Glimpsing speech in temporally and spectro-temporally modulated noise

Daniel Fogerty^{a)}

Department of Communication Sciences and Disorders, University of South Carolina, 1224 Sumter Street, Columbia, South Carolina 29208, USA

Brittney L. Carter and Eric W. Healy

Department of Speech and Hearing Science, The Ohio State University, 1070 Carmack Road, Columbus, Ohio 43210, USA

(Received 11 January 2018; revised 6 April 2018; accepted 30 April 2018; published online 23 May 2018)

Speech recognition in fluctuating maskers is influenced by the spectro-temporal properties of the noise. Three experiments examined different temporal and spectro-temporal noise properties. Experiment 1 replicated previous work by highlighting maximum performance at a temporal gating rate of 4–8 Hz. Experiment 2 involved spectro-temporal glimpses. Performance was best with the largest glimpses, and performance with small glimpses approached that for continuous noise matched to the average level of the modulated noise. Better performance occurred with periodic than for random spectro-temporal glimpses. Finally, time and frequency for spectro-temporal glimpses were dissociated in experiment 3. Larger spectral glimpses were more beneficial than smaller, and minimum performance was observed at a gating rate of 4–8 Hz. The current results involving continuous speech in gated noise (slower and larger glimpses most advantageous) run counter to several results involving gated and/or filtered speech, where a larger number of smaller speech samples is often advantageous. This is because mechanisms of masking dominate, negating the advantages of better speech-information sampling. It is suggested that spectro-temporal glimpsing combines temporal glimpsing with additional processes of simultaneous masking and uncomodulation, and continuous speech in gated noise is a better model for real-world glimpsing than is gated and/or filtered speech. © 2018 Acoustical Society of America.

<https://doi.org/10.1121/1.5038266>

[JLJ]

Pages: 3047–3057

I. INTRODUCTION

Investigating speech recognition in fluctuating backgrounds has been of significant interest over the last several decades. Perhaps best highlighted by the classic study of Miller and Licklider (1950), listening in fluctuating backgrounds involves processing speech based on partial acoustic information distributed across the amplitude dips in the fluctuating masker. Among the various conditions tested by Miller and Licklider (1950), one method was to modulate a steady-state noise by a square wave to create “interrupted noise,” also called “gated noise” (e.g., Nelson *et al.*, 2003). The method has been extended to other types of amplitude modulators, including amplitude modulation by sine waves (e.g., Gnansia *et al.*, 2008) or by the temporal envelope of speech (e.g., Fogerty *et al.*, 2016). A pervasive observation has been that normal-hearing listeners have better speech recognition in modulated noise than in unmodulated (i.e., steady-state) noise (e.g., Miller and Licklider, 1950; Festen and Plomp, 1990; Takahashi and Bacon, 1992; Howard-Jones and Rosen, 1993a; Gustafsson and Arlinger, 1994; Nelson *et al.*, 2003; Füllgrabe *et al.*, 2006; Fogerty *et al.*, 2016). The better performance in modulated noise has been termed “masking release” (e.g., Nelson *et al.*, 2003) or the “fluctuating masker

benefit” (e.g., Bernstein and Grant, 2009). One primary factor that determines the benefit obtained is the gating rate of the masker (i.e., the number of times the noise is turned “on” and “off” in one second). For example, Miller and Licklider (1950) found that performance can vary by as much as 70 percentage-points as a function of the noise temporal gating rate.

In such temporally modulated backgrounds, listeners presumably obtain glimpses of speech preserved at favorable signal-to-noise ratios (SNRs) within the dips of the masker. However, natural environments produce glimpses that are distributed in both time and frequency. Thus, it is important to characterize the effect of spectro-temporal glimpse distributions as a function of the temporal gating rate and the spectral-glimpse bandwidth. Toward this end, Howard-Jones and Rosen (1993b) created a “checkerboard noise” (referring to its representation in the spectrogram view) in which noise was gated on and off within frequency bands, with adjacent bands presented out-of-phase, i.e., uncomodulated. Buss and colleagues (Buss *et al.*, 2004; Hall *et al.*, 2008a,b; Ozmeral *et al.*, 2012) have since followed up this initial work by comparing bands that are modulated synchronously across frequencies (i.e., gated) or asynchronously (i.e., checkerboard). Evidence from these spectro-temporal studies suggests that listeners with normal and impaired hearing are able to integrate glimpses across time and frequency. However, performance

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

declines as the spectral bandwidth of the noise decreases (Howard-Jones and Rosen, 1993b; Ozmeral *et al.*, 2012). What is not known is how the spectral distribution of glimpses interacts with factors associated with the temporal gating rate. This is because these studies have focused on spectro-temporal glimpsing at only one or two temporal gating rates (e.g., 10 or 20 Hz). Given the large variability in performance for temporally gated noise as a function of gating rate (e.g., Miller and Licklider, 1950), it seems likely that the gating rate may influence the spectro-temporal integration of speech glimpses.

Furthermore, as indicated by Ozmeral *et al.* (2012), a limitation of most of these modulated-noise studies has been the use of periodic, and therefore predictable, noise modulations. However, natural-listening environments are likely to involve more random spectro-temporal fluctuations. This was initially recognized by Miller and Licklider (1950) who included some conditions using a random-gating noise. They concluded that such random gating generally results in a smoothing of the performance function across temporal gating rates. However, these conditions were limited. For example, random noise gating was investigated only at an intermediate SNR where listeners were able to resolve some speech information during the energetically masked intervals of the noise as well as during the masker dips.

In the intervening decades, it has become well established that the predictability of auditory information facilitates auditory perception (e.g., Hickok *et al.*, 2015). Indeed, predictable rhythms provide a cue for the future arrival of critical bits of information (e.g., Engel *et al.*, 2001; Giraud and Poeppel, 2012). Findings such as this led to the development of rhythmic-attention theory (Jones, 1976), which suggests that listeners process acoustic regularities over time. Indeed, humans are particularly adept at perceiving and tracking regularities in the acoustic environment (e.g., Jones, 1976; Jones *et al.*, 2002; Arnal *et al.*, 2014; ten Oever *et al.*, 2014). Significantly, listeners also appear specifically attuned to the temporal patterns of a masker. Hickok *et al.* (2015) presented a noise modulated at low frequency, followed by a brief 1-kHz tone embedded in an unmodulated noise. Participants were required to detect the tone. They found that listeners had better detection thresholds when the tone was presented in a position that was out-of-phase with the prior modulated noise. This suggests that the predictable rhythmic context facilitated the processing of acoustic information in phase with the dips of the masker. That is, tone detection was best based on the periodicity of the masker dips when energetic masking was the least—even after the noise modulation was extinguished. Recent neurophysiological evidence demonstrating neural entrainment to an auditory stimulus might underlie these behavioral observations (e.g., Ward, 2003; Neuling *et al.*, 2012; Lakatos *et al.*, 2013; Weisz and Obleser, 2014). These findings together suggest that the predictability of fluctuating backgrounds might have a significant influence on the ability of listeners to successfully listen in the dips of the modulated masker, a hypothesis explicitly investigated within the current paradigm.

