




Listening back in time: Does attention to memory facilitate word-in-noise identification?

T. M. Vanessa Chan^{1,2}  · Claude Alain^{1,2}

Published online: 5 September 2018
© The Psychonomic Society, Inc. 2018

Abstract

The ephemeral nature of spoken words creates a challenge for oral communications where incoming speech sounds must be processed in relation to representations of just-perceived sounds stored in short-term memory. This can be particularly taxing in noisy environments where perception of speech is often impaired or initially incorrect. Usage of prior contextual information (e.g., a semantically related word) has been shown to improve speech in noise identification. In three experiments, we demonstrate a comparable effect of a semantically related cue word placed after an energetically masked target word in improving accuracy of target-word identification. This effect persisted irrespective of cue modality (visual or auditory cue word) and, in the case of cues after the target, lasted even when the cue word was presented up to 4 seconds after the target. The results are framed in the context of an attention to memory model that seeks to explain the cognitive and neural mechanisms behind processing of items in auditory memory.

Keywords Auditory · Attention · Retro-cue · Speech-in-noise · Context

Spoken words are short lived, lasting about a second and then disappearing. Therefore, verbal communication depends on our ability to maintain representations of these words in memory in order to generate sentences and extract meaning. Sequences of speech sounds build expectations based on their established structure and grammar. However, with the many sources of noise in the environment, initial encoding of speech sounds may be imprecise. Hence, in adverse listening situations, speech comprehension could benefit from the incorporation of subsequent contextual information in order to disambiguate speech representations held in short-term memory. This process of incorporating incoming speech sounds with

those held in memory can be investigated by directing attention to previous speech representations in auditory short-term memory (ASTM), a form of reflective attention (Chun, Golomb, & Turk-Browne, 2011).

Reflective attention is studied experimentally using a variant of the delayed match-to-sample task (Astle, Summerfield, Griffin, & Nobre, 2012; Backer & Alain, 2012; Backer, Binns, & Alain, 2015; Griffin & Nobre, 2003; Johnson et al., 2005; Lim, Wöstmann, & Obleser, 2015). Participants are first presented with an array of items (S1) that must be maintained in STM for comparison with a probe (S2) presented after a retention interval. In some trials, participants are presented with a cue after S1, referred to as the retro-cue, because it directs the participant's attention reflectively to a particular item(s) held in STM. In such paradigms, participants are typically more accurate and faster in determining whether the probe matches (or not) the item from the memory array when presented with an informative retro-cue in comparison with an uninformative retro-cue (Griffin & Nobre, 2003). The behavioral advantage of an informative retro-cue suggests that attention can successfully deploy to an item in STM, which brings it to the foreground thereby easing the comparison with the probe item(s) (Johnson et al., 2005).

Golestani and colleagues (Golestani, Hervais-Adelman, Obleser, & Scott, 2013; Golestani, Rosen, & Scott, 2009)

Public significance statement This study demonstrates that in noisy environments, listeners use contextual information presented after the fact to disambiguate words in noise. When developing training programs for listening in adverse situations, encouraging people to use subsequent context may be a useful strategy to incorporate.

✉ T. M. Vanessa Chan
vanessatm.chan@mail.utoronto.ca

¹ Department of Psychology, University of Toronto, Sidney Smith Building, 100 St. George St, Toronto, ON M5S 3G3, Canada

² Rotman Research Institute, Baycrest Health Sciences, Toronto, Ontario M6A 2E1, Canada

examined the role of a “retro-cue” on listeners’ ability to identify a target word embedded in speech-shaped noise. Participants were first presented with the word-in-noise, which was followed by a prime word in quiet and a two-alternative forced-choice question where participants had to select between the target word and a semantically (but not phonologically) related foil word. The effect of prime relatedness on accuracy was inconsistent between studies, with one study reporting a benefit from the prime word (Golestani et al., 2009), while another did not report a benefit in accuracy (Golestani et al., 2013); in the latter study, an effect of relatedness was nonetheless observed on response time, as well as increased activity in the left angular gyrus for semantically related words compared with unrelated words. While the task did not involve an array of multiple items to remember as in typical retro-cue studies, the participants still demonstrated an ability to effectively incorporate subsequent semantic information in the identification of words in noise.

The results from the work of Golestani and colleagues support the notion that speech-in-noise processing involves attention to representations held in memory. However, the inconsistency in behavioral results may be due to differences in strategy being employed in their task, since the relatedness effect on accuracy was not replicated between studies. Moreover, the visual presentation of the target and foil as response choices may have acted as a refresher to the auditory target originally embedded in noise. It was also possible in some cases for the participant to have gathered the semantic context intended by studying the relationship between the visual target and foil presented, which may result in a slower response time but still improved accuracy due to knowledge of the context. In order to test if contextual cues following degraded speech can help with understanding what was said, it is important to control when that context is presented, and remove the possibility of gathering that information through forced-choice response alternatives.

In this study, we tested the hypothesis that orienting attention reflectively to speech representations in memory does improve speech-in-noise identification. We employed a paradigm similar to that of Golestani and colleagues (Golestani et al., 2013; Golestani et al., 2009). Participants were presented with a word in white noise, which was either preceded or followed by a cue word in quiet. Participants were asked to name the word, with the aid of the cue, and we recorded their response using a microphone. We chose to use naming accuracy as our measure of interest instead of a forced-choice task, since it is more analogous to an everyday situation; naming eliminated the need to present the target and potential foils visually, which may serve as an additional or even sufficient cue in and of itself. This allows us to isolate the process of cuing focused attention from the moment of response decision itself.

The hypothesis that a semantically related word can act as a cue to refresh a representation of a word-in-noise, and aid in its

identification, was founded upon a model of auditory attention to memory (Zimmermann, Moscovitch, & Alain, 2016), based upon the initial model in vision (Cabeza et al., 2011; Ciaramelli, Grady, & Moscovitch, 2008). This model in audition notes several aspects that are distinct from how attention to memory is applied to the visual domain, due to the protracted time scale of auditory stimulus presentation and perception; in particular, auditory attention to memory involves a transformation from the initial sensory percept into a more abstract semantic representation in STM, and the ability to redirect attention to auditory representations also persists over a longer time period than in vision. As the attention to memory model in audition has thus far been examined only in nonverbal scenarios (Backer & Alain, 2012; Backer et al., 2015; Zimmermann, Moscovitch, & Alain, 2017), we intended to extend the model’s ability to explain processing of verbal stimuli, an important and commonly perceived set of sounds for humans that also carries semantic information. Importantly, in the previous studies that set out to examine auditory attention to memory, the auditory stimuli used carried high semantic value and could either be cued to focus on the basis of their semantic category (Backer et al., 2015) or were chosen to have semantic reliability such that memory associations would be easily made (Zimmermann et al., 2017), which is in agreement with the finding that semantic information plays a large role in auditory object representations (Gregg & Samuel, 2009).

To this effect, three experiments were conducted. Experiment 1 used a visually presented cue word that could either be presented before the target (as a pre-cue) or after it (as a retro-cue). Visual presentation of the cue was initially inspired by the visual presentation of retro-cues in nonverbal tasks of auditory reflective attention (Backer & Alain, 2012; Backer et al., 2015). Experiment 2 moved from using a visual cue word to an auditory cue word, in line with studies of speech perception and to better emulate a common listening environment. Experiment 3 looked specifically at the retro-cue condition and varied the length of delay between offset of the target and onset of the retro-cue to examine the effectiveness of the cue over a longer time period.

Experiment 1

Method

Participants

Participants were recruited from the research database in the Baycrest Hospital participant pool, as well as through advertisements and word of mouth, selected with the condition that they were young adults between the ages of 18 and 35 years, with self-reports of normal hearing and having learned English before the age of 5 years. English proficiency was

assessed with a questionnaire that included questions on age of acquisition, daily use, and self-rated proficiency. Hearing was assessed with an audiometric pure-tone threshold test, with thresholds ≤ 25 dB in both ears between the tested frequencies of 250 and 8000 Hz. This study was approved by the Research Ethics Board at Baycrest. All participants gave written informed consent prior to beginning the experiment and were monetarily compensated for their time at an hourly rate.

Sixteen participants (mean age 22.4 years, age range: 20–24 years, five males) participated in Experiment 1. Two additional participants were excluded from the analysis, one for exceeding the audiometric thresholds and another for a data error that resulted in the loss of recordings for verification by a second listener (see below). Effect sizes from the main effect of cue type/relatedness observed in Zekveld et al. (2011) and Golestani et al. (2009) were calculated, and the smaller of the two ($\eta_p^2 = 0.58$) was used in a power calculation using G*Power 3.1.9.2 (Faul, Erdfelder, Lang, & Buchner, 2007) to estimate the power of our study; in a repeated-measures design with five levels in our within-subject factor of task condition (see below), estimated power achieved was 0.967.

Stimulus and task

Stimuli were English words selected with the criterion that they had to be one to two syllables in length. Related word pairs were generated with the University of South Florida (USF) free association norms (Nelson, McEvoy, & Schreiber, 2004) and were selected based on the most related pair from the recorded words available. Words were recorded by two male and two female speakers of North American English in a soundproof room. One hundred and fifty pairs were generated in this manner and separated into two lists of 75 pairs, balanced for the level of forward and backward associations in the pairs as reported in the USF norms. Summary statistics for the stimuli were generated from the English Lexicon Project database (Balota et al., 2007) are presented in Table 1. Unrelated words were also selected for each pair, based on the remaining words available with the same criterion of one to two syllables. Word pairs and their corresponding unrelated word were matched for speaker gender and presented by the same speaker where possible (139 out of 150 pairs).

