Perception of interrupted speech: Effects of dual-rate gating on the intelligibility of words and sentences^{a)}

Valeriy Shafiro,^{b)} Stanley Sheft, and Robert Risley

Department of Communication Disorders and Sciences, Rush University Medical Center, 600 South Paulina Street, 1012 AAC, Chicago, Illinois 60612

(Received 26 October 2010; revised 6 August 2011; accepted 8 August 2011)

Perception of interrupted speech and the influence of speech materials and memory load were investigated using one or two concurrent square-wave gating functions. Sentences (Experiment 1) and random one-, three-, and five-word sequences (Experiment 2) were interrupted using either a primary gating rate alone (0.5–24 Hz) or a combined primary and faster secondary rate. The secondary rate interrupted only speech left intact after primary gating, reducing the original speech to 25%. In both experiments, intelligibility increased with primary rate, but varied with memory load and speech material (highest for sentences, lowest for five-word sequences). With dual-rate gating of sentences, intelligibility with fast secondary rates was superior to that with single rates and a 25% duty cycle, approaching that of single rates with a 50% duty cycle for some low and high rates. For dual-rate gating of words, the positive effect of fast secondary gating was smaller than for sentences, and the advantage of sentences over word-sequences was not obtained in many dual-rate conditions. These findings suggest that integration of interrupted speech fragments after gating depends on the duration of the gated speech interval and that sufficiently robust acoustic-phonetic word cues are needed to access higher-level contextual sentence information.

© 2011 Acoustical Society of America. [DOI: 10.1121/1.3631629]

PACS number(s): 43.66.Mk, 43.71.Es [EB]

Pages: 2076–2087

I. INTRODUCTION

Aural communication generally involves perceptual processing of acoustically incomplete speech signals as they are masked by other sounds or distorted in signal transmission. Yet, even when parts of the speech signal are physically missing or completely obscured by noise, speech intelligibility can often remain high (Miller and Licklider, 1950; Warren, 1970; Bashford and Warren, 1987; Jenkins *et al.*, 1983; Strange *et al.*, 1983). In an early seminal study, Miller and Licklider (1950) showed that when portions of speech were periodically removed by gating or masked by modulated noise, high speech intelligibility could be maintained with as little as 25–50% of the original signal as long as the interruption rate was sufficiently fast. Intelligibility was also shown to vary with the interruption duty cycle, which reflects the relative duration of speech on and off times within each interruption cycle. Reducing the duty cycle impaired intelligibility by shortening the speech-on time and lengthening the temporal gaps between the remaining speech fragments. Numerous later investigations have replicated Miller and Licklider's main findings on the relationship between gating rate, duty cycle, and intelligibility (Powers and Speaks, 1973; Huggins, 1975; Powers and Wilcox, 1977; Nelson *et al.*, 2003; Nelson and Jin, 2004; Buss *et al.*, 2009; Jin and Nelson, 2010).

Findings from interrupted speech studies are commonly interpreted by invoking a glimpsing or “multiple looks” process whereby listeners detect spectro-temporal fragments of the original signal and fill in the missing information using memory templates (Miller and Licklider, 1950; Warren *et al.*, 1984; Bashford *et al.* 1988; Howard-Jones and Rosen, 1993; Moore, 2003; Cooke, 2006). The initial “glimpses” or independent “looks” are obtained at the level of auditory periphery, based on a favorable signal-to noise ratio (SNR), and subsequently “intelligently” integrated into higher order perceptual categories at more central processing levels (Moore, 2003). In this view, perception of interrupted speech is based on (1) the detection of preserved speech fragments of the original signal, and (2) the integration of these fragments into higher order perceptual categories (Cooke, 2006). The glimpsing account provides a good fit for many interrupted speech findings, and accounts especially well for the greater intelligibility with faster rates for speech interrupted either by gating or modulated noise maskers (Miller and Licklider, 1950; Powers and Wilcox, 1977; Nelson and Jin, 2004; Jin and Nelson, 2010). Higher intelligibility is also obtained with increasing speech duty cycle, and thus the total proportion of speech retained (Miller and Licklider, 1950). However, previous work also indicates a complex and nonlinear relationship between interruption rate and the proportion of the original speech preserved (Miller and Licklider, 1950; Wang and Humes, 2010).

In a recent study of gated speech, Wang and Humes (2010) examined separate contributions of three interruption parameters to the intelligibility of monosyllabic words. These were (i) interruption rate, (ii) speech-on duration for each interruption interval, and (iii) the total proportion of the original

^{a)}Portions of the data were presented at the 159th Meeting of the Acoustical Society of America.

^{b)}Author to whom correspondence should be addressed. Electronic mail: Valeriy_Shafiro@rush.edu

speech preserved. Among these, the researchers found that the main determinant of speech intelligibility was the proportion of the total speech present after interruption, with a greater intelligibility increase between 25% and 50% than between 50% and 75% proportion of the original speech. The influence of the other two parameters on intelligibility, however, varied with the proportion of speech present. Increasing interruption rate had a positive effect for the lowest (i.e., 25%) speech proportion but a small negative effect for the highest (i.e., 75%). Increasing speech-on time of individual interruptions had a negative effect for the 25% speech proportion and generally no effect for 50% and 75% speech proportions. However, because the last two parameters were not manipulated independently, their individual contributions were less clear.

In addition to the three interruption-specific parameters, Wang and Humes also reported effects of talker gender, stimulus presentation level, and the linguistic difficulty of the stimuli. Interrupted words spoken by a female talker were more intelligible than those spoken by a male talker, interpreted by the authors to result from a greater number of cycles of the fundamental frequency within each interruption interval. Words presented at 85 dB sound pressure level (SPL) were less intelligible than words presented at 65 dB SPL, a finding consistent with other reports of decreased speech intelligibility at presentation levels above 80 dB SPL. More linguistically difficult words, measured by their sparser neighborhood density and lower word frequency, were less intelligible than less linguistically difficult words. According to Wang and Humes, while effects of talker gender and level were likely the result of physical characteristics of the interrupted signal, the effect of stimulus materials reflected the influence of central factors and demonstrated that intelligibility of interrupted speech involves a complex relationship between bottom-up and top-down processes.