The present study was designed to extend the classic results of Miller and Licklider (1950) to comprehensively

examine the effect of gating rate for periodic (predictable) and random temporal gating (experiment 1). Furthermore, it extends these results to examine the effect of spectro-temporal glimpse size (experiment 2) and to examine the interaction of temporal gating rate with spectral bandwidth (experiment 3). Thus, the current study extends the traditional gating methods originated by Miller and Licklider (1950) to more natural listening backgrounds that are spectro-temporally complex and variable.

II. EXPERIMENT 1: TEMPORALLY GATED NOISE

The goal of experiment 1 was to comprehensively examine the effect of gating rate for periodic (predictable) and random temporal gating. Previous studies (e.g., Miller and Licklider, 1950) have suggested a minimal effect of gating predictability. This experiment directly investigated this effect across gating rates and served as a baseline temporal comparison to the subsequent experiments involving spectro-temporal gating.

A. Participants

Two groups of ten normal-hearing young adult listeners participated in the experiment. The first group (mean age = 20.5 yr, 19–23 yr) completed testing at The Ohio State University and the second group (mean age = 20.2 yr, 19–22 yr) was recruited to test additional conditions at the University of South Carolina. All listeners were native speakers of English and had audiometric thresholds of 20 dB hearing level (HL) or better at octave frequencies between 250 and 8000 Hz (ANSI, 2004, 2010). Care was taken to ensure that no subjects had prior exposure to the sentence materials.

B. Stimuli and design

Speech stimuli consisted of sentences spoken by the standard male talker from the Hearing in Noise Test (HINT; Nilsson *et al.*, 1994), which consists of 25 lists of 10 sentences each plus practice lists of 10 sentences each. A continuous speech-shaped noise (SSN) was created that matched the long-term average spectrum of a concatenation of all HINT sentences, 11.275 min in duration. Gated noise for the first group was created by turning the continuous SSN on or off at the following rates (1, 2, 4, 8, 16 Hz) with a duty cycle of 50%. The noise was turned on and off using a 5-ms raised-cosine ramp, such that the “on” period of each noise interval was measured from the 6-dB down point of the modulated rise or fall. This processing resulted in the noise used for the periodic-gating condition. A random-gating condition was created for each interruption rate by circularly shifting each full period of the gating cycle by a random phase. Again, the raised-cosine ramp was applied to the onsets/offsets of the random-gated noise. Figure 1 displays example spectrograms for periodic and random temporal-gating conditions. This resulted in ten conditions (5 gating rates $\times$ 2 predictability levels). All stimuli were presented at a sampling rate of 16 kHz.

FIG. 1. Spectrograms of example noise files for periodic (top) and random (bottom) temporal gating at lower (2 Hz, left) and higher (8 Hz, right) gate rates.

Otherwise identical stimuli were created for the second group of listeners who completed testing with a different range of gating rates. This second group was tested on periodic and random gated noise at 16, 32, and 64 Hz for a total of six conditions.

The presentation level for non-gated SSN was set to 70 dBA and the sentence lists were scaled to produce an overall SNR of -18 dB. As in all experiments described currently, SNR was selected to produce intelligibility scores generally free of floor and ceiling effects.

To demonstrate the modulation properties of the noise, the temporal modulation spectrum for each noise masker was analyzed. To facilitate comparison to the spectro-temporal noise used in the subsequent experiments, analysis was conducted on a $[1/2]$ -ERB_N band (ERB_N=normal equivalent rectangular bandwidth; Glasberg and Moore, 1990) centered at 1943 Hz. The temporal envelope of the noise was extracted by halfwave rectification followed by a sixth-order 128 Hz low-pass Butterworth filter. The temporal envelope was then downsampled to 1000 Hz and the fast Fourier transform (FFT) was calculated to derive the modulation spectrum. The modulation spectrum for each masker resulted in a fundamental modulation peak at the gating frequency, in addition to modulation at odd harmonics of the gate frequency due to the square-wave gating function. Figure 2 displays the energy of the fundamental modulation peak for each of the seven different maskers (i.e., each peak represents the peak modulation for a different gated masker). The top and bottom panels display this modulation analysis for the periodic and random gated maskers, respectively. As expected, random gating was characterized by broader modulation peaks.

C. Procedure

For the first group, the five gating rates were randomized and balanced such that each appeared in each

FIG. 2. Fundamental modulation peak from the gated noise modulation spectrum for each of the seven gating rates (1–64 Hz). The top and bottom panels display the modulation spectrum for periodic and random temporal gating, respectively.

presentation-order position an equal number of times across subjects. The presentation of periodic and random conditions at each rate was juxtaposed, such that half the subjects heard periodic followed by random at each rate, and the other half heard the opposite order. The presentation of the noise began several seconds before speech onset and was continuously on during presentation of each set of ten sentences. This ensured a sufficient temporal buildup for processing any predictability of the noise. A different random start point in the gated noise was employed for each subject, to produce a unique alignment between speech and noise gating for each subject. Further, the sentence list-to-condition correspondence was different for each subject.

Testing began with practice consisting of five HINT sentences in a quiet background, followed by ten HINT sentences in the first-heard condition. Sentences used for practice were not employed for testing. This was followed by the 10 blocks of conditions, with 20 sentences in each condition (HINT lists 1–20). No sentences were ever repeated. Stimuli were presented diotically using a personal computer (PC) equipped with high quality digital-to-analog (D/A) converters and Sennheiser HD 280 Pro headphones (Wedemark, Germany). The headphones were calibrated using a Larson Davis sound-level meter and headphone coupler. Participants were tested while seated with the experimenter in a sound-attenuating booth. They were instructed to repeat each sentence as best they could and to guess if necessary. The experimenter controlled the presentation of stimuli and recorded correct responses.¹

The procedures for the second test group were essentially identical to those of the first group. The Ohio State University and University of South Carolina laboratories were similarly equipped—all software and all described hardware were identical at both locations. The three gating rates employed for this group were randomized and presented in balanced order across subjects, periodic/random were again juxtaposed with half of the subjects hearing periodic before random at each rate and the other half hearing the opposite order, and the sentence list-to-condition correspondence was randomized for each subject. Again, an SNR of -18 dB and 20 sentences per condition were employed (HINT lists 1–12). Practice, calibration, and stimulus presentation methods were the same as those employed for the first test group.

