influences is associated with aging (Mattys & Scharenborg, 2014), divided attention (Mattys & Wiget, 2011), and anxiety (Mattys et al., 2013). All of these factors have been linked to changes in attentional control (e.g., Hartman & Hasher, 1991) and recruitment of prefrontal cortex (Bishop, 2009). These findings align with the emerging evidence that speech processing ability in challenging environments is associated with working memory capacity (Moradi, Lidestam, Saremi, & Rönnberg, 2014; Rönnberg, Rudner, Lunner, & Zekveld, 2010; Zekveld, et al., 2013). Working memory capacity, which predicts performance of not only language processing tasks but also a wide variety of higher-order cognitive assessments, has been argued to index executive attention that regulates the information content activated in the memory (for a review, see Engle, 2002).

Our results cannot completely disambiguate between suppression ability that is lexical-specific or a more generalized suppression ability. Mattys and colleagues (Mattys et al., 2013; Mattys & Scharenborg, 2014) indicated that elevated susceptibility to lexical influences with their designs could be considered from both language-specific and domain-general perspectives (Mattys et al., 2013; Mattys & Scharenborg, 2014). To tease apart the two mechanisms, a follow-up study could include both a lexical negative priming task and a nonlinguistic inhibitory control task in addition to the tasks used here. In a negative priming task, participants will identify target pictures that are presented with competing phonological candidates. On the following trial, the competitors will be the target of response (Blumenfeld & Marian, 2011). The negative priming effect will be measured by the delay in response to the competing phonological candidates on the subsequent trials, which is taken to index both activation and suppression of the competitors (Blumenfeld & Marian, 2011). If susceptibility to lexical influences and speech-in-speech processing is more strongly associated with lexical-specific activation and suppression, performance on these tasks should correlate with only the negative priming effect and thus exhibits little relationship with the nonlinguistic inhibitory control task.

Our findings raise an interesting conundrum with regard to the extent to which reliance on lexical information

enhances speech processing in challenging listening environments. Examining the linkage between susceptibility to lexical influences and speech processing in more complex listening conditions that resemble daily listening scenarios has the potential to inform hearing rehabilitation programs. On one hand, reliance on high-level contextual knowledge (e.g., lexical-semantic) may compensate for degraded signals in less-than-optimal listening conditions (e.g., Van Engen, Phelps, Smiljanic, & Chandrasekaran, 2014). On the other hand, there are also conditions wherein overreliance on high-level information may impair speech recognition (Norris, McQueen, & Cutler, 2000). The extent to which lexically biased listening is an optimal approach to process unclear speech signals may ultimately depend on the signal content and degree of signal degradation. Successful communication depends on flexible shifting of attention according to specific listening conditions: Listeners need to focus on the acoustic details of the degraded signals, yet they also need to utilize long-term lexical knowledge to fill in the missing perceptual information when appropriate.

Potential Implications on the Distinction of Energetic Masker and Informational Masker

Our findings also have potential implications for current discussions about the effect of energetic masking (EM) and informational masking (IM) on speech processing (Brungart, 2001; Mattys et al., 2012). Background noises, in particular competing speech, challenge the speech processing system primarily via EM and IM. IM refers to how irrelevant information content interferes with target speech processing, and EM refers to the overlapping of physical characteristics between the target signal and masker (Brungart, 2001; Sörqvist & Rönnberg, 2011). Relative to EM, IM has been a challenge to define and is often conceptualized as the interference from noise once EM is accounted for (Cooke et al., 2008). In other words, the definition of IM is primarily a "classification by exclusion" (Carlile & Corkhill, 2015). This study does not suggest that the lexical (i.e., two-talker babble) and nonlexical (i.e., pink noise) distinction is synonymous with the IM and EM distinction, as IM does not mandate the presence of lexical interferences. However, as speech maskers introduce a great amount of irrelevant information and lexical interferences that would be absent in EM and more closely resemble IM (Pollack, 1975), our findings indicate that the mechanism underlying susceptibility to lexical influences may be more strongly associated with processing under IM-dominant conditions as opposed to EM-dominant conditions.