Further evidence of a complex relationship between interruption parameters and speech intelligibility comes from a study by Kwon and Turner (2001) who investigated whether the intelligibility of speech interrupted by modulated maskers could be affected by modulation detection or discrimination interference (MDI), where the difficulty in processing modulation in one spectral region decreases with spectral separation of a competing modulation (Yost and Sheft, 1989, 1994). Contrary to expectations based on MDI research with nonspeech tones and noises, Kwon and Turner found that modulated masking bands spectrally remote from filtered speech bands produced greater masking than similar masking bands that were proximal. In Experiment 4 of their study, the intelligibility of highpass filtered speech deprived of typical “syllabic” envelope fluctuations decreased when low-frequency modulated masking noise was present. The researchers concluded that segregating modulated maskers from spectrally sparse speech in a remote frequency band was more difficult than segregating unmodulated maskers, and that listeners were more likely to treat masker fluctuations as part of the speech-signal envelope. Thus, the intelligibility decrement was more akin to auditory induction (Warren 1970; Shafiro and Raphael, 2007) influenced by top-down processes, rather than energetic masking *per se* (cf., Pollack, 1975; Watson, 2005).

Gustafsson and Arlinger (1994) studied the effect of modulated maskers on speech perception, comparing sinusoidal to complex masker envelope patterns more typical of speech maskers. The irregular masking patterns used in their study represented the sum of four inharmonic sinusoidal modulators (i.e., 2.1, 4.9, 10.2, 19.9 Hz) with equal amplitude and random phase. The range of variation in instantaneous noise level produced by complex maskers was similar to that obtained with each of the masker components presented separately, thus preserving the same total proportion of the original speech. Although masker modulation always benefited speech perception, intelligibility was poorer in the presence of complex rather than sinusoidal masker modulation. The researchers interpreted the lower intelligibility scores obtained in the presence of complex maskers as reflecting “... the lower probability of noise intervals of sufficiently low-level and long duration to allow the listener to pick up meaningful fragments of the speech signal, the irregularity [of masker envelopes] having no likely effect.” However, no information was provided on what may constitute “meaningful fragments” or the processes underlying their temporal integration for complex interruption patterns.

With a strong emphasis on the detection of signal glimpses or multiple looks, the glimpsing account provides little *a priori* specification of the underlying integration processes or associated temporal constraints that could affect the intelligibility of complex interruption patterns. Existing evidence is consistent with the notion that perception of individual glimpses into higher order templates that represent linguistic categories involves an intelligent integration process (Moore, 2003; Wang and Humes, 2010). However, the relationship between underlying templates and specific interruption parameters is not fully understood. Furthermore, because processing segmental and suprasegmental aspects of speech takes place on different time scales (Rosen, 1992; Howard-Jones and Rosen, 1993; Greenberg, 1996; Mattys, 1997; Poeppel *et al.*, 2008), it could be expected that qualitatively different perceptual processes may be involved at different rates of interruption as the duration of speech on and off times vary.

Specifically, speech perception at slow gating rates of about 0.5 – 1 Hz, where the duration of the remaining and discarded speech approaches or exceeds that of whole words, could be expected to rely to a greater extent on more central contextual and semantic processing, with the perception of retained words requiring relatively little effort. On the other hand, perception of speech interrupted at faster rates (e.g., above 8 Hz), which sample most phonemic segments within a word at least once, involves greater reliance on temporal smoothing over small silent gaps in the signal. However, for speech gated at a single fast rate, the contributions to intelligibility of perceptual processing on a shorter time scale would overlap with those from a longer time scale, with concurrent involvement of temporal smoothing *and* context-based processes. In its general form, the glimpsing account does not distinguish between perceptual processes that take place on different perceptual time scales. Thus, the first goal of the present study was to examine perceptual processing of interrupted speech across a range of interruption rates that

could be expected to involve qualitatively different perceptual processes.

To separately assess the contributions to intelligibility of slower perceptual processes at low rates from the perceptual processes at higher rates, the present study compared intelligibility of speech interrupted with a single rate to speech interrupted with two concurrent rates. In such dual-rate gating, the second interruption rate was always applied to speech previously gated at a slower primary single rate with a 50% duty cycle. Therefore, the second-rate gating further reduced the proportion of speech within each single-rate speech-on interval by one half. Thus, the total proportion of the original speech in dual-rate gated signal was the same as in a single-rate gated signal with a 25% duty cycle. This manipulation made it possible, for the same total proportion of speech, to examine the effect of secondary interruption rate across a range of speech-on durations of the primary rates. The perceptual processes involved in temporal integration at faster secondary interruption rates could then be assessed independently from those at slower primary rates to indicate possible effects of slower primary rates on the intelligibility changes produced by secondary rates.

Moreover to examine the extent to which rate effects might be constrained by the type of speech materials, the intelligibility of both sentences and words was examined. Previous studies of interrupted speech utilized various types of speech materials, with some using words (Miller Licklider, 1950; Wang and Humes, 2010) and others sentences (Nelson and Jin, 2004; Gustafsson and Arlinger, 1994). While the findings from studies with interrupted words and sentences are generally consistent in terms of the basic effects of key interruption parameters such as rate and duty cycle, differences in results can be observed in the overall performance and shape of the rate-intelligibility function (e.g., steepness of slope, presence of a local minimum). However, intelligibility of interrupted words and sentences has rarely been assessed in a single study with common interruption methods (cf. Dirks *et al.*, 1969). Considering well known differences in the perceptual processing of words and sentences (e.g., coarticulation and context predictability based on syntactic and semantic effects), the second goal of the present study was to examine the effects of dual-rate gating for different types of speech materials.