Sample related word pairs and unrelated words are presented in Table 2.

White noise was generated using MATLAB and converted to .wav format. Files were equated in total root mean square loudness also using MATLAB. Stimuli were presented using Presentation 16.3 (Neurobehavioral Systems), using Presentation's built-in attenuation procedures to adjust the signal-to-noise (SNR) ratio of the target words and white noise. The task was conducted in a sound-attenuated booth. White noise was presented through Etymotic ER-3A insert-earphones (Etymotic Research, Elk Grove, IL, USA) at a level of 80 dB sound pressure level (SPL), measured with a Larson Davis SPL meter (Model 824) using a 2-cc sampler. The task was coded such that depending on the version, one word of the pair would be the related cue and the other the target. A Cyber Acoustics ACM51-B microphone was used to record verbal responses as well as the time of response.

Five cue conditions were presented: No-cue, where the word was presented without any words preceding or following it; before-related cue, where the paired cue word was presented before the target in noise; after related cue, where the paired cue word was presented after the target in noise; and before unrelated and after unrelated, similar to the above, except that the cue word presented was not semantically related to the target. Cuing speech-in-noise perception with semantically related words has previously been demonstrated to be effective in bolstering accuracy (Zekveld et al., 2011); therefore, the inclusion of both precues and retro-cues in the study is to facilitate a within-subject comparison of both cue types in the same study design. Each condition was presented 15 times in a block of 75 trials, resulting in 150 trials over two blocks of trials. The order of conditions, as well as which word pairs were used for each condition, was randomized for each participant so that adaptive strategies could not be formed within a specific block of trials of a given condition. Cue words were visually presented for 1 second as white characters in the center of a black screen. They were presented in size 40-point Times New Roman Font, with participants sitting 60 cm from the screen.

Participants were instructed to repeat back words that were presented in white noise, with the help of cue words that would appear on the screen either before or after the word-

Table 1 Summary statistics for the words used in Experiments 1–3

Measure	List 1			List 2		
	Target 1	Target 2	Unrelated	Target 1	Target 2	Unrelated
Number of syllables	1.75 (0.44)	1.48 (0.50)	1.55 (0.50)	1.63 (0.49)	1.39 (0.49)	1.65 (0.51)
Log frequency (HAL)	8.94 (1.11)	9.17 (0.90)	8.81 (0.98)	9.11 (0.87)	9.18 (0.85)	8.82 (0.94)

Values presented are the mean, and values in parentheses are the standard deviation. HAL = Hyperspace Analogue to Language frequency norms (Lund & Burgess, 1996)

Table 2 Sample word pairs used in Experiments 1–3

Related Word 1	Related Word 2	Unrelated word
Aim	Targets	Springs
Actor	Movie	Beard
Hunter	Killing	Motel
Ocean	Tide	Insight
Suspect	Crime	Textile
Victim	Accused	Folklore

Word pairs and the unrelated word in the set are read horizontally (e.g., *Aim* and *Targets* are a related word pair, with *Springs* being the unrelated word). Each of the related words in the pair is set to be the cue or target, depending on the version of the task presented. Unrelated words were always cues only

in-noise. They were told that the cues may or may not be related to the target, but were encouraged to use them to help guide their listening process. All trials began with a 3-second visual countdown. On before-related and before-unrelated trials, participants were presented the cue word for one second, followed by a 500-ms pause with fixation cross, and then onset of white noise with the fixation cross maintained. After 1 second of white noise, the target word was presented, and visually denoted by the fixation cross turning green. On after-related and after-unrelated trials, the presentation of target and cue was reversed from that of the “before” conditions. The no-cue conditions began after the countdown with onset of white noise and target presentation. Participants were given 1 second to prepare their response, and a 2-second window to say their response into the microphone, prompted by instructions presented on the screen. The schematic of the task is presented in Fig. 1.

The task was presented in two blocks, with each block differing in the SNR at which the target word was played against the white noise: one at a SNR of zero dB, the other at -5 dB. The SNRs were selected to encourage the usage of the cue for younger participants of normal hearing. Previous work has shown a rather large difference in intelligibility between SNRs of zero dB and -5 dB (Davis, Ford, Kherif, & Johnsrude, 2011; see also Du, Buchsbaum, Grady, & Alain, 2014).

The SNR, as well as the cue version and list used for each block, was counterbalanced across participants. The second block always used the other of the two lists of words, to avoid any carryover effects of the previous block. Trials were self-paced, and the task took about 30 minutes to complete.

Experimental procedure

Participants were seated in a sound-proof room with auditory stimuli presented through Etymotic ER3A insert earbuds. Prior to all three experiments, a series of hearing and memory assessments were conducted with each participant, which were standard procedure within our lab. These included

audiometric pure-tone hearing threshold; standardized measures of speech-in-noise comprehension including the QuickSIN (Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004) and Words-In-Noise (WIN; Wilson, 2003) tests; digit-span subtest of the Wechsler Memory Scale–III (WMS-III; Wechsler, 1997); and a two-back memory test of digits, presented first in the auditory modality and then in the visual modality.

Data analysis

Statistical analyses were conducted using R Version 3.4.1 (R Core Team, 2017) and package “ez” (Lawrence, 2016). Generalized eta scores (η_G^2) are reported as a measure of effect size. Post hoc tests of analysis of variance (ANOVA) results were conducted using pair-wise *t* tests with Holm’s correction for multiple comparisons. For all analyses, tests of significance were two-tailed, with a significance level of $p < .05$, after correction, required to achieve statistical significance.

Participants’ responses to each word-in-noise presented were recorded both by the microphone linked to the Presentation software and also by a separate recording device placed in the soundproof room. Two listeners, one of which was the experimenter, transcribed the recordings, which the experimenter then evaluated for accuracy. Complete matches were given a mark of 1, singular/plural deviations were given a mark of 0.5, and all remaining responses (including silence or “don’t know”) were given a mark of zero. The score reported is the average of the evaluations of both transcriptions. To ensure interrater reliability, the resulting scores for each participant and condition were subjected to an intraclass correlation analysis using the ICC function from the package “psy” in R (Falissard, 2012); the mean agreement coefficient across conditions was 0.80 (standard deviation \pm 0.11) in Experiment 1. Response times were also collected, but differences in preparation time between the conditions made the interpretation of these data difficult and unreliable, and so they are not included in the analysis.

Percentage of correct phonemes was calculated by converting one of the transcriptions of each participant’s responses into phonemes using the English Lexicon Project database and comparing these phonemes to the phonemes for the target words. Phonemes were scored as correct if they were positioned in the same position before or after the stressed syllable in the word. Additional phonemes were not deducted from the score.

Results

A 5×2 (Task Condition \times SNR) repeated-measures ANOVA was conducted to evaluate the benefit of cues on the accuracy of identifying words in noise. There was a main effect of task condition, $F(4, 60) = 54.23$, $p < .001$, $\eta_G^2 = 0.51$, and SNR,

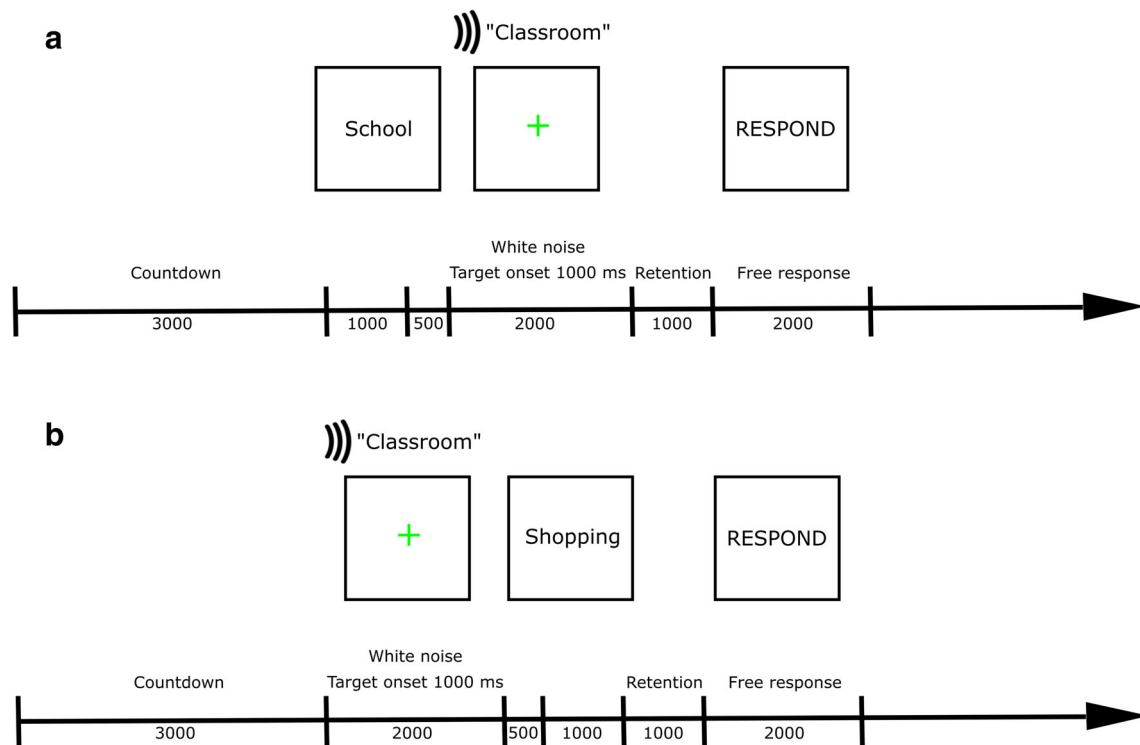


Fig. 1 The general schematic for the experimental task. For both diagrams, time (along the black arrow) is listed in milliseconds. **a** Example of a before-related cue condition. **b** Example of an after-unrelated cue condition

$F(1, 15) = 82.78, p < .001, \eta_G^2 = 0.49$. Post hoc pair-wise t tests for task condition (with Holm's correction for multiple comparisons) showed that while before unrelated and after unrelated did not significantly differ from no-cue (before unrelated, $p = .700$; after unrelated, $p = .860$), before-related ($p < .001$), and after related ($p < 0.001$) significantly improved accuracy from no-cue while not significantly differing from one another ($p = .590$). The results of this analysis are visualized in Fig. 2. Similar to the analyses conducted for accuracy score, a 5×2 (Task Condition \times SNR) repeated-measures ANOVA was conducted on percentage of correct phonemes. There was also a main effect of task, $F(4, 60) = 35.07, p < .001, \eta_G^2 = 0.42$, and SNR, $F(1, 15) = 156.65, p < .001, \eta_G^2 = 0.57$, as well as a Task \times SNR interaction, $F(4, 60) = 3.73, p = .009$.