D. Results and discussion

1. Current study

Figure 3 plots sentence intelligibility for the two listener groups across temporal gating rates for periodic and random temporal gating. For analysis, results were first transformed to rationalized arcsine units (RAUs; [Studebaker, 1985](#)) to stabilize the error variance. RAU performance for the first participant group that was tested at the lower gating rates was examined using a 5 (gating rate) $\times$ 2 (predictability) repeated-measures analysis of variance (ANOVA). Results demonstrated a significant main effect of the temporal gating rate [$F(4,36) = 7.5$, $p < 0.001$, $\eta_p^2 = 0.46$] and predictability, [$F(1,9) = 5.1$, $p < 0.05$, $\eta_p^2 = 0.36$]. A significant interaction was also found [$F(4,36) = 2.7$, $p < 0.05$, $\eta_p^2 = 0.23$], with better performance obtained for periodic interruptions at higher gating rates around 16 Hz. As observed in Fig. 3, performance peaked at 4–8 Hz for periodic and 2–8 Hz for random gating. Paired Bonferroni t -tests between adjacent gating rates demonstrated poorest performance for 1 Hz and 16 Hz for periodic gating and 16 Hz for random gating ($p < 0.05$).

To examine performance at a higher range of gating rates, a 3 (gating rate) $\times$ 2 (predictability) repeated-measures

ANOVA was conducted on RAU data from the second group of participants. Results again demonstrated a significant main effect of rate [$F(2,18) = 121.0$, $p < 0.001$, $\eta_p^2 = 0.93$] and predictability [$F(1,9) = 14.8$, $p < 0.01$, $\eta_p^2 = 0.62$]. But no significant interaction was found. These results together suggest that individuals may receive a small benefit for temporal-glimpse predictability in gated noise. Examination of Fig. 3 suggests that this advantage may be largest when gating between 8 and 32 Hz, but that performance between periodic and random-gating conditions becomes more closely matched at rates outside this range.

Finally, the replication of results across two different laboratories was examined at the 16-Hz gating rate. A 2×2 mixed-model ANOVA was conducted to examine the effect of predictability for the two listener groups at 16 Hz. Results demonstrated a main effect of predictability, [$F(1,18) = 7.1$, $p < 0.05$, $\eta_p^2 = 0.28$]. Whereas the second group of listeners did perform more poorly, no main effect of group or interaction was observed ($p > 0.05$). These results help to confirm that, at certain gating rates, performance is better when temporal glimpses are predictable.

The overall pattern of intelligibility across gating rates may be understood through the limiting cases largely observed currently. It should be expected that performance at a 50% duty cycle (and zero performance in non-gated SSN) will converge at 50% at very low temporal gating rates. Consider the limiting case of an extremely low rate that masks/spares every other sentence or even set of sentences—this would produce 50% intelligibility. At the upper end of the gating-rate continuum, noise-free speech segments become very brief and larger proportions of each segment fall into the time window closely following masker offset where non-simultaneous masking is at a maximum. Thus, they become more susceptible to forward masking by the noise immediately prior (and perhaps subject to backward masking as well). This non-simultaneous masking likely contributes to the very low intelligibility observed at this higher end of the rate continuum (see [Dubno et al., 2003](#)). This is in contrast to gated speech, which reaches a maximum at these very fast rates when masking is absent ([Miller and Licklider, 1950](#)).

2. Comparison to prior work

The results from experiment 1 are highly consistent with previously established findings using gated noise. Figure 4 replots performance measured in the current study along with the results for phonetically balanced words in [Miller and Licklider \(1950\)](#) and for IEEE sentences used by [Nelson and Jin \(2004\)](#), both tested at -18 dB SNR. Due to the overall greater difficulty of IEEE sentences, results from this latter study were shifted to match mean performance in the current study across the gating rates in common. Results across the three studies are highly similar, with performance reaching a maximum peak around 8 Hz. Thus, while overall performance level may be determined by speech materials, the effect of material does not appear to interact with the effect of gating rate. This inverted U-shaped function was also obtained in a consonant-recognition test at -15 dB SNR ([Füllgrabe et al., 2006](#)) and for closed-set matrix sentences

FIG. 3. Speech intelligibility across the different gating rates. The first listener group was tested at rates 1–16 Hz and the second listener group was tested at rates 16–64 Hz. Error bars indicate the standard error of the mean.

FIG. 4. Gated-speech intelligibility for three different studies. Performance for Nelson and Jin (2004) was shifted (+18.6 percentage-points) to equate mean performance with the current study, likely due to different speech materials.

(Gustafsson and Arlinger, 1994), although the latter materials resulted in a performance maximum shifted to a higher gating rate. In general, performance functions for gated noise at higher gating rates (~ 100 Hz) approach that for unmodulated noise (see also Dirks *et al.*, 1969; Wilson and Carhart, 1969; Howard-Jones and Rosen, 1993a).

For random gating, comparisons with the previous literature are less direct. Miller and Licklider (1950) found the random-gating function to be smoother across gating rates than the periodic function. Likewise, Gustafsson and Arlinger (1994) tested an irregular noise modulation composed of multiple sine-wave modulators. This condition resulted in poorer performance compared to periodic interruption, but it is also confounded by decreases in the modulation depth of the stimulus. Using a different method, Ghitza and Greenberg (2009) inserted periodic silent intervals into sentences time-compressed to one-third the original duration. They found that performance was best when inserting silence at 8 Hz, matching the performance peak obtained using the current methods. However, random insertion of silence produced more response errors at this peak rate. This aligns with the current observation that random gating produced the greatest decrement in performance at the rates that maximized performance.

III. EXPERIMENT 2: CHECKERBOARD NOISE

Experiment 1 investigated glimpsing of speech within a temporally gated noise. However, natural listening environments are more likely to provide glimpses that are distributed in both time and frequency. Experiment 2 was designed to investigate spectro-temporal glimpsing as a function of the glimpse size. Accordingly, the primary goal of experiment 2 was to test the alternate hypotheses of whether performance is better for a larger number of smaller glimpse sizes that provide multiple spectro-temporal “looks” or for fewer but larger spectro-temporal glimpses that each contain more information.