Finally, cognitive factors such as working memory have been also found to affect the perception of degraded speech [see Wingfield and Tun (2007) for a review]. However, the effect of working memory load on the intelligibility of interrupted speech has not been investigated. It is unknown whether increasing working memory load would have a similar effect on interrupted speech regardless of interruption rate, or, with respect to the present study, how memory load could influence perceptual processing of dual-rate gated speech. If increasing the memory load leads mainly to greater forgetting of encoded words, it could be expected that negative effects of a higher memory load would be observed for all rates. On the other hand, if increasing memory load leads mainly to reducing the perceptual processing resources needed for encoding of the degraded signal, it could be expected that effects of memory load would differ

across rates, resulting in an interaction of rate and memory load. Effects would be minimal for low rates, especially for single rates with a 50% duty cycle where speech-on times preserve nearly complete undistorted words, but would be observable at higher gating rates which distort speech at (sub)segmental levels, causing encoding difficulty due to input distortion. Thus, the third goal of the present work was to examine the effect of memory load, expressed as the number of isolated words in a stimulus sequence, on the intelligibility of gated speech.

The presents study investigated the relationships between perceptual processes underlying intelligibility of words and sentences interrupted at different rates using single- and dual-rate gating. As in previous research (Wang and Humes, 2010; Miller and Licklider, 1950; Nelson and Jin, 2004), gating was chosen as the method of interruption due to involvement of fewer variables compared with interruption methods which utilize modulated-noise interruptions.¹ It was expected that for dual-rate gated speech (which has the same proportion of the original speech as speech gated at a single rate with a 25% duty cycle) intelligibility would gradually increase with secondary rate, potentially reaching that obtained with a single rate and a 50% duty cycle. If the duration of interrupted speech fragments obtained with primary rates does not interact with secondary rates, the magnitude of secondary rate effects would not be expected to differ across primary interruption rates. These expectations are consistent with most current glimpsing models that specify no *a priori* constraints on the duration of speech intervals subjected to interruption (although such constraints are not explicitly ruled out either). However, if a nonlinear relationship between secondary and primary interruption rates is obtained, such finding would be generally consistent with previous results obtained with complex interruption patterns (Gustafsson and Arlinger, 1994; Kwon and Turner, 2001) and could indicate how the duration of speech-on intervals across primary rates modulates the effects of the secondary rate. In addition, across the two kinds of speech materials, sentences (Experiment 1) were expected to be more intelligible than isolated words (Experiment 2), while a higher working memory load, also manipulated in Experiment 2, was expected to result in reduced intelligibility of single- and dual-rate gated speech and possibly interact with rate.

II. EXPERIMENT 1

A. Method

1. Stimuli and procedure

Speech stimuli were HINT sentences produced by a male talker (Nilsson *et al.* 1994) which were gated either with one rate or two concurrent rates as illustrated in Fig. 1. HINT sentences were chosen for their simple syntactic structure and semantic cues which increase contextual predictability of individual words, providing a contrast to the isolated unrelated words of Experiment 2. Sentences were gated at six single primary rates (0.5–10 Hz) using 50% and 25% duty cycles (see Table I). These rates were selected to span the range of speech-on intervals from word to

FIG. 1. Illustrations of gated speech. The top panel shows the continuous waveform of a sentence; the two middle panels show this sentence gated at a rate of 1 or 10 Hz (50% duty cycle). The bottom panel shows the sentence interrupted at the two rates concurrently. When concurrent, the faster gating rate of 10 Hz affects only the speech remaining after gating at the slower primary rate of 1 Hz, leaving 25% of the original speech signal.

(sub)phonemic durations which could be expected to involve perceptual processing on different time scales. The six primary rates (with a 50% duty cycle) were subsequently gated using secondary rates which also had a 50% duty cycle, reducing the total proportion of the original speech to 25%. In addition, there were two secondary rates for each of the four lower primary rates (0.5–4 Hz) in order to obtain a more comprehensive set of data to more fully indicate the effect of secondary rate on intelligibility.

TABLE I. Primary and associated secondary gating rates used in Experiments 1 and 2. Duty cycles used with primary rates are shown in parentheses next to each rate; a 50% duty cycle was used with all secondary rates. Rates and duty cycles used in both experiments are indicated in bold; rates and duty cycles used only in Experiment 1 are italicized.

Gating rates (Hz)	
Primary	Secondary
0.5 (50%, 25%)	8, 24
1 (50%, 25%)	8, 24
2 (50%, 25%)	8, 24
4 (50%, 25%)	16, 24
8 (50%, 25%)	24
10 (50%, 25%)	24
<i>16 (50%)</i>	<i>None</i>
<i>20 (50%)</i>	<i>None</i>
<i>24 (50%)</i>	<i>None</i>

The primary and secondary gating functions were always in phase, with both square-wave functions starting at 0 degrees. Before gating, a silent interval of random duration was added to the waveform corresponding to each sentence list with interval duration selected from a uniform distribution extending from zero to one second. This manipulation effectively randomized the relationship between the gating functions and the speech signals without altering relationship between primary and secondary gating functions.² In addition, there were three primary rates (16, 20, 24 Hz) gated using only a 50% duty cycle. These rates were included to confirm high intelligibility for these rates when presented alone.

Order of the 25 gating conditions was randomized across sentence lists and subjects. Before running each condition, a digitized version of a randomly selected sentence list was multiplied by either one or two square-wave modulators (i.e., a dc-shifted square wave with peak and trough values of 1.0 and 0.0, respectively, with no additional smoothing), to generate the appropriate gating characteristics. The same gating procedure was applied to stimuli in each gating condition. Under control of a personal computer, modulated stimuli were output at a 22.1-kHz sampling rate through an Edirol UA25 24-bit soundcard with on-board anti-alias filtering. Sentences were presented diotically in quiet through Sennheiser 250 II headphones at 70 dB SPL.

Seated in a double-walled soundproof booth, listeners were asked to repeat what they heard after the presentation

of every sentence. One HINT list of 10 sentences (about 50 words) was used for every gating condition. Correctly repeated keywords for every sentence in the list were summed and divided by the total number of keywords in that list, producing an intelligibility score for a given condition. Condition order was randomized for each subject with each test sentence list preceded by an unscored practice list of five IEEE sentences for that specific gating condition. To familiarize subjects with the testing procedure, before data collection began, subjects practiced with two different sets, each consisting of five IEEE sentences and a HINT list gated with randomly chosen primary and secondary rates.

