Glimpsing speech interrupted by speech-modulated noise

Rachel E. Miller, Bobby E. Gibbs II, and Daniel Fogerty^{a)}

Department of Communication Sciences and Disorders, University of South Carolina, 1229 Marion Street, Columbia, South Carolina 29201, USA

(Received 2 October 2017; revised 24 April 2018; accepted 2 May 2018; published online 23 May 2018)

Everyday environments frequently present speech in modulated noise backgrounds, such as from a competing talker. Under such conditions, temporal glimpses of speech may be preserved at favorable signal-to-noise ratios during the amplitude dips of the masker. Speech recognition is determined, in part, by these speech glimpses. However, properties of the noise when it dominates the speech may also be important. This study interrupted speech to provide either high-intensity or low-intensity speech glimpses derived from measurements of speech-on-speech masking. These interrupted intervals were deleted and subsequently filled by steady-state noise or one of four different types of noise amplitude modulated by the same or different sentence. Noise was presented at two different levels. Interruption by silence was also examined. Speech recognition was best with high-intensity glimpses and improved when the noise was modulated by missing high-intensity segments. Additional noise conditions detailed significant interactions between the noise level and glimpsed speech level. Overall, high-intensity speech segments, and the amplitude modulation (AM) of the segments, are crucial for speech recognition. Speech recognition is further influenced by the properties of the competing noise (i.e., level and AM) which interact with the glimpsed speech level. Acoustic properties of both speech-dominated and noise-dominated intervals of speech-noise mixtures determine speech recognition. © 2018 Acoustical Society of America.

<https://doi.org/10.1121/1.5038273>

[DB]

Pages: 3058–3067

I. INTRODUCTION

Everyday listening conditions frequently involve speech-on-speech masking. Studies investigating the acoustic contributions of speech-on-speech masking, such as the current study, frequently use speech-modulated noise (SMN) to model the masker speech while avoiding linguistic interference (e.g., Brungart, 2001; Brungart *et al.*, 2001; Simpson and Cooke, 2005; Bernstein and Grant, 2009; Humes *et al.*, 2017). When listening to speech in the presence of modulated maskers, the amplitude modulation (AM) of the speech and masker provide the listener with intermittent “dips” where target speech can be perceived at a more favorable signal-to-noise ratio (SNR). Current methods used to model speech understanding in the presence of modulated maskers are based on predicting intelligibility on the basis of these “glimpsed” segments (Cooke, 2003, 2006; Li and Loizou, 2007). Indeed, modifications of speech intelligibility models that capture fluctuations in the SNR over time result in superior performance to models based only on the long-term average speech spectrum (e.g., Rhebergen and Versfeld, 2005). Therefore, the perception of speech in modulated noise appears to involve the integration of glimpsed speech cues that are fragmented across time.

For over half a century, interrupted speech has been used to investigate how listeners perceive speech based on these temporal fragments (Miller and Licklider, 1950). This method involves deleting segments of the target speech and replacing them with silence, noise, or some other type of filler. However, while glimpses of certain speech information

appear essential for determining intelligibility, it is also likely that the acoustic characteristics of the competing signal during the non-glimpsed intervals also plays a role. For example, increasing the overall level of the interrupting noise can facilitate recognition to a certain degree (Powers and Wilcox, 1977; Bashford *et al.*, 1992). Similarly, restoring speech AM cues by modulating the intervening noise intervals with the temporal envelope of the missing speech further aids in recognition (Bashford *et al.*, 1996; Shinn-Cunningham and Wang, 2008). This evidence suggests that in addition to the contribution of speech glimpses, the level and AM properties of the intervening noise also contribute to the recognition of interrupted speech.

The benefit of restoring speech AM cues by modulating interfering noise is perhaps not surprising given that AM properties of speech provide significant information for speech recognition when provided with or without spectral cues (e.g., Shannon *et al.*, 1995; Friesen *et al.*, 2001; Zeng *et al.*, 2005; Fogerty, 2011; Apoux and Healy, 2013). The importance of AM speech cues is further demonstrated when they are distorted, such as through hearing-aid compression. Such AM-distorted conditions limit the perceptual restoration of interrupted speech (Başkent *et al.*, 2009). However, temporal AM speech cues may be difficult to utilize in the presence of temporally fluctuating background noise (Nelson *et al.*, 2003; Nelson and Jin, 2004). This may be due to interruption of the speech AM (Gilbert *et al.*, 2007) and/or to interference between the AM properties of the speech and noise (Stone *et al.*, 2012; Fogerty *et al.*, 2016b; Humes *et al.*, 2017). Thus, SMN may provide significant interference for speech recognition based on its AM properties.

^{a)}Electronic mail: fogerty@sc.edu

In contrast, the AM properties of a background noise can also benefit perception. The recognition of interrupted speech improves when AM cues of the missing speech are restored by amplitude modulating the replacement noise (Bashford *et al.*, 1996; Shinn-Cunningham and Wang, 2008). It appears that this improvement is contingent on what portions of the speech are being restored (e.g., vowels versus consonants, Fogerty, 2014). Furthermore, amplitude modulating the replacement noise by a competing talker may also be beneficial depending on an interaction between the modulation spectra of the preserved speech intervals and of the competing talker (Fogerty *et al.*, 2016a).

This evidence demonstrates that speech recognition is affected by an interaction between the temporal AM of the speech and noise. However, the nature of this interaction is not completely understood. In particular, while arguably more ecological, the effect of noise AM from sources *other* than the missing speech has not been investigated thoroughly within an interrupted speech paradigm. Speech-in-noise perception, particularly in multi-talker environments, is likely dependent on how the AM of competing sources interfere with the recognition of glimpsed speech. Furthermore, while periodic interruption methods are currently used to study a wide variety of effects on speech recognition (e.g., Nagaraj and Knapp, 2015; Wilson and Irish, 2015; Bhargava *et al.*, 2016; Molis and Gallun, 2016; Shafiro *et al.*, 2016; Smiljanic *et al.*, 2016; Smith and Fogerty, 2017); this methodology needs to be verified with more naturalistic conditions where non-periodic speech interruptions are defined according to the temporal intervals of speech glimpses in realistic masking conditions, such as during speech-on-speech masking. This is a more ecologically valid case and necessary to help bridge results from the interrupted speech literature to more naturalistic speech glimpsing models. The current study seeks to extend investigations of interrupted speech to more closely model speech glimpsing in modulated noise. Naturally-defined speech glimpses were alternated with different types of noise to examine how the AM properties of fluctuating maskers affect glimpsed speech recognition.