A. Participants

Ten new normal-hearing young adult listeners (mean age = 20.7 yr, 18–24 yr) from The Ohio State University participated in the experiment. All listener characteristics were identical to those in experiment 1.

B. Stimuli and procedure

HINT sentences were again used as speech materials. For this experiment, checkerboard noise was created with periodic and random gating. Five different spectro-temporal glimpse conditions were created. Because noise gating in every condition involved a 50% duty cycle, these may be thought of as a continuum with a smaller number of large checkerboard squares of noise at one end and a larger number of small squares at the other end. Values were 1 Hz/3 ERB_N , 2 Hz/2.5 ERB_N , 4 Hz/2 ERB_N , 8 Hz/1.5 ERB_N , 16 Hz/1 ERB_N , ranging from the largest spectro-temporal glimpse to the smallest.

Noise processing consisted of first bandpass filtering the SSN from experiment 1 into 60 contiguous $[1/2]-ERB_N$ bands ranging from 80 to 7562 Hz. Filtering was completed using a Hamming window with a linear-phase finite impulse-response bandpass filter. The filter order was equal to $20 \times fs/BW$, where $fs = 16\,000$ Hz and BW = filter bandwidth in Hz. Each band was subsequently turned on or off by the gating rate corresponding to the experimental condition. Frequency bands were also grouped into consecutive bands according to the glimpse bandwidth. For example, the 2- ERB_N condition involved grouping four adjacent $[1/2]-ERB_N$ bands together, with each of these four bands turned on and off in phase. To create periodic checkerboard noise, the second half of the on/off gating cycle involved the same group of four $[1/2]-ERB_N$ bands turned on and off together using a 180° shifted starting phase. For the random-gating condition, the starting phase of the temporal gating within each full gating cycle was circularly shifted by a random phase as in experiment 1. In addition, the spectral dimension was also randomized. Two adjacent spectral regions (e.g., each with a 2- ERB_N bandwidth) were grouped together with half the region on and half the region off. Spectral randomization occurred by circularly shifting whichever group of $[1/2]-ERB_N$ bands were turned on or off within this grouped spectral region. Thus, for the periodic presentation with a 2- ERB_N bandwidth, at a given moment in time, consecutive bands may have had the following state: [1 1 1 1 0 0 0 0], where 1 indicates a band is on and 0 indicates it is off. In contrast, after spectral randomization, this spectral region may have been [0 0 1 1 1 1 0 0]. A different spectral randomization was used for each temporal-gating period and each grouped spectral region. As in experiment 1, the noise was turned on and off using a 5-ms raised-cosine ramp, such that the on period of each noise interval was measured from the 6-dB down point of the modulated rise or fall. Figure 5 displays the spectrograms for different example checkerboard noises. Analysis of the periodic and random masker modulation spectra within each $[1/2]-ERB_N$ band was consistent with the analysis displayed in Fig. 2 for temporal gating. However, due to the unmodulated gating across bands, the wideband temporal envelope resembled that of SSN.

FIG. 5. Spectrograms of example checkerboard noise files for periodic (top) and random (bottom) gating for larger (left) and smaller (right) spectro-temporal glimpses.

Twenty sentences were used in each of the 10 conditions (i.e., 5 glimpse sizes $\times$ 2 levels of predictability) for a total of 200 trials (HINT lists 1–20). These ten conditions were blocked and randomized using the procedures of experiment 1. Again, a different random start point in the checkerboard noise was employed to produce a unique alignment between speech and noise for each subject, and sentence list-to-condition correspondence was different for each subject. The presentation level was 70 dBA and the speech was scaled to produce an overall SNR of -6 dB. All subjects also heard 20 HINT sentences (2 lists drawn randomly from lists 21–25) in non-gated SSN at the same overall SNR as the checkerboard noise. Because the SNR employed in the current experiment was expected to produce intelligibility scores above floor values, half of the subjects were presented with this SSN condition following the ten test blocks, and the other half were presented with this condition prior to the ten test blocks. Practice was the same as in experiment 1, except that 5 practice SSN sentences were added to immediately precede the 20 test SSN sentences. All calibration, apparatus, and other testing procedures were identical to those in experiment 1.

C. Results and discussion

Figure 6 plots sentence intelligibility across spectro-temporal glimpse size for periodic and random gating. Conditions progress left-to-right from largest-to-smallest spectro-temporal glimpse size. In comparison to the results of experiment 1 (see Fig. 3), these findings demonstrate a markedly different function across gating rates, with a performance minimum around 4–8 Hz, compared to the peak performance at these rates obtained in experiment 1. However, results are generally consistent with poorer performance at faster gating rates. Accordingly, spectro-temporal glimpsing appears to

FIG. 6. Speech intelligibility across the different spectro-temporal glimpse sizes. Conditions progress left-to-right from largest-to-smallest spectro-temporal glimpses. The two dashed lines indicate performance in steady-state noise at -6 dB SNR (top) and -9 dB SNR (bottom). Checkerboard noise was presented at a long-term average level of -6 dB SNR with the on portions of noise equated to -9 dB SNR. Error bars indicate the standard error of the mean.

involve factors in addition to temporal glimpsing. The two dashed lines in Fig. 6 indicate performance in SSN at -6 dB SNR (top) and -9 dB SNR (bottom, data from experiment 3). Gated noise was presented at a long-term average level of -6 dB SNR, with the on portions of noise equated to -9 dB SNR. Thus, scores above the bottom dashed line indicates that performance improved due to the introduction of spectro-temporal holes in the continuous noise. From Fig. 6, it can be observed that listeners obtain a benefit from all spectro-temporal glimpses tested. Indeed, even in the poorest conditions, performance in periodic checkerboard noise provided a 3-dB benefit in SNR over continuous noise (performance in checkerboard noise matched that in continuous noise at a 3-dB more favorable SNR). However, performance was best with the largest glimpse size (and correspondingly largest holes), and a clear trend toward better performance is apparent with predictable, periodic spectro-temporal glimpses.

These observations were first examined using a 5 (spectro-temporal glimpse size) $\times$ 2 (predictability) repeated-measures ANOVA on the RAU data. Results demonstrated a significant main effect of glimpse size [$F(4,36)=93.0$, $p<0.001$, $\eta_p^2=0.91$] and predictability [$F(1,9)=65.4$, $p<0.001$, $\eta_p^2=0.88$]. A significant interaction was also observed [$F(4,36)=7.9$, $p<0.001$, $\eta_p^2=0.47$].