2. Subjects

Twenty normal-hearing adults (14 females) with hearing thresholds at or below 20 dB hearing level (HL) between 250 Hz – 8 kHz took part in the experiment. Their mean age was 25 yr (SD=6.7). All subjects were native speakers of American English. Experimental protocol was approved by the Institutional Review Board of the Rush University Medical Center.

B. Results

The mean intelligibility scores of the 20 listeners in all gating conditions of Experiment 1 are presented in Table II. Consistent with past work (e.g., [Miller and Licklider, 1950](#), [Powers and Speaks, 1970](#); [Jin and Nelson, 2010](#)), performance with a single gating rate and a 50% duty cycle (rows in which the secondary rate is labeled “none (50%)” in Table II) improved with rate, reaching an asymptote at 8 Hz. Performance for all primary rates was considerably lower with reduction in the duty cycle to 25% (rows in which the secondary rate is labeled “none (25%)” in Table II), with the greatest drop in accuracy at 2 Hz. The remaining data points in Table II represent performance obtained when secondary gating was applied following primary gating. Application of the secondary rates resulted in increased intelligibility across all primary rates over the 25% duty-cycle single-rate levels, with a tendency for higher intelligibility with higher secondary rate. However, with dual-rate gating, the improvement in intelligibility over the single-rate and 25% duty-cycle performance level varied considerably across primary rates. For some primary rates such as 0.5 and 8 Hz, application of fast secondary rates led to performance approaching that of a single rate with a 50% duty cycle, although the 6.07 point difference in intelligibility between the single and dual rate (with a 24 Hz secondary gating) remained significant at $p < 0.05$ for 0.5 Hz. Nevertheless, it appears likely that for the lowest single gating rate of 0.5 Hz, the 1-s-long speech-on intervals contained more than one word with little, if any, distortion of speech cues (i.e., the average word duration in the sentence material was about 330 ms). With this low primary rate, application of a faster secondary rate such as 24 Hz seemed to have an effect similar to the application of a single rate to speech that had not been already gated, with the small but significant drop in intelligibility for dual-rate gating likely due to some primary rate interruptions at 0.5 Hz randomly producing fragmented words. Thus, for low

TABLE II. Mean identification performance with standard deviations in parentheses of 20 listeners in each gating condition of Experiment 1. When secondary gating rate is labeled “none,” intelligibility corresponds to that of the corresponding primary rate alone with duty cycle shown in parenthesis.

Primary Rate (Hz)	Secondary Rate (Hz)	Percent correct (s.d.)	
0.5	none (50%)	56.6	(7.1)
	none (25%)	27.3	(6.9)
	8.0	42.4	(4.1)
	24.0	50.6	(6.1)
1.0	none (50%)	59.4	(11.3)
	none (25%)	20.4	(8.5)
	8.0	34.0	(8.8)
	24.0	41.8	(8.3)
2.0	none (50%)	71.2	(11.4)
	none (25%)	13.5	(9.3)
	8.0	26.8	(9.7)
	24.0	33.8	(12.9)
4.0	none (50%)	86.3	(7.2)
	none (25%)	35.4	(11.1)
	16.0	50.1	(11.4)
	24.0	50.8	(14.1)
8.0	none (50%)	97.2	(4.0)
	none (25%)	65.6	(13.7)
	24.0	93.8	(5.6)
10.0	none (50%)	98.1	(2.1)
	none (25%)	78.1	(10.0)
	24.0	81.1	(13.9)
16.0	none (50%)	99.0	(2.7)
20.0	none (50%)	99.1	(1.3)
24.0	none (50%)	99.9	(0.4)

primary rates which preserved long intervals of undistorted speech, the rise in intelligibility with increase in secondary rate might mirror that observed with primary-rate gating on its own. Even though half of the original signal was missing in single-rate gated stimuli with a 50% duty cycle, for low gating rates, the perceptual processes underlying temporal integration of small glimpses, following application of a fast secondary rate, appeared to be similar to those of undistorted speech. As the primary rate increased, the effectiveness of a fast 24-Hz secondary rate decreased, especially around 2 – 4 Hz, where a significant local minimum was observed ($p < 0.05$). A further increase in the primary rate to 8 Hz, which on its own produced ceiling-level performance, again became conducive to the recovery of intelligibility with a fast secondary rate, before it declined once more with a primary rate of 10 Hz.

These relationships are displayed more clearly in Fig. 2 which shows a subset of the data from Table II as a function of the primary gating rate. For each primary rate, sentences were perceived most accurately when gated with a single rate and a 50% duty cycle (P50%). Sentences gated with a single rate and a 25% duty cycle (P25%) were overall least intelligible, with intelligibility decreasing by 20 – 58 percentage points below the corresponding P50% performance. For dual-rate gated sentences with the highest secondary rate of 24 Hz (S24Hz), intelligibility remained between the P50% and P25% single-rate functions. Both the P25% and the S24Hz rate-intelligibility functions contained significant

FIG. 2. A subset of sentence intelligibility data from Table II (error bars show ± 1 standard deviation). The top and the bottom curves depict accuracy functions for primary rates alone with a 50% (P50%) and 25% (P25%) duty cycle, respectively. The middle curve indicates intelligibility at each primary rate when the secondary rate was 24 Hz (S24Hz).

local minima at 2 Hz, as intelligibility at 2 Hz was significantly ($p < 0.05$) lower than that at 1 or 4 Hz. This data subset was submitted to a 3×6 repeated-measures ANOVA on the rationalized arcsine-transformed intelligibility scores (Studebaker, 1985) to determine the effects of primary gating rate and gating type (i.e., P50%, P25% and S24Hz). The ANOVA revealed a significant main effect of rate [$F(5, 95) = 444.75, p < 0.0001$] and also a main effect of gating type [$F(2, 938) = 377.45, p < 0.0001$]. Pairwise comparisons with a Bonferroni correction confirmed significant differences ($p < 0.05$) among each of the three levels of gating type.