A. Interrupted speech

The intelligibility of periodically interrupted speech is dependent on the proportion of speech preserved (i.e., duty cycle) and the rate of interruption (e.g., Miller and Licklider, 1950; Powers and Speaks, 1973; Wang and Humes, 2010). Furthermore, a perceptual benefit (i.e., restoration) is commonly observed when the speech interruption intervals are filled with noise instead of silence (Miller and Licklider, 1950; Powers and Wilcox, 1977; Bashford and Warren, 1987; Bashford *et al.*, 1992; Benard *et al.*, 2014). This perceptual restoration is most robust when the noise level is sufficient enough to mask the speech prior to its deletion (Warren, 1970; Bashford *et al.*, 1992), although benefit is reduced for extreme levels of noise that exceed -24 dB SNR (Powers and Wilcox, 1977).

The intelligibility of interrupted speech has also been investigated using non-periodic interruptions, such as defining speech interruptions based on relative root-mean-square

(rms) intensity (e.g., Chen and Loizou, 2012; Gaudrain and Carlyon, 2013; Guan *et al.*, 2016), informational metrics (i.e., cochlea-scaled entropy; Stilp and Kluender, 2010; Chen and Loizou, 2012; Stilp, 2014; Oxenham *et al.*, 2017), and consonant-vowel boundaries (e.g., Cole *et al.*, 1996; Kewley-Port *et al.*, 2007; Fogerty and Kewley-Port, 2009; Chen *et al.*, 2015; Fogerty *et al.*, 2016a). These different interruption methods have been useful in defining how different intervals of speech contribute to intelligibility. A basic summary of this work demonstrates that not all intervals of speech (e.g., speech segments) contribute equally to intelligibility (e.g., Kewley-Port *et al.*, 2007; Stilp and Kluender, 2010; Chen and Loizou, 2012). Importantly, it appears that the recognition of interrupted speech is maximized when speech interruptions minimize distortion of the overall AM cues of the sentence (i.e., temporal envelope) that is best preserved by high-intensity segments, like vowels (Fogerty, 2013). However, speech interruptions defined using the above methods are not necessarily representative of how speech is preserved in the presence of a concurrent AM masker, such as a competing talker. It is particularly important to extend results from speech interruption studies to more ecologically valid interruptions that take into account how speech glimpses are distributed in time, and potential informational content, due to a competing talker. This is because the temporal characteristics of speech interruption (e.g., duration and rate of interruption) also significantly affect intelligibility (e.g., Miller and Licklider, 1950).

B. Noise modulation

The AM of a competing noise defines speech glimpses (i.e., by way of “dip listening”). However, noise modulation may have other effects. For example, in simultaneous masking conditions, AM properties of the noise may interfere with speech recognition (e.g., Stone *et al.*, 2012; Fogerty *et al.*, 2016b; Humes *et al.*, 2017). Significant interference between the modulation properties of speech and noise may also occur in non-simultaneous masking conditions that are present for interrupted speech (Fogerty *et al.*, 2016a). However, in certain conditions, noise AM properties may be beneficial. As already reviewed, listeners obtain higher speech recognition scores by modulating the interrupting noise by the AM of the replaced speech segment (Bashford *et al.*, 1996; Shinn-Cunningham and Wang, 2008). Furthermore, the benefit of this restored speech modulation observed in both studies involved a two-fold improvement in intelligibility compared to interruption by silence. The preserved speech AM may facilitate the perception of speech continuity due to a failure in segregating speech and noise sources (Shinn-Cunningham and Wang, 2008), or possibly provide restored suprasegmental cues (Rosen, 1992) across the target sentence. Interestingly, improvements in speech recognition are only observed when the noise restores the AM of missing high-intensity vowel segments, but not from missing low-intensity consonants (Fogerty, 2014). This demonstrates a potential interaction between the preserved speech glimpses and the particular characteristics of the restored AM, which could either interfere or benefit speech

recognition. However, it is necessary to investigate this effect under interruption conditions that model more natural glimpsing conditions.

C. The current study

As reviewed here, the method of speech interruption, the preservation of speech AM, and the acoustic properties of the replacement noise all appear to affect the intelligibility of interrupted speech. The current study represents an effort to test how these previous results extend to interrupted speech that presents glimpses defined by natural speech-on-speech listening conditions. This interruption method was then used to examine three factors that might determine speech recognition: (1) the level of the preserved speech interruptions, (2) the level of the interrupted noise that alternated with the speech, and (3) the source of modulation for the interrupted AM noise.

Speech interruption intervals were defined using a single-talker SMN presented at 0 dB SNR. Comparison between the original speech and SMN resulted in a running SNR across the sentence that was used to define two levels of speech interruption according to the local SNR (positive or negative). Positive local SNR segments (+SNR) exceeded the level of the competing noise and are consistent with the definition of “speech glimpses.” We also investigated speech interrupted by preserving the opposite, negative local SNR segments (−SNR). This latter speech interruption method afforded the ability to examine if properties of the interrupting noise interact with properties of the preserved speech. Functionally, this resulted in interruptions that preserved either high-intensity speech glimpses (+SNR) or intervals of low-intensity speech (−SNR).

The intervening replacement noise was manipulated in two ways to investigate how the noise level and noise AM affect the perception of interrupted speech. Noise level was controlled by scaling noise interruption intervals according to the level of either the replaced segment or to the overall level of the target sentence. This manipulation allowed us to examine the effect of the noise AM at the level of the signal from which it was derived, as well as at a standard level across noise AM type, respectively.

This intervening noise was amplitude-modulated by one of five different temporal envelopes: unmodulated (condition 1), modulated by the target sentence (conditions 2–3), or modulated by a different sentence (conditions 4–5). (1) The first noise condition used “steady-state” (unmodulated) noise. (2) SMN based on the missing portion of the target speech was used to replicate studies that have demonstrated a benefit of restored speech AM (e.g., [Bashford et al., 1996](#); [Shinn-Cunningham and Wang, 2008](#)) using speech interruptions defined by speech-on-speech masking. (3) SMN based on the preceding (glimpsed) target speech AM examined if this modulation would result in better intelligibility due to perceived AM continuity (following an observation by [Fogerty et al., 2016a](#)). (4) SMN based on a random segment of speech from a different sentence examined the effect of a competing AM signal. (5) Finally, a random segment of SMN as in 4 was time compressed to determine if the altered modulation rate would enhance the segregation of speech and noise. These five noise conditions, one with steady-state noise and four with SMN, were compared to interruption by silence.