To further investigate the independent effects of cue relatedness and cue position, a second analysis was conducted after removing the no-cue conditions from the data set, which allowed for the effects of cue relatedness and position to be entered as factors. A $2 \times 2 \times 2$ (Cue Position \times Cue Relatedness \times SNR) repeated-measures ANOVA was conducted to evaluate the effect of relatedness and position of the cue on target identification. There was a main effect of cue relatedness, $F(1, 15) = 138.28, p < .001, \eta_G^2 = 0.49$, and of SNR, $F(1, 15) = 80.73, p < .001, \eta_G^2 = 0.50$. The Cue Position \times SNR interaction approached significance, $F(1, 15) = 3.25, p = .091, \eta_G^2 = 0.02$, as did the three-way Cue Position \times Cue Relatedness \times SNR interaction, $F(1, 15) = 3.34, p = .088, \eta_G^2 = 0.03$.

For percentage of correct phonemes, when removing the no-cue conditions from the data set, a $2 \times 2 \times 2$ (Cue Position \times Cue Relatedness \times SNR) repeated-measures ANOVA revealed a main effect of cue relatedness, $F(1, 15) = 111.48, p < .001, \eta_G^2 = 0.41$, and SNR, $F(1, 15) = 146.63, p < .001, \eta_G^2 = 0.59$, as well as a Cue Relatedness \times SNR interaction, $F(1, 15) = 9.76, p = .007, \eta_G^2 = 0.06$. This interaction appears to be driven by the increased benefit of cue relatedness on percentage of correct phonemes in the SNR -5 condition compared with the SNR zero condition.

Discussion

We observed a benefit of visually presenting a semantically related cue word on the identification of an auditory target word embedded in white noise, and this benefit was present when the cue was presented either before or after the target. These results are in line with those from Zekveld et al. (2011) as well as Golestani et al. (2009), where semantically related precues and retro-cues, respectively, bolstered performance on a speech-in-noise task. Effects of SNR were also observed, which is consistent with previous findings in the above studies; participants were less accurate when the SNR was lower. The benefit of the after cue on word-in-noise identification is consistent with the hypothesis that listeners use subsequently presented contextual information to disambiguate speech sounds in adverse listening situations. The after cue may help listeners to focus attention on some acoustic details and/or

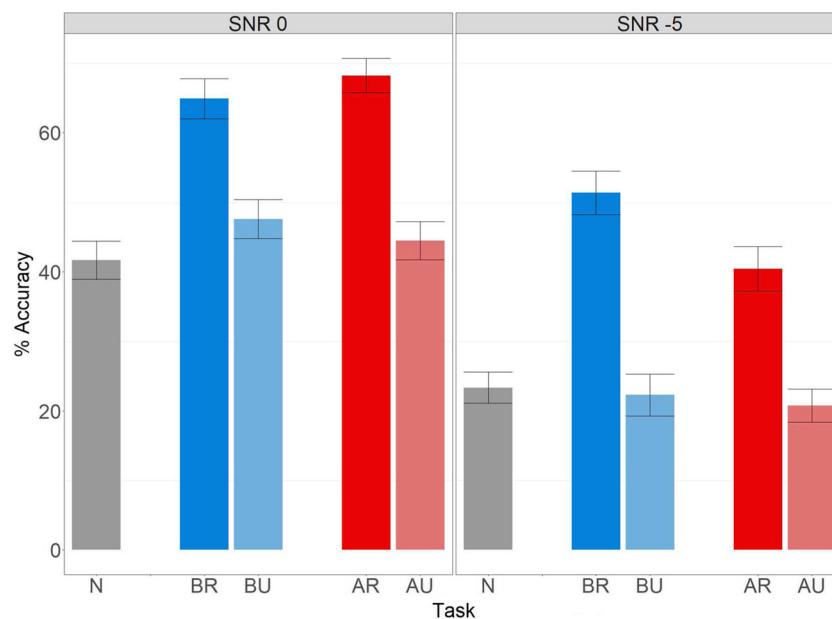


Fig. 2 Percentage of the accuracy of word recognition in Experiment 1. Task type denotes the five cue and target presentation conditions. SNR = signal-to-noise ratio; N = no-cue; BR = before-related; BU = before

unrelated; AR = after related; AU = after unrelated. Error bars indicate the standard error of the mean (*SEM*)

provide response alternatives against which the degraded speech representation in ASTM can be compared.

There was a trend for the before cue to be more effective in improving response accuracy than the retro-cue in the more challenging SNR condition. This may be because the noise level was such that in some cases it precluded the formation of a representation for the cue to refresh, reducing the usefulness of the cue, whereas in the higher SNR condition the word was still distinguishable enough to allow the context provided by the after cue to be useful. The before cue, on the other hand, may have acted as a prime, creating an activation “spread” to related words, thereby easing word-in-noise identification (Kalikow, Stevens, & Elliott, 1977; Pichora-Fuller, Schneider, & Daneman, 1995; Sheldon, Pichora-Fuller, & Schneider, 2008; Zekveld et al., 2011). However, since these interaction trends are small, these results are not sufficient to suggest a difference in cue usage based on cue position.

We have thus far demonstrated that a cue word presented before or after a word-in-noise improved accuracy. The addition of the no-cue condition in this task ensured that it was not merely the conflict arising from unrelated semantic cues that was decreasing accuracy in the unrelated conditions, but that there was also a benefit associated with presenting related semantic cues.

Experiment 2

Experiment 1 demonstrated a cross-modal benefit, where the visual presentation of a semantically related cue word enhanced word-in-noise identification relative to the

presentation of an unrelated word or no word at all. To build on this experiment and to better emulate a speaking environment, we changed the modality of the cue word from a visual cue to an auditory cue and sought to replicate the results of Experiment 1 in an otherwise unchanged design.

Method

Participants

Sixteen participants (mean age 22.3 years, age range: 18–35 years, eight males) participated in Experiment 2. None of the participants who completed Experiment 2 participated in Experiment 1. Criteria for participation and recruitment procedure were the same as in Experiment 1. One additional participant was excluded for exceeding the audiometric thresholds. Estimated power was as in Experiment 1.

Stimulus and task

The stimuli and task used in Experiment 2 were the same as in Experiment 1, except that the cue words were now also presented in the auditory modality at the same overall intensity as the target. Cue duration remained the same as in Experiment 1, since the longest recordings were still 1 second in length. Participants were now instructed that the cue is a spoken word in quiet, which could occur either before or after the word-in-noise.

Experimental procedure

The procedure was identical to that of Experiment 1. Participants were presented with two blocks of trials, one for each SNR level. As for Experiment 1, Experiment 2 included hearing and memory assessments.

Data analysis

Analyses were conducted with the same specifications as in Experiment 1. To compare the effects of cue modality, we also performed an additional analysis that combined the data sets of Experiment 1 and Experiment 2, with cue modality (visual or auditory) as a between-subjects variable. As in Experiment 1, an ICC analysis was conducted to ensure interrater reliability; the mean agreement coefficient was 0.79 (± 0.15) in Experiment 2.

Results

A 5×2 (Task Condition \times SNR) repeated-measures ANOVA was conducted in the same manner as Experiment 1. There was a main effect of task condition, $F(4, 60) = 36.51, p < .001, \eta_G^2 = 0.43$, and SNR, $F(1, 15) = 84.17, p < .001, \eta_G^2 = 0.34$. Post hoc pair-wise t tests for task condition showed that the before-related and after-related cue (BR: $p < .001$; AR: $p = .002$) significantly improved accuracy from no-cue, while before unrelated and after unrelated did not significantly differ from no-cue (both $ps > 0.1$). The results of this analysis are visualized in Fig. 3. Similar to accuracy, the ANOVA on percentage of correct phonemes showed a main effect of task,

$F(4, 60) = 20.16, p < .001, \eta_G^2 = 0.31$, and SNR, $F(1, 15) = 135.24, p < .001, \eta_G^2 = 0.38$, but no Task \times SNR interaction.