The ANOVA also revealed a significant interaction between gating rate and type [$F(10, 190) = 15.65, p < .0001$], driven primarily by the local minimum around 2 Hz, which was present in both the S24Hz and P25% rate-intelligibility functions, but absent in the P50% function. Intelligibility differences in the S24Hz and P25% functions at high and low primary gating rates might have also contributed to the interaction.

C. Discussion

The results of Experiment 1 demonstrated that the effect of a faster secondary rate on the intelligibility of sentences previously gated with a slower primary rate varies considerably across primary rates. For most primary rates, intelligibility of dual-rate gated speech exceeded that obtained with a single rate and a 25% duty cycle. With fast secondary gating of primary rates of 0.5 and 8 Hz, intelligibility approached that obtained with a single rate and a 50% duty cycle. Conversely for primary rates of 1–4 Hz, the negative effect of fast secondary gating was considerably more pronounced. This indicates that speech intelligibility was not determined exclusively by either the total amount of the original speech retained after secondary gating or by the frequency of sampling of the speech cues. Among these two factors, the quantity of the speech cues retained might have a greater influence on performance, as illustrated by the

similarity in the shapes of the S24Hz and P25% intelligibility functions from conditions which shared the same total amount of the original speech.

A tentative interpretation of the dual-rate intelligibility variation across primary rates is that glimpsing of underlying speech cues may work best for those primary gating rates which, on their own, provide sufficient speech cues for accurate identification of whole words. Such is the case for both low and high single gating rates which, while differing in absolute intelligibility, provide sufficiently robust low-level acoustic-phonetic details to reduce uncertainty regarding individual words. For speech gated at low rates, the on-times were long enough to include full undistorted words after gating, and for high-rate gated speech, on-times were so frequent and proximal that each word was sampled multiple times. On the other hand, for the primary gating rate of 2 Hz, speech-on times generally do not contain full words, and are not frequent enough to sample each word more than once. In the cases of gating at a single rate with a 50% duty cycle, the low-level acoustic-phonetic cues may be sufficiently robust to enable access to high-level contextual information in the sentences, resulting in a monotonic rate-intelligibility function. However, as the proportion of the original speech declines with duty cycle or dual-rate gating, the low-level cues may be insufficient for accessing higher level information so that performance with a 2-Hz primary rate declines more than at lower and higher primary rates.

A single exception to this pattern was the gating condition in which the primary rate of 10 Hz, highly accurate on its own, was combined with a secondary rate of 24 Hz, and produced no significant improvement in performance relative to the P25% level. The lack of improvement most likely resulted from the inharmonic relationship between the primary rate of 10 Hz and a secondary rate of 24 Hz. Except for a 10-Hz primary rate, all secondary rates in Experiment 1 were integer multiples of the primary rate. Because the starting phases of the primary and secondary rates were the same, for all rates used other than 10 Hz, the secondary rates would regularly sample speech intervals remaining after primary gating. The inharmonic relationship between a 10-Hz primary and a 24-Hz secondary rate produced a complex pattern of fragmentation with individual speech fragments at times being too brief to be perceptually useful.

Overall, the pattern of intelligibility found in Experiment 1 indicates complex nonlinear relationships based on duty cycle (50% vs 25%) and number of gating functions (one or two). However, this pattern of results may also be affected by the speech materials used. High predictability sentences such as HINT contain strong high-level contextual cues (e.g., semantic, syntactic, and sentence-specific prosodic cues) that can affect intelligibility, particularly when low-level speech cues become degraded and sparse (Bilger *et al.*, 1984; Grant and Seitz, 2000; Boothroyd, 2010; Spitzer *et al.*, 2009). Although intelligibility of gated words and sentences has rarely been tested in a single study (cf. Dirks *et al.*, 1969), at common interruption rates, intelligibility is generally higher for sentences than words. Conceivably, the local minima observed around 2 Hz in the P25% and S24Hz functions for sentences, but not in the P50% conditions,

could reflect differences in accessing higher level sentence cues in the former two conditions. In the absence of contextual sentence cues, listener performance with semantically unrelated words can indicate the extent to which the effectiveness of fast secondary gating is a function of word identification based on separate word cues alone. If the relationship between single- and dual-rate gating functions interact in similarly complex and nonlinear ways for words as they do for sentences, factors other than sentence-level cues would be expected to account for that result. Alternatively, finding no interaction between single- and dual-rate gating functions for words would indicate that sentence level cues were likely involved in an interaction between single- and dual-rate results in Experiment 1. These considerations, and the potential effects of working memory load discussed in the Introduction, were investigated in Experiment 2.

III. EXPERIMENT 2

Experiment 2 examined intelligibility of monosyllabic words gated at one or two concurrent rates. The word stimuli were presented in three different formats: single words, sequences of three randomly selected words, and sequences of five randomly selected words. The single-word condition was included for continuity with previous gating studies with isolated words, which have not explicitly manipulated the number of words in a trial. The additional three- and five-word conditions were included to manipulate word memory load. The goals of Experiment 2 were (1) to determine the rate-intelligibility functions for unrelated words using different types of gating (i.e., P50%, P25%, S24 Hz), and (2) to assess the effect of memory load on performance across gating-type and rate conditions for words.

A. Method

1. Stimuli and procedure

Word stimuli were CNC (consonant-vowel-consonant) words (Peterson and Lehiste, 1962) spoken individually by a male speaker of American English (TigerSpeech Technology). Randomly selected words were arranged into one of the three word-length conditions: one-word (1W), three-word (3W), and five-word (5W) long sequences. Individual words within each three- and five-word sequence were separated on average by 80-ms intervals. Specific interval durations were randomly selected from a uniform distribution extending from 72 to 88 ms. A 3-, 5-, and 7-s interstimulus interval between word sequences was used for listener responses in the 1W, 3W, and 5W conditions, respectively. Stimuli were gated with one or two rates following the methods of Experiment 1. There were five single primary rates (0.5–8 Hz) gated with a 50 or 25% duty cycle when presented alone, or with subsequent gating by a secondary rate of 24 Hz (see Table I). In addition, a condition with a single primary rate of 10 Hz and a 50% duty cycle was included to determine the presence of a possible performance plateau with multiword sequences.