Consistent with rms-interrupted and vowel-interrupted speech, as well as with cochlea-scaled entropy interruption (see [Oxenham et al., 2017](#)), we expected that performance would be best when the high-level +SNR glimpses were preserved. Furthermore, consistent with the literature on perceptual restoration (e.g., [Bashford et al., 1992](#)), we expected that a higher replacement noise level would result in better performance. We also expected an effect of noise AM. We hypothesized that modulating the interruption noise by AM of the target talker would significantly improve intelligibility. AM that preserves the missing speech segment should improve intelligibility due to better perceptual integration across speech glimpses (or failed segregation of the noise intervals, [Shinn-Cunningham and Wang, 2008](#)). Similarly, we hypothesized that AM from the preceding segment could result in better intelligibility by preserving the continuity in the modulation spectrum of the glimpsed speech (see [Fogerty et al., 2016a](#)), and potentially because the cues are derived from the same sentence. Conversely, random AM cues derived from different speech materials were expected to interfere with speech recognition due to conflicting modulation cues. Finally, time compression of this random AM signal was expected to reduce interference due to better modulation-based perceptual segregation from the target speech.

II. METHODS

A. Listeners

Young normal-hearing adults participated in this study ($N = 11$, $M = 23$ years, range = 20–28, 1 male, 10 female). All participants had thresholds at 20-dB hearing level (HL) or below at audiometric frequencies 250–8000 Hz for both left and right ears. They were all native English speakers. Written consent was obtained from all participants. Participants were compensated for their time. An additional participant was not included in the analysis due to having heard the sentences in a prior experiment.

B. Stimuli and design

Sentences were from the IEEE corpus ([IEEE, 1969](#)), which is divided into 72 phonetically-balanced lists of ten sentences (lists 23–49 were used here). Each sentence contained five keywords. Sentences were recorded by a male talker at a sampling rate of 25 000 Hz ([Loizou, 2007](#)). Within each IEEE list, the sentences are syntactically and grammatically correct but have relatively low contextual predictability.

Each sentence was first interrupted based on identified speech glimpses. Glimpses were defined by calculating the running SNR between the target speech and a comparison SMN at an average SNR of 0 dB. Following interruption, speech intervals between the glimpses (for +SNR interruption) or the glimpsed intervals themselves (for −SNR interruption) were deleted and filled with steady-state noise or with one of four different types of SMN at one of two different noise levels. The experimental conditions entailed a 2 (speech interruption method) $\times$ 2 (noise level) $\times$ 5 (noise type) design with two additional silent replacement control

conditions, one for each interruption method (i.e., +SNR and -SNR).

1. Speech interruption

a. SMN used for interruption. The first step for defining the speech interruptions involved creating a competing SMN. A concatenation of 100, randomly selected IEEE sentences was created with silence removed between sentences. The temporal envelope of this concatenation was obtained by low-pass filtering the half-wave rectified signal using a sixth-order Butterworth filter with a cutoff frequency of 50 Hz. This temporal envelope was used to amplitude-modulate a steady-state noise matching the long-term average speech spectrum for the concatenated file, creating a noise that fluctuated based on the AM of spoken sentences.

b. Interruption method. To interrupt the speech signal, a random segment of the SMN matching the sentence duration was added to a second channel of the speech file [Figs. 1(a)–1(b)]. The overall rms of the SMN was equated in level to the overall rms of the speech file. For each sentence, the SMN used for interruption was saved for later analysis. Next, the running rms was calculated separately for the uninterrupted speech and SMN channels using consecutive 16-ms windows. This window size defined the minimum “on” duration for the interrupted speech. This constrained the running SNR to consider temporal fluctuations under 63 Hz that are more characteristic of the AM speech signal and remove the effects of periodicity on the glimpse analysis (see Rosen, 1992). This minimum analysis window size is consistent with other analyses of glimpsing (e.g., Li and Loizou, 2007; Fogerty *et al.*, 2016b) and spectral change (Stilp and Kluender, 2010). Indeed, 16-ms glimpses

appear to be the shortest glimpses used for the perception of phonetic features (Füllgrabe *et al.*, 2006).

The local SNR for each 16 ms window was calculated [Fig. 1(c)]. This running SNR was used to define the time intervals where the local SNR was greater or equal to 0 dB for at least 16 ms, i.e., defined here as the minimum speech glimpse duration. The target speech channel was then interrupted based on these defined glimpse intervals using one of two different interruption methods. The +SNR method preserved speech fragments that occurred above 0 dB SNR (i.e., the glimpses) and replaced the noise-dominated speech intervals with silence or noise [Fig. 1(d) with silent replacement]. Conversely, the -SNR method preserved the sentence file when the running SNR dropped below 0 dB SNR and replaced (with noise or silence) the glimpsed time intervals [Fig. 1(e) with silent replacement]. To minimize the introduction of transients, segmentation occurred at zero-crossings in the speech channel within 2 ms of the identified time marking.

Analysis of the total duration of preserved speech relative to the total sentence duration was near 50% for both interruption methods (i.e., +SNR and -SNR conditions). This matched duration of preserved speech was expected as interruptions were defined by a SMN at 0 dB SNR. In addition, the median duration of preserved speech interruptions was 111 ms (inter-quartile range: 36 ms). Thus, this interruption method resulted in preserved speech intervals of approximately phoneme- and syllable-sized durations (compare to syllable durations in natural speech of 50–400 ms, Greenberg *et al.*, 2003).

2. Noise replacement

The interruption method defined the time intervals when the speech was “on.” During the “off” intervals, speech was

FIG. 1. Waveforms for an example sentence are displayed to demonstrate the speech interruption method and modulated noise types. (a) uninterrupted speech channel, (b) speech-modulated noise channel used to define interruptions, (c) running local SNR based on comparison of (a) and (b), (d) +SNR speech interrupted to preserve intervals (i.e., glimpses) above 0 dB SNR, (e) -SNR speech interrupted to preserve intervals below 0 dB SNR, and (f) different replacement SMNs. Replacement SMNs are paired according to the modulation source: unmodulated, modulated by the target sentence (Target, Preceding), or modulated by a different sentence (Random, Time-Compressed). The vertical boundary lines in (a)–(e) indicate the identification of a noise replacement interval (d) or preserved speech interval (e). The displayed modulated noises in (f) were scaled to the average sentence level.

replaced with either silence [as displayed in Figs. 1(d) and 1(e)] or with one of five different noises [see Fig. 1(f): unmodulated, modulated by the target sentence (target or preceding), or modulated by a difference sentence (random or time-compressed)]. In other words, the final stimulus presented interrupted speech which alternated with either silence or a type of noise.