After removing the no-cue conditions from the dataset, a $2 \times 2 \times 2$ (Cue Position \times Cue Relatedness \times SNR) repeated-measures ANOVA was conducted to evaluate the contributions of various properties of the cue and target in target identification. There was a main effect of cue relatedness, $F(1, 15) = 73.95, p < .001, \eta_G^2 = 0.45$, and SNR, $F(1, 15) = 59.12, p < .001, \eta_G^2 = 0.37$. The main effect of cue position approached significance, $F(1, 15) = 4.23, p = .057, \eta_G^2 = 0.02$. There was also a Cue Relatedness \times Cue Position interaction, $F(1, 15) = 9.09, p = .009, \eta_G^2 = 0.03$. Pair-wise t tests revealed that while before-related was significantly different from after-related ($p = .017$), before-unrelated and after-unrelated were not ($p = .661$). None of the other two-way or three-way interactions were significant. With percentage of correct phonemes, the same analysis conducted revealed a main effect of cue relatedness, $F(1, 15) = 67.63, p < .001, \eta_G^2 = 0.34$, and SNR, $F(1, 15) = 74.96, p < .001, \eta_G^2 = 0.36$, but no interaction effects.

Comparison between cue modalities

In order to examine any potential differences in identifying a word-in-noise with a visual cue word compared with an auditory cue word, the data from Experiments 1 and 2 were collapsed into a single analysis. The no-cue condition was excluded from the ANOVA because we were interested in a potential modality difference with respect to the cue position or relatedness. There were main effects of cue relatedness, $F(1, 30) = 192.32, p < .001, \eta_G^2 = 0.47$; cue position, $F(1, 30) = 6.71, p = .015, \eta_G^2 = 0.02$; and SNR, $F(1, 30) = 139.29, p < .001, \eta_G^2 = 0.43$, as well as an interaction of cue

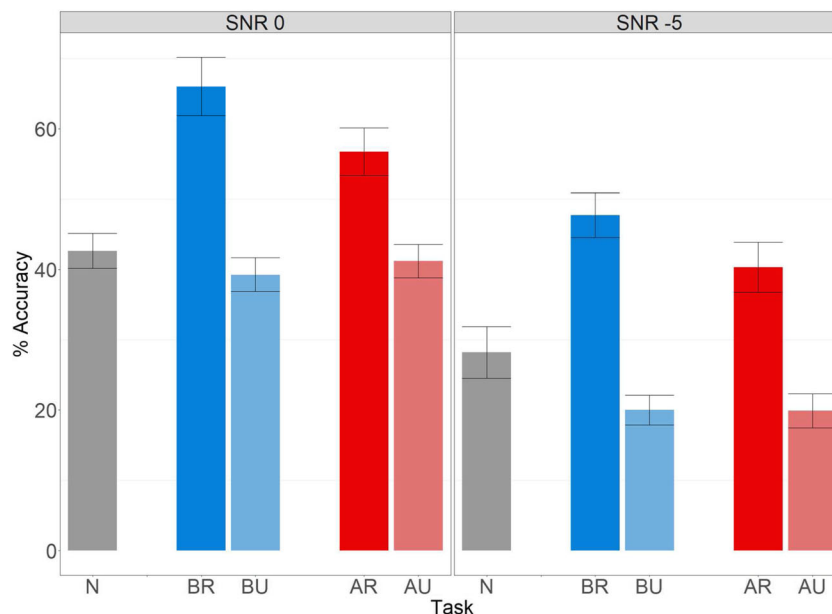


Fig. 3 Percentage of the accuracy of word recognition in Experiment 2. Task type denotes the five cue and target presentation conditions. SNR = signal-to-noise ratio; N = no-cue; BR = before-related; BU = before unrelated; AR = after related; AU = after unrelated. Error bars indicate the SEM

relatedness and cue position, $F(1, 30) = 4.85, p = .04, \eta_G^2 = 0.01$. The main effect of cue modality approached significance, $F(1, 30) = 3.60, p = .067, \eta_G^2 = 0.02$, but none of the interactions between cue modality and the other factors were significant.

Discussion

As in Experiment 1, the presentation of a semantically related cue word improved word-in-noise identification, and this effect was present when the cue was presented before or after the target word. Again, the results are in line with the findings of Golestani et al. (2009), where a relatedness effect was observed with the presentation of cue words after the target, and extends the precuing effect observed in Zekveld et al. (2011) to cue words presented in the auditory modality. Importantly, the improvement in accuracy observed with semantically related cue words is also internally consistent with our findings in Experiment 1, with no group differences observed when the two groups were entered as a between-subjects variable. While the cuing effect with auditory cue words is unsurprising, given the sentence-level context effects observed in studies of speech-in-noise processing (Kalikow et al., 1977; Pichora-Fuller et al., 1995), the usage of otherwise identical materials across the first two experiments allows us to both confirm the efficacy of the task and compare the usage of visual and auditory cues.

The comparison of performance between Experiment 1 (visual cue) and Experiment 2 (auditory cue) revealed no significant differences, although the main effect of modality approached significance and in favor of the visual-cue condition. Two possible paths could emerge for any potential effect of modality. Visually presented words were shown in their entirety from the onset of the cue presentation period, and so participants may have had more time to process the semantic information out of the cue than in an auditory-cue condition where the meaning of the cue may not be clear until most or all of the cue presentation period is over, giving the visual cue a time advantage. Words presented visually also have the benefit of disambiguating any potential homophones arising from auditory cues and clarifying the semantic context more effectively in a single-word cue. However, seeing as the effects did not actually reach statistical significance and were small in effect size, the relative contributions of such putative modality effects are likely minimal. This is consistent with a study showing that any change in semantic interpretation of an ambiguous word is not dependent on the modality of the prime being presented (i.e., the cross-modal prime-target cuing effect was similar to that of unimodal prime-target cuing; Gilbert, Davis, Gaskell, & Rodd, 2018). Therefore, it is possible that any semantic activation that was achieved with our present task is tapping into a modality-independent representation of the target word-in-noise.

In contrast to the results of Experiment 1, a Cue Position \times Cue Relatedness interaction was observed, where accuracy was higher in the before-related condition compared with the after-related condition across both SNR levels. Visual inspection suggests that the behavior in the SNR -5 condition is similar across both experiments, but not for the SNR zero condition. However, it is of note that when collapsing across both Experiment 1 and 2, the Cue Position \times Cue Relatedness interaction was also significant, with no significant modulation by cue modality. This result suggests that correctly identifying a target word-in-noise is enhanced by prior semantic knowledge more than receiving the semantic knowledge afterwards and having to recover the target. While the strength of forward cuing due to semantic relatedness has been repeatedly demonstrated in the domain of speech-in-noise perception (Kalikow et al., 1977; Pichora-Fuller et al., 1995; Sheldon et al., 2008; Zekveld et al., 2011), to our knowledge this is the first study to use semantically manipulated cue words both before and after target words in a single design, allowing for a direct comparison of their efficacy on word-in-noise identification. Importantly, while the after-related cue did not boost accuracy as strongly as the before-related cue, it was still more helpful than an unrelated cue or no-cue at all, which suggests that participants are still able to use the cue in a useful manner, although perhaps not as efficiently as when it is presented before the target.

The presentation of pairs of words is not unlike that used by studies of semantic priming, where the semantic relatedness of a previously presented prime word shows a facilitative effect on the recognition of a probe word, usually on a task of lexical decision (Holcomb, 1988; Meyer & Schvaneveldt, 1971). While primarily conducted in the visual domain, studies have also shown priming effects in the auditory domain (Holcomb & Neville, 1990), and identification of a brief or obstructed prime has also been facilitated with relatedness to the target (Bernstein, Bissonnette, Vyas, & Barclay, 1989).

However, rather than the automatic processes that are purported to be involved in semantic priming, such as spreading activation (Meyer & Schvaneveldt, 1971), we were interested in the direction of attention to mental representations. We attempted this by encouraging the active use of cues through instructions (Holcomb, 1988) and by introducing a longer stimulus onset asynchrony (SOA) than usually reported by studies of semantic priming, which was also necessitated by the nature of auditory stimuli needing time to be presented. Longer SOAs have been associated with a conscious attentional process (Neely, 1977; Rossell, Bullmore, Williams, & David, 2001), although this may also occur at shorter SOAs in individuals with high levels of attentional control (Hutchison, 2007). Therefore, we wanted to delineate the usefulness of a semantically related cue word at various delays after presentation of the degraded target, to examine how delay length would affect a listener's representation of the target, which

the cue could then guide attention to refresh or reinterpret the item in ASTM.

Experiment 3

In Experiment 3, we sought to further characterize the cuing effect we observed. Is the cue effective only because of its temporal proximity, or can it also be effective after a period of actively maintaining the degraded target? Conversations are ongoing streams of information, but a noisy environment may preclude the usage of that continual context until a less masked word or sentence is perceived, which may not necessarily occur close in time to the word-in-noise that is at first unclear to the listener. We anticipated that if the effectiveness of a semantically related cue is sensitive to its timing relative to that of the target, such that it is beneficial for them to be closer together, accuracy should decrease with an increased delay between the target and cue. On the other hand, if the timing of the cue is irrelevant to its usefulness in its ability to aid in identifying the target, then accuracy should not significantly differ across timing delays.

We focus here on the after cue condition from Experiments 1 and 2 and manipulate the temporal delay of this condition for several reasons. Firstly, it is the condition of most relevance to the attention to memory model, and allows us to compare present findings with that of previous work falling under this framework. Secondly, Backer and Alain (2012) found that retro-cuing was beneficial after at least 4 seconds of delay in their paradigm. Finally, since before-related cues are more effective than after-related cues (as found in the Cue Position \times Cue Relatedness effect from Experiment 2), any effect of cue relatedness found in just the after-cue condition can be generalized to the before-cue condition.

Method

Participants

Twenty-four participants (mean age 24.3 years, age range: 19–33 years, nine males) participated in Experiment 3. Criteria for participation and recruitment procedure were the same as in Experiment 1, with the added condition that they did not previously participate in Experiment 1 or 2. One additional participant was excluded for exceeding the audiometric thresholds. Using the same effect size for the power calculation as in Experiment 1, the estimated power with four within-subject levels for cue time (see below) was 0.999.