Stimulus presentation and scoring procedures were the same as in Experiment 1. Performance in each 1W gating

condition was based on a list of 50 randomly selected words. A 25-word practice list preceded every 1W gating condition. In each 3W gating condition, subjects were presented with 17 strings of three randomly selected words and were asked to repeat the words only after presentation of each sequence. A practice list of eight three-word strings preceded each gated 3W condition. Performance in each 5W condition was based on ten five-word strings, following practice with five five-word strings. Gating conditions and word-sequence length were randomized for each listener. Before data collection began for each word-length set, subjects were presented with a practice list gated with a randomly selected primary and secondary rate. In addition, word stimuli were presented without gating to establish baseline performance.

2. Subjects

Seventeen subject-listeners (14 females) who had participated in Experiment 1 were tested. Their mean age was 23 yr ($SD = 1.9$ yr).

B. Results

Baseline results for stimuli presented without gating indicated that words were highly intelligible in each of the three word-length conditions with intelligibility scores of 100 ($SD = 0.8$), 99 ($SD = 1.2$), and 88 ($SD = 9.4$) percent correct for conditions 1W, 3W, and 5W, respectively. The inverse relationship between intelligibility and word-sequence length was also obtained in the gating conditions. Figure 3 depicts the mean rate-intelligibility functions with each word-sequence length condition in a separate panel. In each panel, performance is displayed as a function of the primary gating rate with gating type as the parameter. As in Experiment 1, intelligibility was highest when stimuli were gated at a single rate with a 50% duty cycle, reaching maximum at 8 Hz and becoming asymptotic at 10 Hz for the 1W and 5W conditions. The lowest intelligibility was obtained in all three word sequence length conditions utilizing a 25% duty cycle of the primary gating rate. However, unlike results from Experiment 1, function shapes in Experiment 2 appeared more similar across the three gating types.

A three-way ($3 \times 3 \times 5$) repeated-measures ANOVA on the arcsine-transformed intelligibility scores was conducted to investigate the effects of word-sequence length, gating type, and gating rate. The ANOVA revealed significant main effects of gating type [$F(2, 32) = 596.58, p < 0.0001$] and rate [$F(4, 64) = 391.92, p < 0.0001$], indicating that both factors had a similar overall effect on words as on the sentences of Experiment 1. In addition, a significant main effect of the number of words was also observed [$F(2, 32) = 53.63, p < 0.0001$]. The inverse relationship between word-sequence length and intelligibility is seen more clearly in Fig. 4 which, including the sentence data from Experiment 1, replots results for each gating type separately with stimulus type as the parameter. Pairwise comparisons demonstrated that single words were identified more accurately than three-word stimuli, which, in turn, were identified more accurately than 5W stimuli. The interaction between word-sequence length and gating type was not significant

FIG. 3. Average word identification accuracy (error bars show ± 1 standard deviation) for the one (1W, top), three (3W, middle), and five (5W, bottom) word-sequence-length conditions of Experiment 2. The top and bottom curves in each panel show accuracy functions for primary rates alone with a 50% and a 25% duty cycle (P50% and P25%), respectively. The middle curve indicates intelligibility at each primary rate when the secondary rate was 24 Hz (S24Hz).

($p > 0.55$), confirming the observation that reduction of duty cycle and introduction of a second rate had a similar effect across primary interruption rates (Fig. 3).

Two significant interactions were observed. First, there was an interaction between word-sequence length and rate [$F(8,128) = 12.67$, $p < 0.0001$], indicating that faster gating (i.e., more frequent sampling of speech) was required to achieve high speech intelligibility as word-sequence length increased. As can be seen in Fig. 4, for all gating types, the higher rate ends of the rate-intelligibility functions tended to flatten as the number of words in the stimulus intervals increased, while they overlapped at the lowest rate. The second significant interaction was between gating type and rate [$F(8,128) = 9.85$, $p < 0.0001$] which was likely due to performance differences across primary rates in the S24Hz function, while the P50% and P25% conditions showed more consistent differences across rate (Fig. 3). However, across primary rates, word intelligibility functions for different gating types were more similar to each other than was the case in the corresponding results obtained with gated sentences.

The parametric effects displayed in Fig. 4 were further examined with a 4×5 repeated-measures ANOVA on stimulus type (sentences, one-, three-, five- word sequences) and primary gating rate (0.5–8 Hz) using data from both experi-

FIG. 4. Intelligibility of gated sentences and one, three, and five word-length sequences (1W, 3W, and 5W) for three types of gating with either a single rate and 50% duty cycle (P50%), single rate and 25% duty cycle (P25%), or dual-rate gating with both duty cycles 50% (S24Hz) in the top, middle, and bottom panels, respectively.

ments. A separate ANOVAs was conducted for each of the three gating types (i.e., P50%, P25%, S24Hz). There was a main effect of stimulus type for each of the three gating types [P50%: $F(3,48) = 124.00$, $p < 0.0001$; P25%: $F(3,48) = 39.00$, $p < 0.0001$; S24Hz: $F(3,48) = 45.49$, $p < 0.0001$]. Pairwise comparisons indicated that sentences were significantly more intelligible than words ($p < 0.01$), except for the single-rate 25% duty-cycle condition in which there was no significant performance difference between sentences and one-word stimuli. Pairwise comparisons further confirmed significant differences among the three word-sequence length conditions, with single-word stimuli having the highest intelligibility, followed by three-word and five-word stimuli. However, sentence superiority over words varied across gating conditions. It was most apparent with the P50% single-rate stimuli (Fig. 4, top panel), but decreased for the P25% and S24Hz stimuli. For some intermediate gating rates, sentence intelligibility even declined below single word levels (Fig. 4, middle and bottom panels) and was comparable to that of three- and five-word stimuli. A main effect of rate was also found for all three gating types [P50%: $F(4,64) = 400.35$, $p < 0.0001$; P25%: $F(4,64) = 199.36$, $p < 0.0001$; S24Hz: $F(4,64) = 348.63$, $p < 0.0001$], further confirming the positive effect of higher gating rates.