As perceptual restoration of interrupted speech is influenced by the level of the replacement noise (e.g. Bashford *et al.*, 1992), the level of each noise type was scaled to two different intensity levels. This manipulation examined interactions between the effects of masker AM type and intensity level on speech recognition. For the first scaling method, the *segment-scaled* replacement noise level was defined according to the level of the segment used for modulation. For SMN, this was the speech segment from either the same (i.e., target) or different sentence that was used to modulate the noise masker. For the second scaling method, the *sentence-scaled* replacement noise level was defined according to the average level of the target sentence prior to interruption.

For all conditions, a 1-ms raised cosine ramp was applied to the beginning and end of each replacement noise segment. Processing for the five noise replacement conditions (one unmodulated and four types of SMN) are detailed below.

a. Unmodulated noise. For each sentence, a steady-state noise was created that matched the speech spectrum of the target sentence. A random segment of this unmodulated noise was used to fill in each replacement interval. To examine the effect of noise level, the unmodulated noise was scaled to be either at the average level of the uninterrupted sentence, or 3 dB below this average level. This scaling allowed for a comparison of recognition in unmodulated noise based on a consistently higher or lower intensity. A scale factor of 3 dB approximated the average level of the segment-scaled noise in the other SMN conditions when the replacement noise occurred below 0 dB (re: average uninterrupted speech level).

b. Target SMN. Target SMN was created by modulating a steady-state noise with the temporal envelope of the deleted target speech. Therefore, this noise AM condition, to a certain extent, preserved the speech modulations of the original sentence throughout the noise replacement intervals. In the segment-scaled condition, the level of the deleted speech was also preserved, meaning that the level was necessarily lower or higher depending on the +SNR or -SNR condition, respectively. The temporal envelope of the deleted speech was extracted using half-wave rectification and low-pass filtering using a second order Butterworth filter at 32 Hz.¹

c. Preceding SMN. Preceding noise was created by modulating a steady-state noise based on speech modulations from the interval immediately previous to the replacement interval. Thus, this condition resulted in the noise modulated to repeat the previously heard modulations present during the preserved (i.e., glimpsed) speech interval. The preceding modulation was extracted from the original target speech channel [Fig. 1(a)] and therefore was not influenced by any

preceding interruption that may have occurred in the stimulus. Temporal envelope filtering was identical to the target SMN condition (see 2 above).

d. Random SMN. The randomly selected SMN used to define interruption [Fig. 1(b)] was used as the replacement signal by selecting the interruption noise corresponding to the time interval of the deleted (i.e., noise-replaced) target speech segment. As this was the SMN used to define speech interruptions, it follows that for segment-scaled +SNR conditions, the intensity level of the Random SMN was low, and for -SNR, the relative level was high. However, noise levels were equalized for the sentence-scaled conditions.

e. Time-compressed SMN. To facilitate modulation-based segregation of the replacement noise, this last condition used a time-compressed SMN. The IEEE speech concatenation was time-compressed to 25% of the original speed (i.e., four times as fast) using the Pitch Synchronous Overlap and Add (PSOLA) method in Praat (Boersma and Weenink, 2014). This time-compression factor of 25% was previously associated with a release from modulation masking (Fogerty *et al.*, 2016b) which, potentially, could facilitate perceptual segregation of the SMN from the target speech. The temporal envelope of this time-compressed signal was extracted using identical filtering to the random SMN condition. Finally, a steady-state speech-shaped noise was modulated by the temporal envelope of the time-compressed speech concatenation to create the time-compressed SMN. Interrupted speech stimuli were then alternated with random selections of this SMN.

C. Procedure

Participants were seated in a sound-attenuating booth and presented sentence stimuli monaurally via the right ear (Sennheiser HD 280PRO) routed through an Edirol UA 25-EX soundcard. The level of the stimuli was calibrated at 70 dB sound pressure level (SPL) using a Larson-Davis 800B sound level meter with linear weighting, using a speech-shaped noise matching the rms level of the uninterrupted speech stimuli. Participants were instructed to listen to the sentences and repeat back what they heard as accurately as possible. The task was self-paced and the participant pressed a button using a custom MATLAB interface to play the next stimulus. Participants only heard each sentence one time and sentences were never repeated in a different condition. Responses were audio recorded and keywords were scored by the first author, a native English speaker.

Stimuli were blocked by interruption type and intensity level, with an additional block for the silence conditions. The blocks contained 50 sentences; ten sentences for each of five noise types (one unmodulated and four SMN). The blocks were composed of five preselected IEEE sentence lists. Stimuli containing the silence replacement condition were blocked together regardless of glimpse type, creating a block of 20 sentences. Block order and sentence order within the blocks were randomized for each participant. Each experimental block was immediately preceded by a demo exercise that presented two example sentences for each

experimental condition (i.e., ten sentences for each block). All demo sentences were unheard IEEE sentences spoken by the same talker as the experimental sentences. No feedback was provided to participants.

III. RESULTS

Participant accuracy scores are displayed in Fig. 2 for +SNR interruption (panel A) and -SNR interruption methods (panel B) that preserve the different relative speech levels. In general, participants had much higher accuracy for +SNR interruption. Recall that this interruption method preserved the intense portions of speech, as would be typical of naturally glimpsed speech. Results were analyzed using a 2 (interruption method: +SNR, -SNR) $\times$ 2 (noise level: sentence-scaled, segment-scaled) $\times$ 5 (noise type: unmodulated noise,

target SMN, preceding SMN, random SMN, time-compressed SMN) repeated measures analysis of variance. Significant main effects were observed for the interruption method [$F(1,10) = 582.0$, $p < 0.001$, $\eta_p^2 = 0.98$], noise level [$F(1,10) = 15.0$, $p < 0.01$, $\eta_p^2 = 0.59$], and noise type [$F(1,10) = 8.2$, $p < 0.001$, $\eta_p^2 = 0.45$]. No interaction was found for interruption method and noise level ($p > 0.05$), but interactions were found for interruption method by noise type [$F(1,10) = 50.0$, $p < 0.0001$, $\eta_p^2 = 0.83$] and noise type by noise level [$F(1,10) = 9.5$, $p < 0.0001$, $\eta_p^2 = 0.49$]. A significant three-way interaction of interruption method, noise level, and noise type was also found [$F(4, 40) = 12.5$, $p < 0.0001$, $\eta_p^2 = 0.56$]. To further investigate the nature of this interaction, the effect of noise level, as well as noise type was examined below using subsequent *post hoc* tests.