Stimulus and task

Stimuli used in Experiment 3 were the same as in Experiment 2, except that three sets of words from each list were removed

to accommodate for the number of trials needed to balance the conditions. The cue was always presented in the auditory modality and after the target (emulating the after-related and after-unrelated conditions from Experiment 2), but the time between presentation of target and cue was varied over four timing conditions: 500 ms (like in Experiments 1 and 2), 1 s, 2 s and 4 s. Due to its similarity in performance to the unrelated condition in Experiments 1 and 2, the no-cue condition was removed in Experiment 3. Participants were again instructed to withhold their response until prompted. Each timing condition was crossed with two cue conditions, for a total of eight conditions. Since the amount of time given for the usage of a retro-cue has been shown to affect performance (Backer & Alain, 2012), the amount of time between retro-cue offset and response window onset was kept consistent with Experiments 1 and 2. Each condition was presented nine times in a block, resulting in 72 trials per block, with the conditions randomized within a given block such that participants could not anticipate a given delay. All other aspects of the task remained the same between this experiment and Experiments 1 and 2.

Experimental procedure

Since the number of task dimensions had to be reduced in order to have enough trials for each of the four cue delay lengths, the experimental task was presented only at one SNR (−5 dB) throughout both blocks of this task. The more difficult of the SNR conditions from Experiments 1 and 2 was selected to ensure a sufficient level of difficulty, and to encourage participants to use the cue in completing the task despite the longer delay intervals being implemented. The cue version and list used for each block was counterbalanced across participants, and both blocks employed a different list. All other procedures, including the hearing and memory assessments, remained the same as in Experiments 1 and 2.

Data analysis

Analyses were conducted with the same specifications as in Experiment 1. The ICC mean agreement coefficient for interrater reliability was 0.92 (\pm 0.04).

Results

A 4×2 (Cue Time \times Cue Relatedness) ANOVA yielded a main effect of cue relatedness, $F(1, 23) = 134.20$, $p < .001$, $\eta_G^2 = 0.44$, and a Cue Relatedness \times Cue Time interaction, $F(3, 69) = 4.58$, $p = .006$, $\eta_G^2 = 0.07$. Pair-wise t tests showed that the interaction effect was driven by differences in word accuracy when comparing the effect of a related word cue across the time conditions; this difference approached significance when comparing cue presentation at 500 ms and 2,000

ms, $p = .051$. None of the other comparisons between the different cue times at each level of relatedness were significant. The results are visualized in Fig. 4. A 4×2 (Cue Time \times Cue Relatedness) ANOVA on the percentage of correct phonemes showed a main effect of cue relatedness, $F(1, 23) = 55.41$, $p < .001$, $\eta_G^2 = 0.17$, and a Cue Time \times Cue Relatedness interaction, $F(3, 69) = 2.81$, $p = .046$, $\eta_G^2 = 0.02$. Pair-wise t tests revealed that there was no significant difference in percentage of correct phonemes between related and unrelated conditions at the 2,000-ms delay ($p = .718$), despite a difference at all other delay conditions (all $ps < .05$).

Discussion

As in Experiments 1 and 2, participants in Experiment 3 were more accurate in identifying a word-in-noise when the target word was followed by a semantically related word compared with an unrelated word. This after cue effect of relatedness extended to up to 4 seconds posttarget presentation. Importantly, the differences in timing presented here are with respect to the interstimulus interval (ISI), and not the SOA, as is often the case in studies of semantic priming; with respect to the SOA, even the shortest timing condition in Experiment 3 (500 ms ISI) is equivalent to a 1,500 ms SOA, which is longer than most priming studies. Therefore, along with the instructions given to pay attention to the cue sound as well, it is likely that attentional processes were engaged in the performance of this task. That the related cue word could still facilitate identification at 4,000 ms after target presentation is less surprising if we consider that participants are no longer relying on sensory information, but instead on rehearsal of a verbal representation of what they initially perceived the target word to be.

In this case, at the point of cue perception, the rehearsed target would be compared with the cue word and allow for either a verification or a reassessment of the word that was initially perceived.

The decrease in accuracy when a related cue was presented at 2,000 ms after stimulus offset, compared with other cue timings, was unexpected. If there were a decrease in accuracy due to the delay of the cue, it would have been anticipated that this decrease occurred in a manner proportional to the amount of delay, such that accuracy would be lowest when presenting a related cue at 4,000 ms posttarget offset. However, performance with a related cue presented 4,000 ms posttarget was similar to that at 500 and 1,000 ms, which would support the hypothesis that the effect of semantic relatedness of the cue was invariant to the temporal delay of the cue. The observed pattern of results does not appear consistent with either hypothesis, but a replication of this particular finding would be beneficial before giving additional consideration to this particular interaction effect. However, it is possible that the change in performance marks a switch in strategy. We attempted to examine changes in strategy by looking at changes in error distribution (see Additional Analyses, below).

With respect to the relationship between these findings and that of the previous two experiments, our observation of a replicated relatedness effect for the after-cue conditions alone suggests that whatever representations were maintained in STM from the noisy target were sufficient for the cue to have some level of efficacy, even beyond the time frame dictated in Experiments 1 and 2. Since it is easier to maintain a clear speech representation than a noisy or distorted one (Hervais-Adelman, Davis, Johnsrude, & Carlyon, 2008), we can anticipate that a hypothetical before-cue condition in this

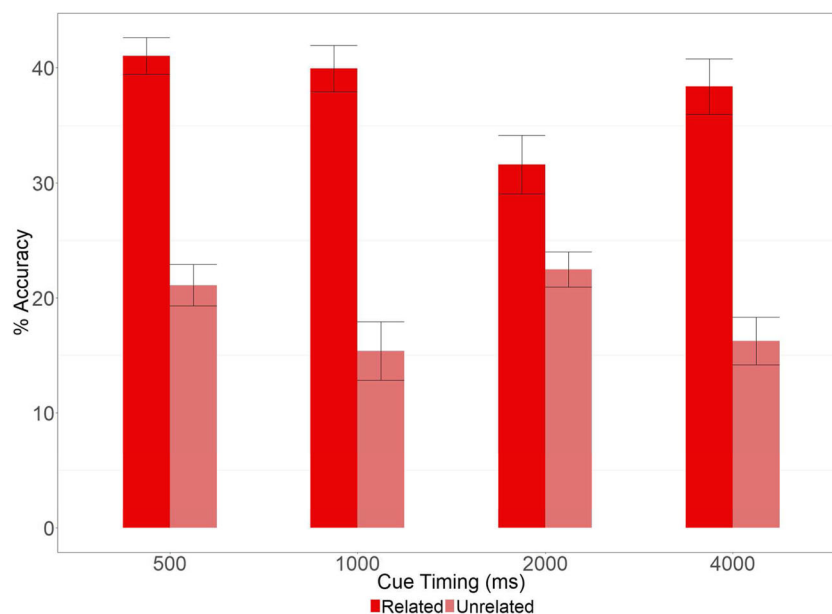


Fig. 4 Percentage of the accuracy of word recognition in Experiment 3. Cue timing is presented in milliseconds. Error bars indicate the *SEM*

experiment would also show a relatedness effect that is invariant across temporal delays between the cue and the target. This comparison across delays for both before and after cues is worth future consideration.

Additional analyses

In order to further characterize the pattern of responses and potential cognitive functions associated with task performance, we conducted two additional analyses with the data across all three experiments. With respect to certain models of speech processing that implicate the involvement of working memory such as the ease of language understanding model (Rönnberg et al., 2013), we conducted correlations between the standard working memory measures collected and performance on the experimental task. To further categorize the types of responses that were being made by participants, an error analysis was conducted classifying the types of errors being made under each condition.

Correlational analysis with standard working memory measures

An analysis was conducted to determine whether performance on the experimental task correlated with the standard working memory assessments conducted. To avoid correcting for multiple comparisons, an aggregate score of the collected working memory measures was calculated, from averaging percentage scores from the digit span and two-back tests. For the digit-span test, the longest span length reached for each participant was divided by the maximum span length, in both forward and backward span tests. For the two-back test, a hits minus false alarms percentage was calculated for each modality. Thus, each aggregate score was made of an average of four scores: the forward span length, backward span length, auditory two-back, and visual two-back.

In order to maximize the number of participants used in the analysis, we selected accuracy scores for certain conditions in Experiments 1 and 2 such that they were best representative of the conditions in Experiment 3. Thus, we conducted the correlations on the SNR -5 scores in Experiments 1 and 2 and averaged the accuracy scores across the four delays in Experiment 3 to generate one score for the after-related and after-unrelated cue conditions. This made for a correlation with 56 participants in total using the after-related and after-unrelated scores. Because there were no no-cue, before-related, and before-unrelated conditions presented in Experiment 3, correlations conducted with these three conditions were done on only the participants of Experiments 1 and 2, for a total of 32 participants.

The results of the correlational analyses are presented in Table 3. None of the correlations produced using the aggregate score were significant.

Correlational analysis with standard speech-in-noise measures

Similar to the analysis for the working memory measures, a correlational analysis was done to compare performance on the experimental task with performance on the QuickSIN and WIN tests. Because participants were all young and of normal hearing, performance on the standardized tests was generally high; scores on the QuickSIN and WIN were log transformed before the correlations were completed to more fully capture individual differences in performance on these tests. The same scores on the experimental task were used for this analysis as in the working memory correlational analysis. Corrections were done for multiple comparisons across all 10 correlations (five task conditions \times 2 speech-in-noise tests).