The ANOVAs also revealed significant interactions between gating rate and stimulus type for all three gating

types [P50%: $F(12,192)=7.32$, $p<0.0001$; P25%: $F(12,192)=6.06$, $p<0.0001$; $F(12,192)=16.67$, $p<0.0001$]. These interactions suggest that the advantages of sentence-level cues and reduced memory load were not equally distributed across primary rates. This was especially so in the S24Hz dual-rate condition which had the highest interaction effects size: $\eta_p^2=0.57$, compared to $\eta_p^2=0.31$ and $\eta_p^2=0.27$ for P50% and P25% single rate conditions, respectively. As can be seen in Fig. 4, for all three gating types (P50%, P25%, and S24Hz), intelligibility differences based on the number of words in a sequence, absent at the lowest rate of 0.5 Hz, appeared as rate increased. Moreover, for P50% gated stimuli, the interaction was driven primarily by the flattening of the high-rate portion of the rate-intelligibility curve with increasing memory load and a significant local minimum ($p<0.05$) for the three-word sequences only. For P25% gated stimuli, the increased separation of the high-rate portions of the four rate-intelligibility curves as well as the different locations of the local minima [significant ($p<0.05$) for all four functions] between words and sentences contributed to the observed interaction. For the S24Hz function, the interaction was also driven by the differences in performance obtained with primary-rate gating of words or sentences, further suggesting that access to sentence-level cues for dual-rate stimuli was primary-rate dependant. Compared to results obtained with sentences in Experiment 1, word stimuli gated with two rates also showed a reduced range of variation relative to single-rate performance. Whereas for sentences the differences between the P50% and S24Hz conditions ranged between 3 and 37 percentage points across primary rates, this range was reduced by almost one half for words, regardless of memory load [i.e., 7–19 (1W), 7–20 (3W), 4–21 (5W)].

C. Discussion

The results of Experiment 2 demonstrated that the intelligibility of gated speech varied with speech materials (i.e., words versus sentences) and the number of words in a gated sequence. As expected, high-context sentences were generally more intelligible than unrelated words. However, the intelligibility advantage of sentences varied with gating type and was the highest and most consistent for stimuli gated at a single rate with a 50% duty cycle. On the other hand, for the stimuli gated with a 24-Hz secondary rate, the intelligibility advantage of sentences depended on primary rate, being highest for the low and high primary gating rates and significantly less for the mid-range rates of 2–4 Hz (Fig. 4).

Overall, in comparison with sentences, the intelligibility of words gated with two rates, although still nonmonotonic, was more linearly related to the intelligibility obtained when gating with the primary rates alone (Fig. 3). This pattern of results may indicate that for sentences, access to higher level context cues can be constrained by the availability of low-level speech cues. After dual-rate gating or when the duty cycle was reduced to 25%, with both cases reducing the efficacy of the remaining low-level cues, access to additional sentence-level cues would be impaired so that intelligibility remained at the individual word level. This effect, however, would not be as detrimental with individual semantically

unrelated words which *a priori* lack any contextual predictability across words. However, other factors that might have also contributed to the differences in intelligibility between words and sentence include cross-talker variation, the absence of coarticulation across words, and sentence-specific prosodic cues (Spitzer *et al.*, 2009; Wang and Humes, 2010).

The effect of word-sequence length on the shape of the rate-intelligibility functions suggests that memory load can affect the ability to utilize cues that remain after either single- or dual-rate gating. This is especially evident in the single-rate 50% duty-cycle condition (Fig. 4 top) where increasing the number of words in the sequence led to a flattening of the intelligibility curves at higher rates. A similar trend was observed for the single-rate 25% duty-cycle and the S24Hz dual-rate conditions, although the differences across the word-length functions were much smaller, potentially due to overall lower performance levels (Fig. 4, middle and bottom panels).

In addition to memory load effects, another factor that can influence the intelligibility of word sequences of variable length in interrupted speech is the listener's ability to segment individual words from the sequence by finding appropriate word boundaries (Spitzer *et al.*, 2009). This could be particularly challenging in a naturally produced multi-word utterance due to coarticulation and prosodic contour effects. However, in the present stimuli the challenges associated with word segmentation were alleviated by the relatively rigid linguistic and temporal structure of the word stimuli. All words were monosyllabic with the same phonological structure (i.e., CVC) and of similar duration separated by an average 80-ms interstimulus interval (ISI) without any coarticulatory overlap. The relatively long ISIs and stable beginning and end points of all stimuli would be expected to assist proper segmentation of the multi-word sequences.

The greater differences in intelligibility for different word-length sequences at higher rates are consistent with the interpretation that the greater memory load negatively affected the encoding of the words, as encoding of words with multiple interruptions at higher rates demanded greater resources. On the other hand, lack of significant effect of gating type in the intelligibility of words gated at 0.5 Hz, and in the three- and five-word conditions gated at 1 Hz, indicate that when speech-on times were equal to or greater than word durations so that gating produced little distortion of individual words, increased memory load had relatively little effect.

IV. GENERAL DISCUSSION AND SUMMARY

The ability of listeners to integrate temporally distributed audible speech fragments into coherent percepts has been well documented for speech interrupted at a single rate either by gating or the introduction of a modulated masking noise. The findings of the present study extend this experimental approach and indicate that for a given proportion of the original speech, rate effects can be constrained by the duration of the gated speech segments. Specifically, systematic interruptions of words and sentences with two concurrent gating rates resulted in a highly nonmonotonic performance

across primary rates. Even with application of a fast secondary rate of 24 Hz, which on its own results in ceiling-level performance, there was considerable variation in intelligibility across primary gating rates. However, such cross-rate variations in intelligibility were affected by contextual predictability (unrelated words versus sentences) and memory load (number of words in a stimulus). For fast secondary gating of sentences with low or high primary rates, performance, except at 10 Hz, approached that obtained with the primary rate alone and a 50% duty cycle. On the other hand, for intermediate primary gating rates of 2–4 Hz, performance with the 24-Hz secondary rate remained substantially lower than that obtained with a single rate and a 50% duty cycle. However, the nonlinear effect of fast secondary gating was attenuated for less predictable semantically unrelated word stimuli where the differences between single- and two-rate functions were smaller.