Paired comparisons were made using paired Bonferroni t -tests. Comparisons between adjacent glimpse sizes demonstrated that performance in periodic noise dropped significantly only from 2 Hz/2.5 ERB_N to the smaller glimpse size of 4 Hz/2 ERB_N, $p<0.001$. Performance was then maintained at the smaller glimpse sizes. The observed difference in performance between 1 Hz/3 ERB_N and 2 Hz/2.5 ERB_N was not significant ($p>0.05$). In contrast, for random spectro-temporal glimpsing, performance was poorest for 4 Hz/2 ERB_N and 8 Hz/1.5 ERB_N, with better performance at progressively smaller and larger glimpse sizes ($p<0.001$).

Paired comparisons between the two levels of predictability demonstrated better performance for periodic spectro-temporal glimpses for glimpse sizes in the middle of the tested range ($2\text{ Hz}/2.5\text{ ERB}_N$ – $8\text{ Hz}/1.5\text{ ERB}_N$, $p < 0.001$). These results indicate large effects of spectro-temporal glimpse predictability at these glimpse sizes, ranging from 12 to 23 percentage-points.

Spectro-temporal glimpses in the current experiment were randomized according to both the spectral-band cutoff and the temporal interval. Thus, the independent contribution of either dimension is unclear. However, given that some effect of randomization was observed for the temporal dimension in experiment 1, and that the effect observed here was larger, it is likely that both dimensions contributed to the overall poorer performance of random spectro-temporal glimpses. Indeed, random spectral-band selection based on glimpses in multi-talker babble also results in poorer performance compared to when glimpses are taken from the same frequency region over time (Li and Loizou, 2007). It is also important to note that the spectro-temporal randomization method employed here used a single temporal randomization across all spectral bands, to best match the wideband temporal gating from experiment 1. However, this methodological choice means that spectro-temporal glimpses were not entirely random, but instead, involved temporal coherence across bands. Alternatively, we could have chosen to independently randomize temporal gates for each spectral band, which would have introduced spectral coherence across time. The potential effect of this temporal coherence (or of potential spectral coherence) is not clear from the present data. However, performance does appear to be detrimentally influenced by increased glimpse variability (i.e., decreased predictability) in the spectro-temporal masker. Future studies are required to investigate the perceptual effect of different properties of randomization that might give rise to glimpse variability.

It is further notable that all conditions used a 50% duty cycle and were matched in terms of SNR. However, large differences in performance were obtained based on the size of the spectro-temporal glimpses. Again, performance is clearly best with large glimpses. But the difference between periodic and random glimpses is attenuated at the largest glimpse size, potentially due to the interaction between the gating rate and sentence duration. HINT sentences have an approximate duration of 1.5 s. With a gating rate of 1 Hz, one period would have encompassed two-thirds of the total sentence duration. Thus, random deviations across periods of the gated noise may have a limited effect on participant performance. Nevertheless, performance with this large spectro-temporal glimpse size resulted in sizable improvements in performance relative to SSN. As glimpse sizes get smaller, performance appears to approximate performance in SSN normalized to the long-term average level of the gated noise. This is especially true for periodic spectro-temporal gating. Thus, whereas the on portions of the noise were presented at -9 dB SNR, as only 50% of the noise was presented, there was an average drop in noise level of 3 dB. Faster spectro-temporal gating rates approximated performance in SSN at this -6 dB level. This may be because the higher-rate checkerboard noise is continuously masking the speech,

albeit in different spectral bands at different time points. This is in contrast to the far superior performance observed for temporally gated noise (experiment 1) when listeners always receive some temporal intervals of clean broad-frequency speech. This latter point, the superior masking of speech by spectro-temporal noise relative to temporally gated noise, both at 50% duty cycle, is made clear by the large differences in SNR required to approximately equate performance across experiments 1 and 2 (-18 dB for gated noise and -6 dB for checkerboard noise) and the large difference across experiments in the gated- versus continuous-noise performance difference (large difference for gated noise and smaller difference for checkerboard noise). These differences can be understood in terms of the mechanism of masking involved. In temporally gated noise, all of the masking is non-simultaneous (forward and perhaps backward masking), and this is true in every frequency band. This contrasts with spectro-temporally gated noise, in which masking is both non-simultaneous within frequency bands and simultaneous across frequency bands. The considerable and well established superiority of simultaneous masking over non-simultaneous masking likely accounts for the overall difference in masking potential observed across experiments 1 and 2.

IV. EXPERIMENT 3: SPECTRO-TEMPORAL GLIMPSE SIZE

Experiment 2 demonstrated a general decline in performance as the spectro-temporal glimpse size decreased. However, the creation of checkerboard noise having different sized squares confounds spectral-hole size with temporal-gating rate. Experiment 3 was designed to dissociate these two factors to obtain a better understanding of the independent spectral and temporal factors underlying glimpsing in periodic checkerboard noise. Therefore, the goal of experiment 3 was to delineate whether the better performance for larger glimpse sizes observed in experiment 2 was due to the effect of a slower gating rate, narrower spectral bands, or a combination of both factors.

A. Participants

Nine new normal-hearing young adult listeners (mean age = 20.2 yr, 20–21 yr) from The Ohio State University participated and had characteristics matching those employed in experiment 1.

B. Stimuli and procedure

HINT sentences were again used for experiment 3. Noise processing followed the same stimulus processing procedures as in experiment 2 for the periodic checkerboard noise. This experiment examined the effect of spectral-hole size across different temporal-gating rates using a $5\text{ (1, 2, 4, 8, 16 Hz)} \times 3\text{ (1, 2, 3 ERB}_N)$ factorial design. These 15 conditions were tested with 15 sentences per condition. The conditions were blocked by ERB_N , with gating rates within each ERB_N width presented in a new random order for each subject. The three ERB_N conditions were balanced such that

each appeared in each presentation-order position an equal number of times across subjects (hence, nine subjects). The presentation level was again set to 70 dBA and the speech was scaled to produce an overall SNR of -6 dB.² A non-gated SSN condition was also included, with the noise scaled to match the level of the on portions of the gated noise, resulting in an overall SNR of -9 dB. As in experiment 2, half the subjects heard this condition precede the test conditions and half heard it last. There were therefore 16 conditions and 240 sentences total (HINT lists 1, 3–25). Practice was as in experiment 2, and all calibration, apparatus, and other testing procedures were identical to the previous two experiments.