Shinn-Cunningham's (2008) model of speech processing in noise postulates that listeners need to segregate the target source from maskers in EM-dominant conditions, a mechanism known as *object formation*. In contrast, listeners need to also listen to the target while ignoring competing maskers in IM-dominant conditions, a mechanism known as *object selection* (Shinn-Cunningham, 2008). Our findings suggest that individuals who are more susceptible

to lexical influences may also experience greater difficulty with object selection in competing-talker conditions that result in a considerable amount of IM. Studying the locus of susceptibility to lexical influences may potentially advance understandings about the impact of IM on target speech processing.

Limitations of the Present Study

To summarize, our study demonstrated that two-talker babble, which introduced intense lexical competition and interferences (Boulenger et al., 2010; Hoen et al., 2007), presented selectively greater challenges for listeners who exhibited greater reliance on top-down lexical processes. However, it is preliminary to postulate that greater reliance on bottom-up processing may provide a protective effect against the interferences from speech maskers. Many questions remain regarding the nature of the association. The questions that concern the mechanism underlying the Ganong effect also apply to speech-in-speech processing. For instance, does activation of lexical competitors embedded in speech maskers interfere with bottom-up processing of the target speech, or do the lexical competitors interfere with postperceptual reporting of the target speech? Previous studies have shown that manipulating listeners' attention to the target phonemes and entire words could modulate the amount of lexical influences on phonemic processing (McClelland, Mirman, & Holt, 2006; Mirman, McClelland, Holt, & Magnuson, 2008; Norris et al., 2000). Extending these studies, a potential follow-up study would be to examine the extent to which speech-in-speech processing is associated with both the Ganong effect and susceptibility to attention manipulation. If greater reliance on bottom-up processing provides a protective effect against the interferences from speech maskers, listeners who exhibit a small Ganong effect and high resilience toward attention manipulation should obtain the best performance on speech-in-speech processing. The current study used only four sentences for each condition to estimate the speech perception performance. Administering a larger number of stimuli may yield a more stable estimate of participants' performance for each condition and improve the reliability of our results. Future studies should include a larger number of stimuli in the measure of speech perception abilities in adverse listening conditions.

Conclusion

Although cocktail-party-like scenarios present a formidable computational challenge to listeners on a daily basis, our knowledge about speech-in-speech processing remains incomplete. To our knowledge, this is the first study to demonstrate that greater susceptibility to lexical influences in healthy young adults, as quantified by the magnitude of lexical influences in the Ganong paradigm, also predicts selectively poorer abilities to process speech accurately in noisy conditions that are ubiquitous in daily settings. Previous studies on speech-in-speech listening have

largely focused on stimulus characteristics, but our study highlights the importance of studying listeners' inherent properties in order to understand abilities to process target speech in the presence of competing talkers, specifically activation of lexical knowledge and its suppression.

Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders (Grant R01 DC-013315 to B. Chandrasekaran). We thank Kristin J. Van Engen, Kirsten Smayda, Han-Gyol Yi, and Jasmine E. B. Phelps for their invaluable assistance in stimulus preparation, data management, and data analysis. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- Anderson, S., White-Schwoch, T., Parbery-Clark, A., & Kraus, N. (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hearing Research*, 300, 18–32.
- Audacity Developer Team. (2008). *Audacity* (Version 1.2.6) [Computer software]. Available from <http://audacity.sourceforge.net/download>.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using Eigen and R syntax. Version 1.1–7 in R package Version 3.2.0.
- Bishop, S. J. (2009). Trait anxiety and impoverished prefrontal control of attention. *Nature Neuroscience*, 12, 92–98.
- Blumenfeld, H. K., & Marian, V. (2011). Bilingualism influences inhibitory control in auditory comprehension. *Cognition*, 118, 245–257.
- Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 5.1.25) [Computer program]. Retrieved from <https://www.praat.org>
- Boulenger, V., Hoen, M., Ferragne, E., Pellegrino, F., & Meunier, F. (2010). Real-time lexical competitions during speech-in-speech comprehension. *Speech Communication*, 52, 246–253.
- Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, 121, 2339–2349.
- Brouwer, S., Van Engen, K. J., Calandruccio, L., & Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *The Journal of the Acoustical Society of America*, 131, 1449–1464.
- 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.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, 110, 2527–2538.
- 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, 1342–1355.
- Carlile, S., & Corkhill, C.** (2015). Selective spatial attention modulates bottom-up informational masking of speech. *Scientific Reports*, 5, 8662.
- Chandrasekaran, B., Van Engen, K., Xie, Z., Beevers, C. G., & Maddox, W. T.** (2015). Influence of depressive symptoms on speech perception in adverse listening conditions. *Cognition and Emotion*, 29, 900–909.
- Cherry, E. C.** (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25, 975–979.
- Cooke, M., Lecumberri, M. G., & 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, 414–427.
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M.** (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16, 507–534.
- Engle, R. W.** (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11(1), 19–23.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S.** (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 109, 2112–2122.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S.** (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 115, 2246–2256.
- Ganong, W. F.** (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110–125.
- Gow, D. W., Segawa, J. A., Ahlforth, S. P., & Lin, F. H.** (2008). Lexical influences on speech perception: A Granger causality analysis of MEG and EEG source estimates. *NeuroImage*, 43, 614–623.
- Grant, K. W., & Seitz, P. F.** (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the American Academy of Audiology*, 108, 1197–1208.
- Hartman, M., & Hasher, L.** (1991). Aging and suppression: Memory for previously relevant information. *Psychology and Aging*, 6, 587–594.
- Helfer, K. S., & Freyman, R. L.** (2005). The role of visual speech cues in reducing energetic and informational masking. *The Journal of the Acoustical Society of America*, 117, 842–849.
- Hoen, M., Meunier, F., Grataloup, C. L., Pellegrino, F., Grimault, N., Perrin, F., . . . Collet, L.** (2007). Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. *Speech Communication*, 49, 905–916.
- Kapnoula, E. C., & McMurray, B.** (2016). Training alters the resolution of lexical interference: Evidence for plasticity of competition and inhibition. *Journal of Experimental Psychology: General*, 145(1), 8–30.
- Li, L., Daneman, M., Qi, J. G., & Schneider, B. A.** (2004). Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults? *Journal of Experimental Psychology: Human Perception and Performance*, 30(6), 1077–1091.
- Luce, P. A., & Pisoni, D. B.** (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36.
- Mainela-Arnold, E., Evans, J. L., & Coady, J. A.** (2008). Lexical representations in children with SLI: Evidence from a frequency-manipulated gating task. *Journal of Speech, Language, and Hearing Research*, 51, 381–393.
- Marslen-Wilson, W. D.** (1987). Functional parallelism in spoken word recognition. *Cognition*, 25, 71–102.
- Marslen-Wilson, W., & Zwitserlood, P.** (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 576–585.
- Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K.** (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, 27, 953–978.
- Mattys, S. L., & Scharenborg, O.** (2014). Phoneme categorization and discrimination in younger and older adults: A comparative analysis of perceptual, lexical, and attentional factors. *Psychology and Aging*, 29, 150–162.
- Mattys, S. L., Seymour, F., Attwood, A. S., & Munafò, M. R.** (2013). Effects of acute anxiety induction on speech perception: Are anxious listeners distracted listeners? *Psychological Science*, 24, 1606–1608.
- Mattys, S. L., & Wiget, L.** (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, 65, 145–160.
- McClelland, J. L., & Elman, J. L.** (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McClelland, J. L., Mirman, D., & Holt, L. L.** (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, 10, 363–369.
- McMurray, B., Munson, C., & Tomblin, J. B.** (2014). Individual differences in language ability are related to variation in word recognition, not speech perception: Evidence from eye movements. *Journal of Speech, Language, and Hearing Research*, 57, 1344–1362.
- McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B.** (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, 60, 1–39.
- Mesgarani, N., & Chang, E. F.** (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485(7397), 233–236.
- 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.
- Moradi, S., Lidestam, B., Saremi, A., & Rönnerberg, J.** (2014). Gated auditory speech perception: Effects of listening conditions and cognitive capacity. *Frontiers in Psychology*, 5, 531.
- Norris, D.** (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52, 189–234.
- Norris, D., McQueen, J. M., & Cutler, A.** (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23, 299–325.
- Pitt, M. A., & Samuel, A. G.** (2006). Word length and lexical activation: Longer is better. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 1120–1135.
- Pollack, I.** (1975). Auditory informational masking. *The Journal of the Acoustical Society of America*, 57, S5–S5.
- Rimikis, S., Smiljanic, R., & Calandruccio, L.** (2013). Nonnative English speaker performance on the basic English lexicon (BEL) sentences. *Journal of Speech, Language, and Hearing Research*, 56, 792–804.
- Rogers, C. S., Jacoby, L. L., & Sommers, M. S.** (2012). Frequent false hearing by older adults: The role of age differences in metacognition. *Psychology and Aging*, 27, 33–45.
- Rönnerberg, J., Rudner, M., Lunner, T., & Zekveld, A. A.** (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise and Health*, 12, 263–269.

- Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, 17, 1147–1153.
- Samuel, A. G. (2011). Speech perception. *Annual Review of Psychology*, 62, 49–72.
- Schwartz, R. G., Scheffler, F. L., & Lopez, K. (2013). Speech perception and lexical effects in specific language impairment. *Clinical Linguistics & Phonetics*, 27, 339–354.
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, 12, 182–186.
- Shinn-Cunningham, B., & Best, V. (2015). Auditory selective attention. In A. Kingstone, J. M. Fawcett, & E. F. Risko (Eds.), *The Handbook of Attention* (pp. 99–118). Cambridge, MA: MIT Press.
- Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *The Journal of the Acoustical Society of America*, 118, 2775–2778.
- Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychology and Aging*, 11, 333–341.
- Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging*, 14, 458–472.
- Sörqvist, P., & Rönnerberg, J. (2011). Episodic long-term memory of spoken discourse masked by speech: What is the role for working memory capacity? *Journal of Speech, Language, and Hearing Research*, 55, 210–218.
- Stewart, M. E., & Ota, M. (2008). Lexical effects on speech perception in individuals with “autistic” traits. *Cognition*, 109, 157–162.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 26, 212–215.
- Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first-and second-language multi-talker babble. *Speech Communication*, 52, 943–953.
- Van Engen, K. J. (2012). Speech-in-speech recognition: A training study. *Language and Cognitive Processes*, 27, 1089–1107.
- Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, 121, 519–526.
- Van Engen, K. J., Phelps, J. E., Smiljanic, R., & Chandrasekaran, B. (2014). Enhancing speech intelligibility: Interactions among context, modality, speech style, and masker. *Journal of Speech, Language, and Hearing Research*, 47, 1908–1918.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing: CTOPP* [Measurement instrument]. Austin, TX: Pro-Ed.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.
- Xie, Z., Maddox, W. T., Knopik, V. S., McGeary, J. E., & Chandrasekaran, B. (2015). Dopamine receptor D4 (DRD4) gene modulates the influence of informational masking on speech recognition. *Neuropsychologia*, 67, 121–131.
- Xie, Z., Yi, H. G., & Chandrasekaran, B. (2014). Nonnative audiovisual speech perception in noise: Dissociable effects of the speaker and listener. *PloS ONE*, 9, e114439.
- Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Rönnerberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America*, 134, 2225–2234.