Tables 4 and 5 summarize the results of these correlational analyses for the QuickSIN and WIN, respectively. Performance on the QuickSIN correlated with performance in the before-related and after-unrelated conditions, while the correlation between after-unrelated and WIN performance was marginally significant. This relationship held even when the after-unrelated correlations were done with only the participants from Experiments 1 and 2.

Visual inspection of the data revealed several outliers that appeared to drive the correlations observed. Thus, the correlations were redone after removing participants who performed two standard deviations above or below the mean in the before-related and after-unrelated conditions. After this, the correlation of after-unrelated with QuickSIN performance disappeared, although the correlation of before-related with QuickSIN remained significant ($p = .024$). Although the sign value of the correlation is negative, this denotes a positive relationship between QuickSIN performance and accuracy on the before-related condition, since better performance on the QuickSIN results in a lower SNR loss value.

Table 3 Correlation between performance on experimental task and aggregate working memory score

Cue position	Related			Unrelated		
	<i>df</i>	<i>r</i>	<i>p</i>	<i>df</i>	<i>r</i>	<i>p</i>
Before	30	.029	.875	30	.268	.137
After	54	.059	.667	54	.109	.424
None	30	-.101	.584			

The no-cue condition falls under neither related nor unrelated cue category since there is no-cue word present at all, and it is only listed under “Related” for convenience

Table 4 Correlation between performance on experimental task and log-transformed QuickSIN score

Cue position	Related			Unrelated		
	<i>df</i>	<i>r</i>	<i>p</i>	<i>df</i>	<i>r</i>	<i>p</i>
Before	30	-.547	.012	30	-.078	1.000
After	54	-.204	.716	54	-.409	.016
None	30	.281	.716			

The no-cue condition falls under neither related nor unrelated cue category since there is no cue word present at all, and it is only listed under “Related” for convenience. Reported *p* values have been corrected for multiple comparisons (across both QuickSIN and WIN measures)

Error type analysis

Using one set of transcriptions, each participant’s responses were categorized as either correct or one of five types of errors: phonemic (partially matching phonemes with target), semantic (related to cue but not the target), both (semantically related to the cue but also has partially matching phonemes to target), other (responses that did not fit any of the above), and omission (when a participant didn’t respond or said “I don’t know”). Errors of each type were tallied up for each task and SNR condition, and a proportion score was generated for each participant by dividing the number of errors in each condition by the total number of errors committed for each participant. The proportion score was used as a measure of taking into account the total number of errors made by each participant when looking at how many of a given type were made under a certain condition.

For Experiments 1 and 2, similar to the accuracy calculations, a 5×2 (Task Condition \times SNR) repeated-measures ANOVA was first conducted, followed by a $2 \times 2 \times 2$ (Cue Position \times Cue Relatedness \times SNR) repeated-measures ANOVA after removing the no-cue conditions. The second ANOVA was particularly important for the conditions of semantic errors and both errors, since they could not be scored in No-cue conditions on account of those conditions not having a cue that could semantically bias a response. The ANOVAs

Table 5 Correlation between performance on experimental task and log-transformed WIN score

Cue position	Related			Unrelated		
	<i>df</i>	<i>r</i>	<i>p</i>	<i>df</i>	<i>r</i>	<i>p</i>
Before	30	-.307	.606	30	-.249	.716
After	54	-.016	1.000	54	-.345	.073
None	30	-.041	1.000			

The no-cue condition falls under neither related nor unrelated cue category since there is no cue word present at all, and it is only listed under “Related” for convenience. Reported *p* values have been corrected for multiple comparisons (across both QuickSIN and WIN measures)

were conducted separately on each error type while correcting for multiple comparisons (three comparisons in the 5×2 ANOVA, and five in the $2 \times 2 \times 2$ ANOVA). In the same manner, Experiment 3, a 4×2 [Cue Time \times Cue Relatedness] ANOVA was conducted on each error type and corrected for multiple comparisons.

Experiment 1 In the 5×2 ANOVA, a main effect of task condition was found for phonemic ($p < .001$), other ($p < .001$), and omission errors ($p = .001$), while the main effect of SNR approached significance for phonemic and omission errors (both $ps = .067$) and was significant for other errors ($p < .001$), with a greater proportion of errors occurring in the SNR -5 condition. The Task Condition \times SNR interaction was not significant. Pair-wise *t* tests showed that for the phonemic errors, a significantly greater proportion occurred in the no-cue condition than either of the related conditions (vs. before-related and after related, $p < .001$) as well as the before unrelated condition ($p = .021$). For the other errors, there was a significant difference between no-cue and the related conditions (BR: $p = .002$, AR: $p = .019$), but not the unrelated conditions (BU: $p = .352$, AU: $p = .926$). There were no significant differences between the no-cue and any of the cued conditions on omission errors (all $ps > .1$).

In the $2 \times 2 \times 2$ ANOVA, there was a main effect of cue relatedness for phonemic ($p < .001$), other ($p = .003$), and omission ($p = .022$) errors, but not for semantic ($p = .492$) or both ($p = .583$) errors. There was a main effect of cue position for other errors only ($p = .029$), and there was a main effect of SNR for semantic ($p = .015$) and other ($p < .001$) errors. None of the interaction effects were statistically significant (Fig. 5).

Experiment 2 In the 5×2 ANOVA, a main effect of task condition was found for phonemic and other errors (both $ps < .001$), with the effect approaching significance for omission errors ($p = .097$). The main effect of SNR was significant only for other errors ($p < .001$) with a greater proportion of errors occurring in the SNR -5 condition. The Task Condition \times SNR interaction was not significant. Pair-wise *t* tests showed that for the phonemic errors, a significantly greater proportion occurred in the no-cue condition than either of the related conditions (vs. before-related and after related, $p < .001$) but not in either unrelated condition (BU: $p = .226$, AU: $p = .691$). For the other errors, there was a significant difference between no-cue and the before-related condition ($p = .024$), the difference with after related approaching significance ($p = 0.065$), and no significant difference between the no-cue and the unrelated conditions (both $ps = 1$).

In the $2 \times 2 \times 2$ ANOVA, there was a main effect of cue relatedness for phonemic ($p < .002$) and other ($p < .001$) errors, while the effect for omission errors approached significance ($p = .059$), but not for semantic or both errors (both $ps =$

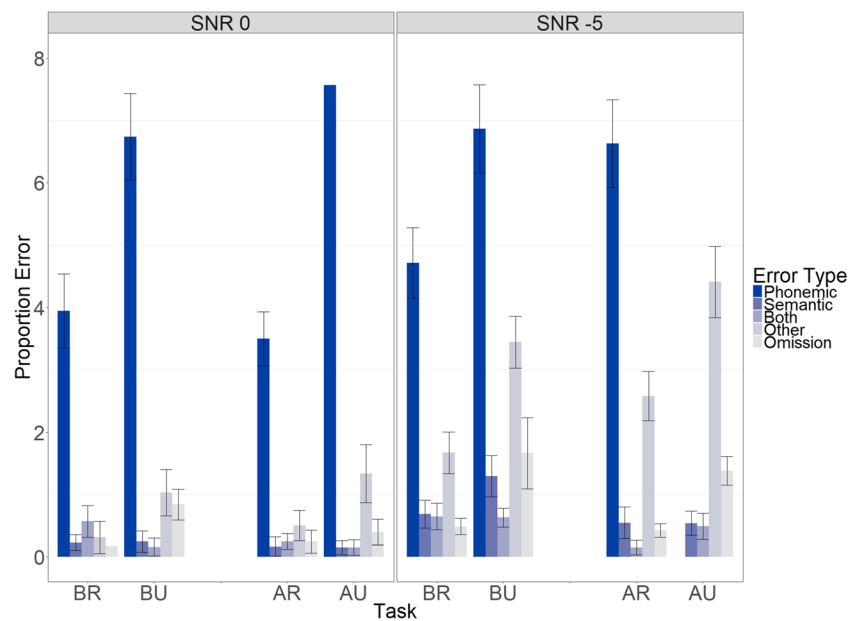


Fig. 5 Proportion of errors (in %) made of each error type in the cued conditions of Experiment 1. “No-cue” is not listed since it could not have semantic/both errors. Error bars indicate the SEM. (Color figure online)

1). There was a main effect of SNR for both ($p = .007$) and other ($p < .001$) errors. The main effect of cue position was not significant, nor were any of the interaction effects (Fig. 6).

Experiment 3 The 4×2 ANOVA revealed a main effect of cue relatedness on phonemic ($p < .001$), other ($p < .001$), and omission errors ($p = .029$), but not semantic ($p = .111$) or both errors ($p = .531$). The interaction of cue relatedness and cue timing was also not significant for any of the error types (Fig. 7).