A. Noise level effects

A main effect of noise level and significant interactions with noise level were found in the omnibus analysis of variance (ANOVA). This finding indicates that the intensity level of replacement noise influences the recognition of interrupted speech. Therefore, paired *t*-tests were used to compare performance between sentence-scaled and segment-scaled conditions for each interruption method and noise type. Tests were conducted using a Bonferroni-adjusted alpha level for multiple comparisons. For +SNR, performance was better in sentence-scaled conditions for unmodulated noise [$t(10) = 4.11$, $p < 0.05$], target SMN [$t(10) = 5.19$, $p < 0.01$], preceding SMN [$t(10) = 4.77$, $p < 0.01$], and time-compressed SMN [$t(10) = 3.93$, $p < 0.01$]. For -SNR, participants performed better in the sentence-scaled conditions for random SMN [$t(10) = 3.13$, $p < 0.01$] and time-compressed SMN [$t(10) = 4.55$, $p < 0.01$]. No other comparisons reached significance after Bonferroni correction ($p > 0.05$).

B. Comparison to silence conditions

Stimulus conditions were compared to silence in order to examine whether or not listeners received any perceptual benefit from hearing noise during the replacement intervals. The top row of Fig. 3 displays the average difference between the experimental conditions and the silence condition.

Bonferroni-adjusted paired *t*-tests were used to compare performance on the silence condition to the other noise replacement conditions. No benefit of noise was observed for the sentence-scaled +SNR conditions ($p > 0.05$). Note that in these conditions, the preserved speech is at or above the average sentence-level (and noise level). The only segment-scaled +SNR condition that resulted in a significant change in performance compared to silence was with preceding SMN, where performance was actually poorer with noise replacement [$t(10) = -3.89$, $p < 0.05$].

In sentence-scaled -SNR conditions, performance was better for time compressed SMN [$t(10) = 3.91$, $p < 0.05$] and target SMN [$t(10) = 4.78$, $p < 0.05$] compared to silence. In segment-scaled -SNR conditions, performance with target SMN was again significantly better than silence [$t(10) = 4.88$, $p < 0.05$]. However, random SMN [$t(10) = -6.50$, $p < 0.05$]

FIG. 2. Keyword recognition accuracy for the experimental conditions. (A) +SNR conditions preserved speech at positive local SNRs; (B) -SNR conditions preserved speech at negative local SNRs. Replacement noise levels were scaled to either the sentence or segment levels. Error bars = standard error of the mean. Asterisks indicate significantly different performance between the two scaling types (Bonferroni-adjusted $p < 0.05$). UNMOD refers to unmodulated noise. TARG (target SMN) and PRE (preceding SMN) had noise modulated by the target sentence. RAND (random SMN) and TC (time-compressed SMN) had noise modulated by a different sentence.

FIG. 3. The top panels display difference scores between noise and silent replacement for the different noise type conditions. +SNR conditions preserved speech at positive local SNRs, while -SNR conditions preserved speech at negative local SNRs. The bottom row displays the average level of the replacement noise relative to the average level of the original, uninterrupted target sentence. Note that all sentence-scaled conditions were equal to the average sentence level and are therefore at 0 dB. Error bars = standard error of the mean. UNMOD refers to unmodulated noise. TARG (target SMN) and PRE (preceding SMN) had noise modulated by the target sentence. RAND (random SMN) and TC (time-compressed SMN) had noise modulated by a different sentence.

and time-compressed SMN [$t(10) = -3.38$, $p < 0.05$] resulted in significantly lower performance than silence.

In summary, perceptual restoration due to noise replacement was only observed for a subset of -SNR conditions. These results can, in part, be explained by the level of the noise for each condition. These levels are shown for each condition in the bottom row of Fig. 3. Notice that noise levels were only above the average level of the uninterrupted sentence for a few conditions.

C. Comparison to unmodulated noise conditions

SMN conditions were also compared to unmodulated noise conditions to determine whether any of the experimental modulation conditions resulted in better or worse performance relative to “steady-state” modulation. For sentence-scaled +SNR, the only significant difference was observed for preceding SMN, which resulted in significantly poorer performance [$t(10) = -4.23$, $p < 0.05$]. In segment-scaled +SNR conditions, preceding SMN again resulted in worse performance than unmodulated noise [$t(10) = -5.22$, $p < 0.05$]. The only other condition significantly different from unmodulated noise was random SMN, which resulted in better performance [$t(10) = 5.72$, $p < 0.05$], likely due to a higher overall level of the noise.

Sentence-scaled -SNR conditions resulted in similar accuracy compared to unmodulated noise, except that better performance was obtained for the target SMN [$t(10) = 4.10$, $p < 0.05$]. Better performance was also observed for target SMN in the segment-scaled -SNR condition [$t(10) = 4.19$, $p < 0.05$]. In contrast, random SMN resulted in significantly worse performance than unmodulated noise [$t(10) = -9.08$, $p < 0.05$] along with significantly worse performance with time-compressed SMN [$t(10) = -5.11$, $p < 0.05$]. These

findings demonstrate significant effects of the type of SMN dependent on the noise level and relative level of the preserved speech segment.

IV. DISCUSSION

Interrupted speech provides a model for studying how listeners recognize speech based on partial information, which is necessary for understanding the perception of speech in noise. Properties of the interrupting background, such as the level and temporal properties of the noise (e.g., Bashford *et al.*, 1996), interact with the preserved speech to facilitate the continuity and recognition of the partial acoustic signal. The current study examined how the AM properties of an interrupting noise interact with the preserved speech cues during the recognition of interrupted speech. Results indicated that the AM properties of the interrupting noise significantly affect speech recognition. Importantly, AM properties of the noise interact with the specific speech cues that are preserved in the signal, and in some cases, determine the direction of the interaction such that the properties of both the noise and speech must be known for predicting speech recognition.

Overall, level effects of the alternating speech and noise intervals resulted in most performance changes, potentially because the relative levels of short phoneme- and syllable-duration intervals reflect the primary AM of the sentence (Greenberg *et al.*, 2003). Further enhancements in perception do occur for within-interval target speech modulations for high intensity (+SNR) speech intervals.