C. Results and discussion

A 5 (gating rate) $\times$ 3 (spectral bandwidth) repeated-measures ANOVA was conducted on the participant RAU-intelligibility scores. Results demonstrated a significant main effect of gating rate [$F(4,32) = 59.1$, $p < 0.001$, $\eta_p^2 = 0.88$] and spectral bandwidth [$F(2,16) = 62.0$, $p < 0.001$, $\eta_p^2 = 0.88$]. A significant interaction was also observed [$F(8,64) = 5.0$, $p < 0.001$, $\eta_p^2 = 0.38$]. From Fig. 7 it can be observed that performance was best with larger spectral-glance bandwidths (i.e., holes). Performance functions for all three spectral bandwidths demonstrated a minimum at a 4-Hz gating rate, potentially indicating an interaction with the primary syllabic rate of speech. The interaction occurred due to a larger effect of gating rate for larger spectral bands. Indeed, paired Bonferroni t -tests between the extreme gating rates and the performance minima at 4 Hz demonstrated that 1 Hz and 4 Hz were significantly different for 2-ERB_N and 3-ERB_N bandwidths ($p < 0.05$). As the bandwidth decreased to 1 ERB_N, performance was still poorest at 4 Hz, but the effect of gating rate was not significant ($p > 0.05$). No significant difference was observed between 4 Hz and 16 Hz for any bandwidth, as observed by the flatter function at higher rates ($p > 0.05$). Furthermore, in

FIG. 7. Speech intelligibility across different temporal-gating rates. Different symbols represent the different spectral conditions tested. The two dashed lines indicate performance in steady-state noise at -6 dB SNR (top) and -9 dB SNR (bottom). Checkerboard noise was presented at a long-term average level of -6 dB SNR with the on portions of noise equated to -9 dB SNR. Error bars indicate the standard error of the mean.

conditions with smaller spectral holes (e.g., 1-ERB_N conditions or 2-ERB_N conditions at or above 4 Hz), performance largely approximated performance in continuous SSN at the long-term average level of the checkerboard noise. As these conditions used a 50% duty cycle, this reflects a 3-dB gain in performance relative to the level of the on portions of the noise.

The different levels of performance across the three ERB_N conditions may be partially understood in terms of differing amounts of simultaneous masking. The 1-ERB_N speech bands were most susceptible to simultaneous masking, primarily spreading upward from the noise bands immediately lower in frequency. In contrast, the larger bandwidth conditions were less susceptible to this effect because smaller proportions of each speech band fell into the spectral region immediately above each noise band, which is responsible for maximum simultaneous masking.

Overall, these results are largely consistent with those obtained in experiment 2. Performance is better with larger spectro-temporal glimpses. These results also highlight again how spectro-temporal glimpsing involves additional processes to those with temporally gated noise. First, spectro-temporal “checkerboard” noise produces far more masking than temporally gated noise, as evidenced by the 12-dB difference in the level of the masking noise. Second, spectro-temporal glimpsing results in a U-shaped function with the poorest performance at 4–8 Hz, whereas performance in temporally gated noise is an inverted U-shaped function with *maximum* performance obtained around 4–8 Hz. Both types of noise appear to be associated with the primary 4-Hz modulation rate of speech, but with fundamentally different effects on overall intelligibility. These results together indicate that different intelligibility predictions are required for temporally fluctuating versus spectro-temporally fluctuating maskers.

V. GENERAL DISCUSSION

Recognizing speech in noise often requires the listener to process speech based only on partial acoustic information. The distribution in time and frequency of this partial information is, in part, determined by the temporal and spectro-temporal properties of the noise. The current study confirms that the ability of listeners to resolve partial information is dependent on the rate of this distribution across time, i.e., gating rate. Now replicated across many different studies using different speech materials, speech recognition appears to be maximized for temporal gating rates close to 4–8 Hz. The current results further demonstrate that this pattern is maintained for random gating, which may arguably better represent natural environments (although it is still clearly artificial). The effect of random gating becomes much more apparent for spectro-temporal glimpsing where the variability is extended in dimensions of time and frequency. Experiment 2 demonstrated particularly poorer performance with random gating for the middle spectro-temporal glimpse sizes.

Interestingly, spectro-temporal glimpsing is maximized for glimpse sizes that are the largest (i.e., 1-Hz/3-ERB). As glimpse size is reduced, performance approaches that in continuous noise matching the overall power of the gated noise. However, significant masking release is still observed when

compared to performance in continuous noise that matches the level of the on portions of the spectro-temporal gated noise. When the influence of gating rate is tested for different spectral bandwidths (i.e., experiment 3), we again found that performance is best with the widest spectral bandwidth. This is consistent with Howard-Jones and Rosen (1993b) who found that performance was best for two-band compared to four-band checkerboard noise.

Other studies have presented spectro-temporally filtered speech, which shares similarities and clear differences with speech in checkerboard noise. For example, Buss *et al.* (2004) created “checkerboard” speech by spectro-temporally filtering it into 2–16 logarithmic bands gated at 10 or 20 Hz. They found that performance increased with an increasing number of bands (except that performance with two bands was also better). The results from Buss *et al.* (2004) are consistent with Humes and Kidd (2016) for spectrally filtered speech without temporal modulation. The latter authors found that performance was better with a larger number of smaller spectrally sparse speech bands than with a smaller number of wider contiguous bands (i.e., performance improved monotonically from a bandwidth of approximately 8 ERB_N to around 1.5 ERB_N). This was found for filtered speech in quiet and when presented in broadband or notch-filtered noise (i.e., noise filling the spectral gaps of missing speech). These results are consistent with a long lineage of studies by Warren and colleagues demonstrating a super-additivity of sparsely represented speech (e.g., Warren *et al.*, 1995; Warren *et al.*, 1997; Warren and Bashford, 1999; Healy and Warren, 2003; Warren *et al.*, 2005; Bashford *et al.*, 2000). That is, significant speech recognition can be obtained with narrow speech bands if they sufficiently sample the frequency spectrum.