General discussion

In three experiments, we showed that the presentation of a visual or auditory semantic cue both before and after the presentation of a degraded auditory word stimulus increases accuracy of word identification, and that performance on the task with either modality of cue is comparable. This extends on the findings of Bernstein et al. (1989) as well as Golestani and colleagues (Golestani et al., 2013; Golestani et al., 2009) by showing that the cue and target do not have to belong to the

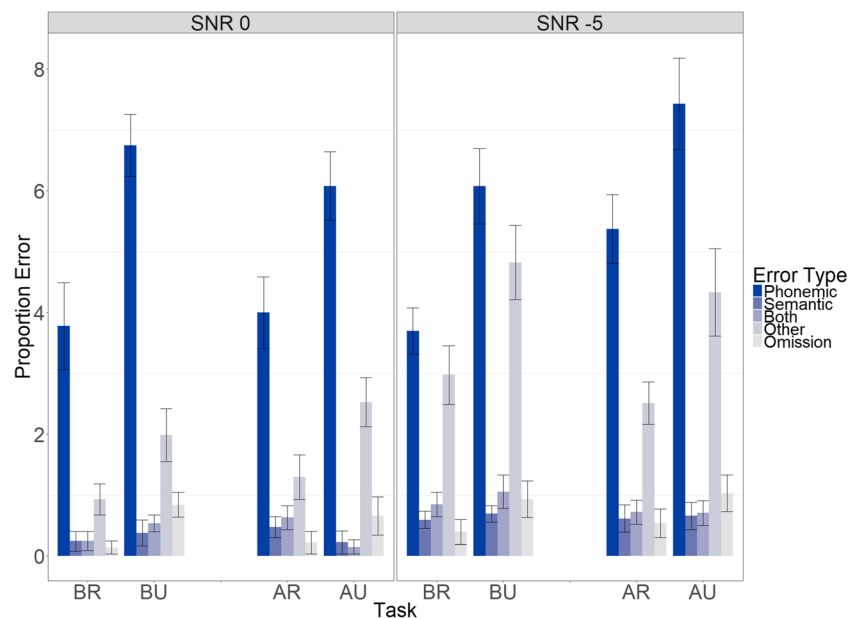


Fig. 6 Proportion of errors (in %) made of each error type in the cued conditions of Experiment 2. “No-cue” is not listed since it could not have semantic/both errors. Error bars indicate the SEM. (Color figure online)

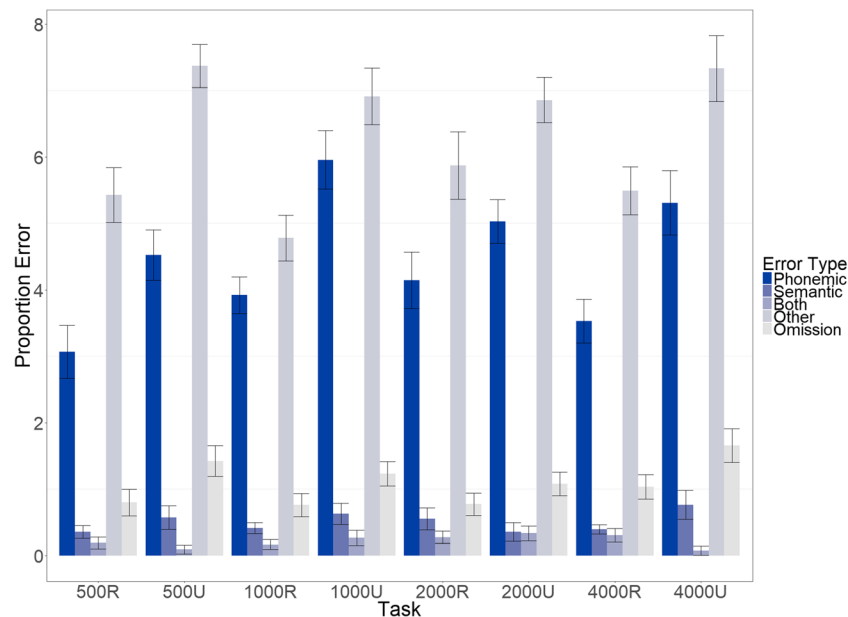


Fig. 7 Proportion of errors (in %) made of each error type in each condition of Experiment 3. Error bars indicate the *SEM*. (Color figure online)

same sensory modality for the cue to retain its efficacy, and that performance with a cue presented *after* the target can still be comparable to when a cue is presented *before* the target, although perhaps not to the same degree. We further demonstrate that cue relatedness effects are present even up to four seconds poststimulus, with the level of accuracy at four seconds comparable to when the cue is presented 500 ms poststimulus.

The results are consistent with predictions made based on the attention to memory model (Zimmermann et al., 2016), in which a listener deploys attention to elements held in STM. In the present study, the cue word provides semantic information that help disambiguate the word-in-noise. In the retro-cue condition, participants may have attempted to maintain what information they could (acoustic, phonological, and semantic) from the noisy word, and the presentation of a related word afterwards would not only refresh this representation but also aid in the selection of a suitable response from alternatives generated from the initially available phonological information. An unrelated word, on the other hand, may have simply refreshed the representation of the target word by virtue of being a cue, but not aid in selecting a proper response from generated alternatives.

The benefit of the cue word presented before the word-in-noise is expected, given the many studies that have shown the advantage of various forms of context in the processing of words in noise (Guediche, Reilly, Santiago, Laurent, & Blumstein, 2016; Johnsrude et al., 2013; Obleser & Kotz, 2011; Pichora-Fuller et al., 1995; Sohoglu, Peelle, Carlyon, & Davis, 2012; Strauß, Kotz, & Obleser, 2013; Zekveld et al., 2011). These results have largely been attributed to top-down mechanisms altering the neural processing of incoming

auditory information, such as in predictive coding models (Friston, 2010; see discussion of Sohoglu et al., 2012) or integrative models (McClelland & Elman, 1986), although work from Davis et al. (2011) suggests that the flow of information in sentence-level context is not necessarily top down, but rather bottom up. Although there is strong evidence of predictive coding taking place in processing incoming acoustic information, such as the body of research on the mismatch negativity (for a review, see Garrido, Kilner, Stephan, & Friston, 2009), the model is unclear about how degraded or unclear past information can be fed into the present for reinterpretation after a retro-cue. In the present study, a degraded target word is presumably transformed into its phonological and/or semantic components to be held in STM. Subsequent information, provided by the retro-cue, helps to disambiguate it. Within the predictive coding framework, the degraded representation in itself is grounds for a prediction to form, albeit a weak prediction. Upon receipt of the retro-cue, if the prediction is incorrect, there will be a prediction error that signals a need to revise the existing model, which now requires the retrieval of the information from STM in order to be matched to a new prediction that is formed based on the retro-cue. The attention to memory model enters here, as it accounts for the maintenance and refocusing of attention onto representations held in STM as opposed to incoming sensory information. Another model of speech perception, the TRACE model (McClelland & Elman, 1986; McClelland, Mirman, & Holt, 2006), posits that speech identification is facilitated by the preceding and following contexts. However, the model does not incorporate postlexical information and does not easily account for the results of our study. Similar to the predictive coding framework, integrative accounts like TRACE allow

for representations in memory to be retrieved and reinterpreted; we propose that the attention to memory model as applied in speech can be the mechanism by which subsequent context is applied to speech representations.

An important distinction between the current task and tasks of auditory attention orienting is that unlike previous tasks, the current task did not involve the recognition of a previously presented stimulus but the identification of the stimulus presented. The results suggest that there is a facilitative effect of presenting a semantically related cue word on the accuracy of identifying a degraded target word held in ASTM. Although there is only a single presented item to be maintained, its ambiguity gives rise to multiple response alternatives that can be derived; therefore, it could be said that the presentation of the semantically related word can act as a retro-cue in selecting one of the possible responses, in a similar manner to that of the retro-cue in previous studies where an array of items had to be maintained and then selected from (Backer & Alain, 2012; Griffin & Nobre, 2003). In the same manner that semantic priming may have an attentional component in which a prime generates expectancies for a subsequent target word, we propose that a semantically related word following a degraded target can inform the processing of the target by guiding the selection of a contextually relevant word that matches the perceptual information initially captured by the listener. Our task has captured aspects of both the prospective and retrospective relations between words, and has shown a comparable benefit between both.

One possible alternative explanation for the increase in accuracy with respect to cue relatedness is that in actively using all cues provided, participants were confused with the presentation of unrelated cues, even though they otherwise heard the word properly. We attempted to mitigate this possibility as much as possible by informing the participants that cue words could be either related or not related to the target, but to always state what they believed the target to be regardless of cue relatedness. The results of the error analysis suggest that participants were doing this consistently, since the majority of errors made still contained phonemes of the target word, and the effect of relatedness on phonemic, other, and omission errors (but not on semantic or both) suggests that the increase in errors for unrelated conditions is driven by an increased number of such responses. If participants were using the cue words without consideration for their relation to what target signals they heard, we would have expected to a greater proportion of semantic or both errors made, but the overall proportions of such errors were small. This relationship was maintained in the error analysis of Experiment 3, in which the profile of proportion of errors made did not significantly vary over time of delay, but only in relation to cue relatedness.

Our results, when comparing the related cue conditions with the no-cue condition, suggest that the benefit of relatedness is not dependent on impairment due to an unrelated cue,

but the unrelated words could still be an additional source of error. A feature for future consideration would be adding a measure of confidence after each response, so that we can assess whether the increase in accuracy is reflective of an increase in confidence of response, as well as the degree of bias that participants may experience due to the presentation of what they perceive to be a related cue (Rogers, Jacoby, & Sommers, 2012). In addition, it can aid in separating responses in which the cue word served a confirmatory role (i.e., the participant clearly heard the word and did not need additional aid in deciphering it) from responses where the cue word was informative (i.e., the participant did not hear the word in its entirety and needed a contextual aid to generate a response or choose from one of several alternatives).

Regarding the results of the working memory correlation analyses, the lack of correlation between performance on the experimental task reported and working memory assessments may be surprising, given the ELU model of speech comprehension that implicate working memory capacity as an index of amount of comprehension (Rönnberg et al., 2013). However, it has been argued that some simple span tests, such as the digit span test used here, are inadequate for capturing differences in memory ability that are important for speech comprehension (Daneman & Merikle, 1996), although it may be due to the way such tests are scored (Unsworth & Engle, 2007). We attempted to compensate here by combining our various working memory assessments, but this method still showed no significant correlation. It may be that the experimental task did not tax working memory much for a group of healthy young adults, such that small differences in working memory could not explain differences in performance. Increasing the number of stimuli to recall could lead to a correlation with measures of working memory, while still cueing either all or a subset with a semantically related word to observe the effect of a cue.