A. Interruption method effects

Intelligibility was higher for +SNR conditions than for -SNR conditions overall, regardless of modulation type or

noise level, with accuracy hovering around 60% correct for +SNR conditions and around 40% for -SNR conditions. Importantly, both interruption methods preserved 50% of the total sentence duration. Therefore, the difference in sentence recognition suggests that high-intensity portions of speech (preserved in the +SNR condition) contribute more to sentence recognition than low-intensity portions of speech. This is in line with previous studies that have utilized other interruption techniques. For example, studies that have interrupted speech based on vowel/consonant boundaries have demonstrated better performance for high-intensity vowel segments, compared to low-intensity consonant segments (e.g., Kewley-Port *et al.*, 2007; Fogerty and Kewley-Port, 2009). Similarly, studies interrupting speech based on the rms intensity found that speech recognition was best with the middle and high rms segments preserved (Chen and Loizou, 2012; Guan *et al.*, 2016). Indeed, high and middle rms levels correspond to vowel-dominated segments and vowel-consonant transitions (Guan *et al.*, 2016), respectively. Furthermore, even results based on cochlea-scaled entropy that indexes information change can be explained by simple differences in speech level (Oxenham *et al.*, 2017). Thus, evidence across a variety of interruption paradigms suggest the importance of high-intensity cues for recognizing interrupted speech, as observed here with the +SNR interruption method. This may help to explain why the perception of speech can be robust in noisy real-world conditions, because +SNR segments, defined here using a model of speech-on-speech masking, are most likely to be preserved.

B. Noise level effects

In general, if amplitude scaling increased the intensity of the replacement noise (for either segment- or sentence-scaled conditions) beyond the level of the preserved speech, performance improved (albeit not always significantly above silence or unmodulated noise performance). This is consistent with previous work indicating improvements in the recognition of periodically interrupted speech when the intensity level of the replacement unmodulated noise is at or above the average speech level (e.g., Bashford *et al.*, 1996). For the +SNR conditions tested here, this occurred very infrequently. As would be expected, when the noise was scaled to the average, uninterrupted sentence level (i.e., sentence-scaled) the noise level was less than the high-intensity +SNR segments. However, better performance was observed for the high-intensity noise created by segment-scaled random SMN, which resulted in better performance. In contrast, these same noise replacement levels were typically greater than the low-intensity -SNR preserved segments. Thus, the greatest improvements in intelligibility based on noise level were observed for the -SNR interruption method. Interestingly, the segment-scaled random SMN in this condition resulted in a very low-level noise and correspondingly lower performance. Thus, these differences in performance for random SMN are likely due to effects of the noise level rather than AM properties of the replacement noise. Random SMN is the best model for speech modulations of a competing talker. Here we found that interactions

for this type of masker are primarily level-dependent, rather than modulation-dependent. Furthermore, for the most realistic condition, +SNR, the high-intensity random SMN segments actually improved performance. However, this might not occur for perceptually segregated competing talkers or when considering potential additional factors of informational masking (for example, see Iyer *et al.*, 2007, regarding a difference between noise and speech maskers with interrupted speech).

C. Noise modulation effects

There are a number of conditions where the AM properties of the noise did seem to influence performance. Restoring the AM of the missing speech (i.e., target SMN) improves accuracy for -SNR conditions with both segment and sentence noise levels. This demonstrates that restoring the AM cues of high-intensity speech segments (i.e., those deleted in -SNR conditions) is beneficial for speech recognition. This is particularly apparent for sentence-scaled conditions, which controlled for level across noise types. However, restoring AM of the missing speech did not improve performance in +SNR conditions. This suggests that AM cues from low-intensity speech segments (i.e., the segments used to modulate the target SMN in +SNR conditions) provide more limited perceptual information for the listener. This observation is consistent with Fogerty (2014), who found that listeners performed significantly better when the replacement noise restored AM cues from vowels (i.e., high-intensity segments) but had limited benefit when the replacement noise restored consonant AM (i.e., low-intensity segments). Furthermore, this observation suggests that the advantage of restoring AM during the noise replacement intervals of periodically interrupted speech (Bashford *et al.*, 1996; Shinn-Cunningham and Wang, 2008) was likely due to restoration of AM for these high-intensity segments.

Using AM from a previous portion of the target speech (i.e., preceding SMN) resulted in poorer performance in +SNR conditions. This suggests that repeating high-intensity AM is detrimental to speech intelligibility, potentially due to distorting the typical modulatory rate of speech. In contrast, repeating low-intensity segment modulation trended toward improved performance for -SNR conditions. This later observation is consistent with findings from Fogerty *et al.* (2016a) who found that performance was improved for sentences interrupted to preserve low-intensity consonant segments when the AM characteristics of the replacement noise matched that of the preserved low-intensity speech.

Using time-compressed SMN in general resulted in the same or poorer performance compared to silence or unmodulated noise. This time-compressed condition was investigated to examine if significantly faster noise modulation would facilitate segregation from the slower modulations of speech. The current observation is consistent with this hypothesis in that introducing modulation cues which induce segregation results in poorer perceptual performance (i.e., due to a larger difference between the target and masker temporal modulation spectra). This is supported by current

theories of perceptual restoration that suggest it arises from a failure of stream segregation (e.g., Shinn-Cunningham and Wang, 2008). Furthermore, the faster modulated noise may have exaggerated the discontinuity of the interrupted speech, which could be detrimental to speech intelligibility (Bashford *et al.*, 1996). However, one of the four time-compressed conditions, sentence-scaled –SNR, resulted in improved performance. This may have been due to a level effect of the noise which for this condition was, on average, greater than the preserved –SNR segments. While the segment-scaled version was of similar average intensity, random variation in level across the sentence combined with the faster noise modulation may have facilitated segregation cues and reduced perceptual restoration.

D. Summary and conclusions

Speech is often heard in the presence of modulated noise that results in partial availability of the speech signal. The naturalistic glimpses of speech, defined by a model of speech-on-speech masking, were used in the current study to interrupt the speech resulting in the preservation of primarily high- or low-intensity speech segments. Using this interruption paradigm, this study investigated the interaction between the AM properties of the preserved speech and the intervening replacement noise. Results demonstrated that AM cues, particularly from high-intensity speech segments, are important for speech-in-noise perception. The benefit of these cues depends on the properties of both the preserved speech and of the interrupting noise.