Interestingly, a very different pattern of results emerged in the present study when listening to continuous speech in spectro-temporally filtered noise. Performance was better with fewer, wider noise bands that spared fewer, wider speech regions. It may be that the bands used in the current study (i.e., 5 3-ERB_N to 15 1-ERB_N unmasked speech bands distributed across the spectrum) were sufficient to adequately sample the frequency spectrum and that performance would decline if larger bandwidths (i.e., fewer frequency bands) were employed. However, there was some overlap in the number of bands tested by Buss *et al.* (2004) and in the current study, yet performance trended in opposite directions for spectro-temporal speech versus spectro-temporal noise according to spectral bandwidth. A similar dichotomy for temporally gated speech versus speech in temporally gated noise is observed in the work of Miller and Licklider (1950). For example, in that study performance for gated speech is maintained at a relatively high level for rates above 16 Hz (see Fig. 4 in Miller and Licklider, 1950), whereas, performance in gated noise quickly rolls off to minimal speech intelligibility for noise gating above 16 Hz (see Fig. 8 in Miller and Licklider, 1950).

The results from this and prior studies together indicate that gated and/or filtered speech may not be the best model for real-world listening. In real-world listening, gated and/or filtered speech does not typically occur. But continuous speech often occurs in the presence of fluctuating noise, potentially obliterating various portions of the spectro-temporal signal.

Accordingly, performance is determined not only by temporal and spectral gating properties, but also by processes of simultaneous and non-simultaneous masking that can actually reverse the conclusions. Whereas performance for gated speech is maximized when faster temporal gating and/or narrower spectral bands are employed, performance in gated noise is maximized with slower gating and/or wider spectral bands. This is because the mechanisms of masking negate the advantages associated with better sampling of speech information.

A final observation involves the difference in performance across temporal gating rate for temporally gated noise versus spectro-temporally gated noise. Whereas the former demonstrates peak performance around 4–8 Hz, the latter resulted in a dip in performance at these rates. Although speculative, it is possible that this observation is a result of interactions with the temporal modulation of speech. Speech modulations peak at a syllabic rate around 4 Hz (Greenberg *et al.*, 2003). Furthermore, this predominant speech-modulation rate is correlated across spectral bands (Crouzet and Ainsworth, 2001). A temporal gating rate of noise around 4 Hz may preserve access to these important speech fluctuations that are comodulated across the spectrum. In contrast, spectro-temporal gating results in uncomodulated glimpses of this primary 4-Hz rate across the spectrum. Therefore, it might be advantageous for speech recognition to have access to comodulated speech information across the spectrum, particularly around 4 Hz. Further work is needed to examine this potential interaction between uncomodulated glimpses and perception of speech modulation.

The current results involving modulated noise should not be confused with issues associated with glimpsing speech in noise by listeners having sensorineural hearing impairment and resulting broad frequency tuning. The current results describe the temporal (experiment 1) and spectro-temporal (experiments 2 and 3) characteristics of noise that are most advantageous or disruptive to speech recognition. Accordingly, they indicate what aspects of a modulated masker result in glimpses of speech that convey the best/poorest information to the listener and therefore yield the best/poorest intelligibility. Clearly, the current results show that larger and slower glimpses in background noise convey more effective speech information. The situation is quite different for hearing-impaired listeners attempting to recognize speech in complex noise backgrounds. In this situation, it has been argued that poor frequency resolution limits the ability to locate glimpses of clean speech in natural noises, because all of the broad glimpses available to these listeners contain noise (e.g., Apoux and Healy, 2009). The current results suggest that, in the right noise (one containing speech and noise in different relatively large regions), hearing-impaired listeners should perform quite well, unless their ability to integrate speech information across glimpses is somehow limited (e.g., Healy and Bacon, 2002; Hall *et al.*, 2008a,b).

ACKNOWLEDGMENTS

The authors would like to thank Jordan Vasko and Bobby Gibbs for research assistance. This work was supported, in part, by the National Institutes of Health, National Institute on Deafness and Other Communication Disorders, Grant Nos. R01-DC015465 (D.F.) and R01-DC015521 (E.W.H.).

¹Intelligibility in non-gated SSN was also assessed. The first group of subjects in experiment 1 heard a single list of ten HINT sentences following all other conditions. The level of the SSN matched that of the on portions of the gated noise, producing an overall SNR of -21 dB. To also assess intelligibility in non-gated SSN at the overall SNR of -18 dB employed in experiment 1, a control group of five additional listeners (mean = 24 yr, 23–26 yr) completed testing at The Ohio State University. These listeners had characteristics identical to those in the main groups, and experienced the same practice conditions followed by a single list of ten HINT sentences. Subjects reported being unable to detect the presence of speech at either SNR and reported no keywords, resulting in scores of 0 for both groups.

²The non-gated SSN from each experiment was used to calibrate the head-phone level. In experiments 1 and 3, the level of this signal matched the level of the on portions of the noise. In experiment 2, this signal matched the overall level of the gated noise. Therefore, presentation level was 3 dB higher in experiment 2 than in experiment 3, where SNRs were equal. This difference in overall level is expected to be inconsequential, an assumption that is supported by a comparison of conditions that are identical across experiments 2 and 3—average intelligibility across these three conditions differed by less than 1 percentage-point across experiments.