As for the speech-in-noise tests, usage of context is different between the QuickSIN and WIN tests. The WIN tests are constructed strictly to repeat back a single target word with no additional context provided in the preceding sentence; any correlation would be expected to occur either with the no-cue or unrelated-cue conditions. The QuickSIN asks participants to repeat back entire sentences; although they were designed such that predictions of sentences could not be generated from one word to the next, sentences were not completely nonsensical and so some words can provide context for subsequent words in a similar way as the cue words in the experimental task. This could explain the correlation found between QuickSIN performance and the before-related condition.

Active maintenance of the target-in-noise is proposed as the mechanism by which participants completed the task, but it is possible that the associations were implicitly triggered without the cue being a guide for selection. Higgins and Johnson (2013) found that the presentation of masked words

that were semantically related interfered with the refreshing of a target word. However, their task was such that the masked words were imperceptible, and their findings were in support of their theory that this interference was observed when conscious control was not exercised; their initial study (Higgins & Johnson, 2009) did not demonstrate an impaired ability to refresh an immediately presented target. On the other hand, performance was improved in our task, which suggests that the exercise of cognitive control was employed and therefore implicit interference was not induced.

It is asserted, therefore, that there were attentional processes engaged in each of the three experiments, due to the timing conditions used and the difficulty of word identification. Without the ability to maintain the possible alternative responses based on what was perceived, and then select from one of them based on a subsequent semantic context, participants would not demonstrate a benefit from the semantically related retro-cue as compared with the unrelated cue or no-cue at all, especially after a delay of 4 seconds between target and cue. To test this assertion in the future, it will be necessary to either test this paradigm against measures of attentional control, or disrupt attention during the task by asking participants to perform an intervening task. The usage of neuroimaging methods, such as EEG, will also allow for comparisons between performance on this task and on previous tasks of auditory reflective attention, which have identified indices of maintenance and selection of information from ASTM. These methods, combined with the abovementioned measure of confidence, will allow us to separate various neural processes involved in different stages of speech-in-noise processing, as well as the temporal dynamics of maintaining verbal memory.

In conclusion, we have shown that a related cue word relative to an unrelated cue can boost the accuracy of identification of a word energetically masked by white noise, and the related word is effective whether it is placed either before or after the word-in-noise. The cuing effect persists even when the cue is presented up to 4 seconds after the target. These findings support a model of speech-in-noise perception where context aids word identification either prospectively by generating expectancies, or retrospectively by aiding selection from among self-generated alternatives. Future work should endeavor to distinguish between situations when a cue word merely serves a confirmatory role in identification of a word-in-noise, and when a cue word influences the comprehension of a word not able to be identified from a set of possibilities. The use of neuroimaging techniques with this design will contribute to understanding how we attend reflectively to speech representations, and how this differs between when we have confident and vague representations of what was just said.

Author note T. M. Vanessa Chan, Department of Psychology, University of Toronto, Toronto, Ontario, Canada, and Rotman Research Institute, Baycrest, Toronto, Ontario, Canada; Claude Alain, Rotman Research Institute, Baycrest, Toronto, Ontario, Canada, Department of Psychology, Institute of Medical Sciences and Faculty of Music, University of Toronto, Toronto, Ontario, Canada.

This research was supported by a National Sciences and Engineering Research Council (NSERC) grant awarded to C.A.

References

- Astle, D. E., Summerfield, J., Griffin, I., & Nobre, A. C. (2012). Orienting attention to locations in mental representations. *Attention, Perception, & Psychophysics*, *74*(1), 146–162.
- Backer, K. C., & Alain, C. (2012). Orienting attention to sound object representations attenuates change deafness. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1554–1566.
- Backer, K. C., Binns, M., & Alain, C. (2015). Neural dynamics underlying attentional orienting to auditory representations in short-term memory. *Journal of Neuroscience*, *35*(3), 1307–1318.
- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459.
- Bernstein, I. H., Bissonnette, V., Vyas, A., & Barclay, P. (1989). Semantic priming: Subliminal perception or context? *Perception & Psychophysics*, *45*(2), 153–161.
- Cabeza, R., Mazuz, Y. S., Stokes, J., Kragel, J. E., Woldorff, M. G., Ciaramelli, E., . . . Moscovitch, M. (2011). Overlapping parietal activity in memory and perception: Evidence for the attention to memory model. *Journal of Cognitive Neuroscience*, *23*(11), 3209–3217.
- Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101.
- Ciaramelli, E., Grady, C. L., & Moscovitch, M. (2008). Top-down and bottom-up attention to memory: A hypothesis (AtoM) on the role of the posterior parietal cortex in memory retrieval. *Neuropsychologia*, *46*(7), 1828–1851.
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433.
- Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through “top-down” processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, *23*(12), 3914–3932.
- Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences*, *111*(19), 7126–7131.
- Falissard, B. (2012). psy: Various procedures used in psychometry (R Package Version 1.1) [Computer software]. Retrieved from <https://CRAN.R-project.org/package=psy>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127.
- Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*(3), 453–463.
- Gilbert, R. A., Davis, M. H., Gaskell, M. G., & Rodd, J. M. (2018). Listeners and readers generalize their experience with word

- meanings across modalities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. <https://doi.org/10.1037/xlm0000532>
- Golestani, N., Hervais-Adelman, A., Obleser, J., & Scott, S. K. (2013). Semantic versus perceptual interactions in neural processing of speech-in-noise. *NeuroImage*, *79*, 52–61.
- Golestani, N., Rosen, S., & Scott, S. K. (2009). Native-language benefit for understanding speech-in-noise: The contribution of semantics. *Bilingualism: Language and Cognition*, *12*(3), 385–392.
- Gregg, M. K., & Samuel, A. G. (2009). The importance of semantics in auditory representations. *Attention, Perception, & Psychophysics*, *71*(3), 607–619.
- Griffin, I. C., & Nobre, A. C. (2003). Orienting attention to locations in internal representations. *Journal of Cognitive Neuroscience*, *15*(8), 1176–1194.
- Guediche, S., Reilly, M., Santiago, C., Laurent, P., & Blumstein, S. E. (2016). An fMRI study investigating effects of conceptually related sentences on the perception of degraded speech. *Cortex*, *79*, 57–74.
- Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: Effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(2), 460.
- Higgins, J. A., & Johnson, M. K. (2009). The consequence of refreshing for access to nonselected items in young and older adults. *Memory & Cognition*, *37*(2), 164–174.
- Higgins, J. A., & Johnson, M. K. (2013). Lost thoughts: Implicit semantic interference impairs reflective access to currently active information. *Journal of Experimental Psychology: General*, *142*(1), 6.
- Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and language*, *35*(1), 66–85.
- Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*(4), 281–312.
- Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 645.
- Johnson, M. K., Raye, C. L., Mitchell, K. J., Greene, E. J., Cunningham, W. A., & Sanislow, C. A. (2005). Using fMRI to investigate a component process of reflection: Prefrontal correlates of refreshing a just-activated representation. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(3), 339–361.
- Johnsrude, I. S., Mackey, A., Hakymez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science*, *24*(10), 1995–2004.
- Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, *61*(5), 1337–1351.
- Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *116*(4), 2395–2405.
- Lawrence, M. A. (2016). ez: Easy analysis and visualization of factorial experiments (R Package Version 4.4-0) [Computer software]. Retrieved from <https://CRAN.R-project.org/package=ez>
- Lim, S.-J., Wöstmann, M., & Obleser, J. (2015). Selective attention to auditory memory neurally enhances perceptual precision. *Journal of Neuroscience*, *35*(49), 16094–16104.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. doi: 10.1016/0010-0285(86)90015-0
- McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, *10*(8), 363–369.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226.
- Nelson, D., McEvoy, C., & Schreiber, T. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*, 402–407.
- Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage*, *55*(2), 713–723.
- Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*(1), 593–608.
- R Core Team. (2017). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.
- 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*(1), 33.
- Rossell, S. L., Bullmore, E. T., Williams, S. C., & David, A. S. (2001). Brain activation during automatic and controlled processing of semantic relations: A priming experiment using lexical-decision. *Neuropsychologia*, *39*(11), 1167–1176.
- Rönnerberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., . . . Rudner, M. (2013). The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*(31). <https://doi.org/10.3389/fnsys.2013.00031>
- Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *The Journal of the Acoustical Society of America*, *123*(1), 489–499.
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, *32*(25), 8443–8453.
- Strauß, A., Kotz, S. A., & Obleser, J. (2013). Narrowed expectancies under degraded speech: Revisiting the N400. *Journal of Cognitive Neuroscience*, *25*(8), 1383–1395.
- Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038.
- Wechsler, D. (1997). WMS-III: Wechsler memory scale administration and scoring manual. San Antonio, TX: Psychological Corporation.
- Wilson, R. H. (2003). Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. *Journal of the American Academy of Audiology*, *14*(9), 453–470.
- Zekveld, A. A., Rudner, M., Johnsrude, I. S., Festen, J. M., Van Beek, J. H., & Rönnerberg, J. (2011). The influence of semantically related and unrelated text cues on the intelligibility of sentences in noise. *Ear and hearing*, *32*(6), e16–e25.
- Zimmermann, J. F., Moscovitch, M., & Alain, C. (2016). Attending to auditory memory. *Brain Research*, *1640*, 208–221.
- Zimmermann, J. F., Moscovitch, M., & Alain, C. (2017). Long-term memory biases auditory spatial attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(10), 1602.