These results are of course limited to the understanding of interrupted speech. Natural listening conditions involve continuous sources of speech and noise. Continuous noise, although not continuous competing speech, results in poorer speech recognition than the interrupted noise used here (Iyer *et al.*, 2009), likely due to effects of simultaneous masking. However, the method used here interrupted speech based on naturally defined glimpses that would have been preserved at favorable SNRs, which would limit the effect of simultaneous masking. Therefore, it is expected that a similar processes will apply for conditions of continuous noise masking, although future empirical studies are needed.

Overall, performance was best when high-intensity speech glimpses were preserved. Performance also generally improved when the replacement noise level exceeded the level of the preserved speech—not the average sentence level. In addition, the characteristics of the AM of the noise also influenced performance. Performance improved most often when the noise restored missing AM cues from high-intensity, but not low-intensity, segments of speech. However, modulations from preserved high-intensity segments were detrimental to performance when based on a different segment of target speech. Finally, fast-rate noise modulations induced through time-compression facilitated the segregation of speech and noise and interfered with perceptual restoration processes. These results indicate that the recognition of speech based on partial information depends on the particular intervals of speech that are glimpsed, in addition to the level and AM properties of the interfering

noise source. Significantly, the specific influence of noise AM depends on properties of the preserved speech, such that performance can only be explained by knowing the properties of both the speech and noise signals. Overall, conditions that preserve the slow varying changes in level across phonemes and syllables in the sentence result in the best recognition of interrupted speech.

ACKNOWLEDGMENTS

This work was supported, in part, by the National Institutes of Health, National Institute on Deafness and Other Communication Disorders, Grant Nos. R03-DC012506 and R01-DC015465. Portions of this data were presented at the 173rd meeting of the Acoustical Society of America, Boston, MA.

¹The lowpass filter envelope cutoff frequency used here for noise modulation is different from the envelope cutoff implemented in the SMN signal used to define speech interruptions. While both cutoffs were used to avoid the influence of faster periodic modulations in the analysis, here we were interested in restricting the modulation range to the primary energy associated with typical phonetic features (Ghitza, 2011). We expect the effect of this difference to be minimal as both cutoffs capture the primary energy in the speech modulation spectrum. Furthermore, the lower filter order used here, necessary for appropriate filtering of the short replacement segments, had a shallower filter slope and so potentially included some energy of faster AM beyond the cutoff frequency.

- Apoux, F., and Healy, E. W. (2013). “A glimpsing account of the role of temporal fine structure information in speech recognition,” in *Basic Aspects of Hearing* (Springer, New York), pp. 119–126.
- Bashford, J. A., Riener, K. R., and Warren, R. M. (1992). “Increasing the intelligibility of speech through multiple phonemic restorations,” *Percept. Psychophys.* **51**, 211–217.
- Bashford, J. A., and Warren, R. M. (1987). “Multiple phonemic restorations follow the rules for auditory induction,” *Percept. Psychophys.* **42**, 114–121.
- Bashford, J. A., Warren, R. M., and Brown, C. A. (1996). “Use of speech-modulated noise adds strong ‘bottom-up’ cues for phonemic restoration,” *Percept. Psychophys.* **58**, 342–350.
- Başkent, D., Eiler, C., and Edwards, B. (2009). “Effects of envelope discontinuities on perceptual restoration of amplitude-compressed speech,” *J. Acoust. Soc. Am.* **125**, 3995–4005.
- Benard, M. R., Mensink, J. S., and Başkent, D. (2014). “Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities,” *J. Acoust. Soc. Am.* **135**, EL88–EL94.
- Bernstein, J. G., and Grant, K. W. (2009). “Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners,” *J. Acoust. Soc. Am.* **125**, 3358–3372.
- Bhargava, P., Gaudrain, E., and Başkent, D. (2016). “The intelligibility of interrupted speech: Cochlear implant users and normal hearing listeners,” *J. Assoc. Res. Otolaryngol.* **17**, 475–491.
- Boersma, P. and Weenink, D. (2014). “Praat: Doing phonetics by computer (version 5.3.80) [Computer program],” <http://www.praat.org/> (Last viewed 29 June 2014).
- Brungart, D. S. (2001). “Informational and energetic masking effects in the perception of two simultaneous talkers,” *J. Acoust. Soc. Am.* **109**, 1101–1109.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (2001). “Informational and energetic masking effects in the perception of multiple simultaneous talkers,” *J. Acoust. Soc. Am.* **110**, 2527–2538.
- Chen, F., and Loizou, P. C. (2012). “Contributions of cochlea-scaled entropy and consonant-vowel boundaries to prediction of speech intelligibility in noise,” *J. Acoust. Soc. Am.* **131**, 4104–4113.
- Chen, F., Wong, M. L., Zhu, S., and Wong, L. L. (2015). “Relative contributions of vowels and consonants in recognizing isolated Mandarin words,” *J. Phon.* **52**, 26–34.