- ANSI (2004). ANSI S3.21 (R2009), *American National Standard Methods for Manual Pure-Tone Threshold Audiometry* (American National Standards Institute, New York).
- ANSI (2010). ANSI S3.6, *American National Standard Specification for Audiometers* (American National Standards Institute, New York).
- Apoux, F., and Healy, E. W. (2009). "On the number of auditory filter outputs needed to understand speech: Further evidence for auditory channel independence," *Hear. Res.* **255**, 99–108.
- Arnal, L. H., Doelling, K. B., and Poeppel, D. (2014). "Delta-beta coupled oscillations underlie temporal prediction accuracy," *Cereb. Cortex* **25**, 3077–3085.
- Bashford, J. A., Jr., Warren, R. M., and Lenz, P. W. (2000). "Relative contributions of passband and filter skirts to the intelligibility of bandpass speech: Some effects of context and amplitude," *Acoust. Res. Lett. Online* **1**, 31–36.
- 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.
- Buss, E., Hall, J. W., III, and Grose, J. H. (2004). "Spectral integration of synchronous and asynchronous cues to consonant identification," *J. Acoust. Soc. Am.* **115**, 2278–2285.
- Crouzet, O., and Ainsworth, W. A. (2001). "On the various influences of envelope information on the perception of speech in adverse conditions: An analysis of between-channel envelope correlation," in *Workshop on Consistent and Reliable Cues for Sound Analysis*, September 2, Aalborg, Denmark.
- Dirks, D. D., Wilson, R. H., and Bower, D. R. (1969). "Effect of pulsed masking on selected speech materials," *J. Acoust. Soc. Am.* **46**, 898–906.
- Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (2003). "Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with normal hearing," *J. Acoust. Soc. Am.* **113**, 2084–2094.
- Engel, A. K., Fries, P., and Singer, W. (2001). "Dynamic predictions: Oscillations and synchrony in top-down processing," *Nat. Rev. Neurosci.* **2**, 704–716.
- Festen, J. M., and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *J. Acoust. Soc. Am.* **88**, 1725–1736.
- Fogerty, D., Xu, J., and Gibbs, B. E. (2016). "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.
- Füllgrabe, C., Berthommier, F., and Lorenzi, C. (2006). "Masking release for consonant features in temporally fluctuating background noise," *Hear. Res.* **211**, 74–84.
- Ghitza, O., and Greenberg, S. (2009). "On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence," *Phonetica* **66**, 113–126.
- Giraud, A. L., and Poeppel, D. (2012). "Cortical oscillations and speech processing: Emerging computational principles and operations," *Nat. Neurosci.* **15**, 511–517.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- Gnansia, D., Jourdes, V., and Lorenzi, C. (2008). "Effect of masker modulation depth on speech masking release," *Hear. Res.* **239**, 60–68.
- Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). "Temporal properties of spontaneous speech—A syllable-centric perspective," *J. Phon.* **31**, 465–485.
- Gustafsson, H. A., and Arlinger, S. D. (1994). "Masking of speech by amplitude-modulated noise," *J. Acoust. Soc. Am.* **95**, 518–529.
- Hall, J. W., III, Buss, E., and Grose, J. H. (2008a). "The effect of hearing impairment on the identification of speech that is modulated synchronously or asynchronously across frequency," *J. Acoust. Soc. Am.* **123**, 955–962.
- Hall, J. W., III, Buss, E., and Grose, J. H. (2008b). "Spectral integration of speech bands in normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **124**, 1105–1115.
- Healy, E. W., and Bacon, S. P. (2002). "Across-frequency comparison of temporal speech information by listeners with normal and impaired hearing," *J. Speech Lang. Hear. Res.* **45**, 1262–1275.
- Healy, E. W., and Warren, R. M. (2003). "The role of contrasting temporal amplitude patterns in the perception of speech," *J. Acoust. Soc. Am.* **113**, 1676–1688.
- Hickok, G., Farahbod, H., and Saberi, K. (2015). "The rhythm of perception: Entrainment to acoustic rhythms induces subsequent perceptual oscillation," *Psychol. Sci.* **26**, 1006–1013.
- Howard-Jones, P. A., and Rosen, S. (1993a). "The perception of speech in fluctuating noise," *Acoustica* **78**, 258–272.
- Howard-Jones, P. A., and Rosen, S. (1993b). "Unmodulated glimpsing in 'checkerboard' noise," *J. Acoust. Soc. Am.* **93**, 2915–2922.
- Humes, L. E., and Kidd, G. R. (2016). "Speech recognition for multiple bands: Implications for the Speech Intelligibility Index," *J. Acoust. Soc. Am.* **140**, 2019–2026.
- Jones, M. R. (1976). "Time, our lost dimension: Toward a new theory of perception, attention, and memory," *Psychol. Rev.* **83**, 323–355.
- Jones, M. R., Moynihan, H., MacKenzie, N., and Puente, J. (2002). "Temporal aspects of stimulus-driven attending in dynamic arrays," *Psychol. Sci.* **13**, 313–319.
- Lakatos, P., Musacchia, G., O'Connell, M. N., Falchier, A. Y., Javitt, D. C., and Schroeder, C. E. (2013). "The spectrotemporal filter mechanism of auditory selective attention," *Neuron* **77**, 750–761.
- Li, N., and Loizou, P. C. (2007). "Factors influencing glimpsing of speech in noise," *J. Acoust. Soc. Am.* **122**, 1165–1172.
- Miller, G. A., and Licklider, J. C. R. (1950). "The intelligibility of interrupted speech," *J. Acoust. Soc. Am.* **22**, 167–173.
- 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.
- Neuling, T., Rach, S., Wagner, S., Wolters, C. H., and Herrmann, C. S. (2012). "Good vibrations: Oscillatory phase shapes perception," *Neuroimage* **63**, 771–778.
- Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Am.* **95**, 1085–1099.
- Ozmeral, E. J., Buss, E., and Hall, J. W. III (2012). "Asynchronous glimpsing of speech: Spread of masking and task set-size," *J. Acoust. Soc. Am.* **132**, 1152–1164.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech, Lang., Hear. Res.* **28**, 455–462.
- Takahashi, G. A., and Bacon, S. P. (1992). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," *J. Speech Lang. Hear. Res.* **35**, 1410–1421.
- ten Oever, S., Schroeder, C. E., Poeppel, D., van Atteveldt, N., and Zion-Golumbic, E. (2014). "Rhythmicity and cross-modal temporal cues facilitate detection," *Neuropsychologia* **63**, 43–50.
- Ward, L. M. (2003). "Synchronous neural oscillations and cognitive processes," *Trends Cogn. Sci.* **7**, 553–559.
- Warren, R. M., and Bashford, J. A., Jr. (1999). "Intelligibility of 1/3-octave-speech: Greater contribution of frequencies outside than inside the nominal passband," *J. Acoust. Soc. Am.* **106**, L47–L52.

- Warren, R. M., Bashford, J. A., Jr., and Lenz, P. W. (2005). "Intelligibilities of 1-octave rectangular bands spanning the speech spectrum when heard separately and paired," *J. Acoust. Soc. Am.* **118**, 3261–3266.
- Warren, R. M., Hainsworth, K. R., Brubaker, B. S., Bashford, J. A., Jr., and Healy, E. W. (1997). "Spectral restoration of speech: Intelligibility is increased by inserting noise in spectral gaps," *Percept. Psychophys.* **59**, 275–283.
- Warren, R. M., Riener, K. R., Bashford, J. A., Jr., and Brubaker, B. S. (1995). "Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits," *Percept. Psychophys.* **57**, 175–182.
- Weisz, N., and Obleser, J. (2014). "Synchronisation signatures in the listening brain: A perspective from non-invasive neuroelectrophysiology," *Hear. Res.* **307**, 16–28.
- Wilson, R. H., and Carhart, R. (1969). "Influence of pulsed masking on the threshold for spondees," *J. Acoust. Soc. Am.* **46**, 998–1010.