- Cole, R. A., Yan, Y., Mak, B., Fanty, M., and Bailey, T. (1996). "The contribution of consonants versus vowels to word recognition in fluent speech," in *Proceedings of the ICASSP'96*, May 7–10, Atlanta, GA, pp. 853–856.
- Cooke, M. (2003). "Glimpsing speech," *J. Phon.* **31**, 579–584.
- Cooke, M. (2006). "A glimpsing model of speech perception in noise," *J. Acoust. Soc. Am.* **119**, 1562–1573.
- Fogerty, D. (2011). "Perceptual weighting of individual and concurrent cues for sentence intelligibility: Frequency, envelope, and fine structure," *J. Acoust. Soc. Am.* **129**, 977–988.
- Fogerty, D. (2013). "Acoustic predictors of intelligibility for segmentally interrupted speech: Temporal envelope, voicing, and duration," *J. Speech Lang. Hear. Res.* **56**, 1402–1408.
- Fogerty, D. (2014). "Importance of envelope modulations during consonants and vowels in segmentally interrupted sentences," *J. Acoust. Soc. Am.* **135**, 1568–1576.
- Fogerty, D., Ahlstrom, J. B., Bologna, W. J., and Dubno, J. R. (2016a). "Glimpsing speech in the presence of nonsimultaneous amplitude modulations from a competing talker: Effect of modulation rate, age, and hearing loss," *J. Speech Lang. Hear. Res.* **59**, 1198–1207.
- Fogerty, D., and Kewley-Port, D. (2009). "Perceptual contributions of the consonant-vowel boundary to sentence intelligibility," *J. Acoust. Soc. Am.* **126**, 847–857.
- Fogerty, D., Xu, J., and Gibbs, B. E. (2016b). "Modulation masking and glimpsing of natural and vocoded speech during single-talker modulated noise: Effect of the modulation spectrum," *J. Acoust. Soc. Am.* **140**, 1800–1816.
- Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.* **110**, 1150–1163.
- Füllgrabe, C., Berthommier, F., and Lorenzi, C. (2006). "Masking release for consonant features in temporally fluctuating background noise," *Hear. Res.* **211**, 74–84.
- Gaudrain, E., and Carlyon, R. P. (2013). "Using Zebra-speech to study sequential and simultaneous speech segregation in a cochlear-implant simulation," *J. Acoust. Soc. Am.* **133**, 502–518.
- Ghitza, O. (2011). "Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm," *Front. Psychol.* **2**, 130.
- Gilbert, G., Bergeras, I., Voillery, D., and Lorenzi, C. (2007). "Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues," *J. Acoust. Soc. Am.* **122**, 1336–1339.
- Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). "Temporal properties of spontaneous speech—A syllable-centric perspective," *J. Phon.* **31**, 465–485.
- Guan, T., Chu, G. X., Tsao, Y., and Chen, F. (2016). "Assessing the perceptual contributions of level-dependent segments to sentence intelligibility," *J. Acoust. Soc. Am.* **140**, 3745–3754.
- Humes, L. E., Kidd, G. R., and Fogerty, D. (2017). "Exploring use of the coordinate response measure in a multitalker babble paradigm," *J. Speech Lang. Hear. Res.* **60**, 741–754.
- IEEE (1969). "IEEE recommended practice for speech quality measures," *IEEE Trans. Audio Electroacoust.* **17**, 225–246.
- Iyer, N., Brungart, D. S., and Simpson, B. D. (2007). "Effects of periodic masker interruption on the intelligibility of interrupted speech," *J. Acoust. Soc. Am.* **122**, 1693–1701.
- Kewley-Port, D., Burkle, T. Z., and Lee, J. H. (2007). "Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners," *J. Acoust. Soc. Am.* **122**, 2365–2375.
- Li, N., and Loizou, P. C. (2007). "Factors influencing glimpsing of speech in noise," *J. Acoust. Soc. Am.* **122**, 1165–1172.
- Loizou, P. C. (2007). *Speech Enhancement: Theory and Practice* (CRC Press, Taylor and Francis, Boca Raton, FL), p. 608.
- Miller, G. A., and Licklider, J. C. (1950). "The intelligibility of interrupted speech," *J. Acoust. Soc. Am.* **22**, 167–173.
- Molis, M. R., and Gallun, F. J. (2016). "The intelligibility of interrupted, time-compressed speech," *J. Acoust. Soc. Am.* **139**, 2190–2190.
- Nagaraj, N. K., and Knapp, A. N. (2015). "No evidence of relation between working memory and perception of interrupted speech in young adults," *J. Acoust. Soc. Am.* **138**, EL145–EL150.
- Nelson, P. B., and Jin, S. H. (2004). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **115**, 2286–2294.
- Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (2003). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **113**, 961–968.
- Oxenham, A. J., Boucher, J. E., and Kreft, H. A. (2017). "Speech intelligibility is best predicted by intensity, not cochlea-scaled entropy," *J. Acoust. Soc. Am.* **142**, EL264–EL269.
- Powers, G. L., and Speaks, C. (1973). "Intelligibility of temporally interrupted speech," *J. Acoust. Soc. Am.* **54**, 661–667.
- Powers, G. L., and Wilcox, J. C. (1977). "Intelligibility of temporally interrupted speech with and without intervening noise," *J. Acoust. Soc. Am.* **61**, 195–199.
- Rhebergen, K. S., and Versfeld, N. J. (2005). "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," *J. Acoust. Soc. Am.* **117**, 2181–2192.
- Rosen, S. (1992). "Temporal information in speech: Acoustic, auditory and linguistic aspects," *Philos. Trans. R. Soc. London, Ser. B* **336**, 367–373.
- Shafiro, V., Sheft, S., and Risley, R. (2016). "The intelligibility of interrupted and temporally altered speech: Effects of context, age, and hearing loss," *J. Acoust. Soc. Am.* **139**, 455–465.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wyganski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Shinn-Cunningham, B. G., and Wang, D. (2008). "Influences of auditory object formation on phonemic restoration," *J. Acoust. Soc. Am.* **123**, 295–301.
- Simpson, S. A., and Cooke, M. (2005). "Consonant identification in N-talker babble is a nonmonotonic function of N," *J. Acoust. Soc. Am.* **118**, 2775–2778.
- Smiljanic, R., Sheft, S., Chandrasekaran, B., and Shafiro, V. (2016). "Effect of language experience on the intelligibility of interrupted speech," *J. Acoust. Soc. Am.* **139**, 2189–2189.
- Smith, K. G., and Fogerty, D. (2017). "Speech recognition error patterns for steady-state noise and interrupted speech," *J. Acoust. Soc. Am.* **142**, EL306–EL312.
- Stilp, C. E. (2014). "Information-bearing acoustic change outperforms duration in predicting intelligibility of full-spectrum and noise-vocoded sentences," *J. Acoust. Soc. Am.* **135**, 1518–1529.
- Stilp, C. E., and Kluender, K. R. (2010). "Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility," *Proc. Natl. Acad. Sci.* **107**, 12387–12392.
- Stone, M. A., Füllgrabe, C., and Moore, B. C. (2012). "Notionally steady background noise acts primarily as a modulation masker of speech," *J. Acoust. Soc. Am.* **132**, 317–326.
- Wang, X., and Humes, L. E. (2010). "Factors influencing recognition of interrupted speech," *J. Acoust. Soc. Am.* **128**, 2100–2111.
- Warren, R. M. (1970). "Perceptual restoration of missing speech sounds," *Science* **167**, 392–393.
- Wilson, R. H., and Irish, S. E. (2015). "Recognition performance on words interrupted (10 ips, 50% duty cycle) with two interruption patterns referenced to word onset: Young listeners with normal hearing for pure tones and older listeners with sensorineural hearing loss," *Int. J. Audiol.* **54**, 933–941.
- Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargava, A., Wei, C., and Cao, K. (2005). "Speech recognition with amplitude and frequency modulations," *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2293–2298.