Interrupted sentences were overall more intelligible than interrupted words, a finding consistent with previous research that demonstrated differences in the intelligibility of interrupted speech for different types of speech materials varying in linguistic structure (Dirks *et al.*, 1969; Dirks and Bower, 1971; Wang and Humes, 2010). However, the sentence advantage was most consistent in conditions with a single gating rate and a 50% duty cycle. For single rates with a 25% duty cycle and for stimuli gated with a 24-Hz secondary rate, the sentence advantage was observed only for low and high primary gating rates. For the intermediate rates (2–4 Hz), the intelligibility of sentences and word sequences was approximately the same. In general for word stimuli, increasing the number of words in a sequence had an adverse effect on intelligibility with an inverse relationship between word-sequence length and performance. This effect was independent of the type of gating performed, but interacted with gating rate such that increasing the sequence length from three to five words had a negative effect only for primary rates above 1 Hz. This interaction suggests that negative consequences of increased memory load resulted from greater encoding difficulty at higher rates as opposed to a greater likelihood of forgetting as sequence length increased. At lower rates that were more likely to preserve complete individual words during the speech-on times, the memory load effects were not obtained for three- and five-word strings.

The nonmonotonic variation across primary rate in the intelligibility of dual-rate gated stimuli may reflect the effect of gating rate on low-level speech cues. For a low rate of 0.5 Hz, stimuli are characterized by a long speech-on time which may contain one or more undistorted words, and also a long speech-off time during which no information is available. When the relatively long speech-on intervals are gated with fast secondary rates, perceptual integration of short temporal fragments into words may be similar to that of undistorted speech gated at a single fast rate. However, the small but significant difference of 6.07 points between P50% and S24Hz intelligibility at 0.5 Hz for sentences also suggests limits to that analogy, likely due to the presence of incomplete words arising from the misalignment of speech-on- and speech-off times and word boundary locations. In the case of a high

primary rate of 8 Hz, both speech-on and speech-off times are short enough to sample most of the perceptually salient speech cues, so that full words can be reconstructed completely based primarily on low-level acoustic-phonetic cues. Thus, a fast secondary rate would have a minimal effect, comparable to that observed with fast primary rates alone. On the other hand, for primary rates of 2–4 Hz, the duration of the speech-on intervals is less than the average word duration so that more than half of the following word may be omitted during the speech-off time. In that case, syllable-sized word fragments that remain after primary gating may not contain sufficient information to constrain lexical access to a small number of word candidates and the additional distortion introduced by the secondary rate could make the task even more difficult. This interpretation is consistent with previous studies which report local minima in single rate-intelligibility functions around 2 Hz with increased sensory degradation of the stimuli (e.g., use of lowpass, time-compressed, or vocoded speech, or testing of hearing-impaired listeners) or with increased cognitive and linguistic difficulty of the material (Miller and Licklider, 1950; Nelson and Jin, 2004; Jin and Nelson, 2010; Shafiro *et al.*, 2011).

The poor performance obtained with fast secondary gating of primary rates of 2–4 Hz may also indicate a processing bottleneck created by the need to conduct a more exhaustive higher order lexical search, while simultaneously integrating finer low-level temporal cues that remain following secondary gating. This explanation is consistent with the Reverse Hierarchy Theory (RHT) account of speech perception that predicts performance penalties when simultaneous perceptual access to finely grained low-level stimulus details and higher order information (such as sentence-level cues) is needed (Nahum *et al.*, 2008; Ahissar *et al.*, 2008). According to RHT, speech perception typically proceeds from global to local levels, where existing higher order categories at the global level interact with local sensory input in a top-down manner. For undistorted stimuli presented in quiet, higher levels of perceptual hierarchy that pertain to the meaning of a spoken message are accessed before lower sensory levels. Numerous and redundant acoustic cues in an undistorted signal typically allow for a quick and accurate determination of correct high-level response categories, such as words, without scrutinizing the low-level sensory information. However, when such direct word-level access is impaired, as in interrupted speech, a slow backward search is initiated to resolve the ambiguity by examining low-level stimulus details which are then recursively checked against higher level categories. This iterative process is slow and impairs the simultaneous use of higher level semantic information. Within the RHT framework, the intelligibility of sentences interrupted with 2–4 Hz single rates and a 50% duty cycles can be maintained due to the slower iterative matching process with a greater reliance on bottom-up cues. However, introduction of an additional fast secondary gating impairs the ability to use the sensory evidence and leads to a significant drop in accuracy.

The presence of local minima around 2–4 Hz for sentences gated with a single rate and a 25% duty cycle or a 24-Hz secondary rate is also consistent with the view that

emphasizes the importance of the low-rate modulation spectrum (Houtgast and Steeneken, 1985; Greenberg, 1996). These gating rates produce periodic interruptions that affect the region of the speech modulation spectrum that has been previously shown to be critical for speech perception (Drullman *et al.*, 1994). The rate-intelligibility functions of the present work that show local minima resemble the inverse of the typical speech modulation spectrum. The effect of speech gating on intelligibility may in part relate to modulation masking, that is, gating rates that coincide with the dominant region of the speech modulation spectrum introduce the greatest interference. Decreasing duty cycle from 50 to 25% does increase the peak in the modulation spectrum at the gating rate. The effect of duty cycle on intelligibility is thus consistent with involvement of modulation masking. However, the introduction of a secondary gating rate has little effect on the modulation spectrum beyond the region proximal to the secondary rate.³ In this case, involvement of modulation masking would not predict the local minima around 2–4 Hz in the S24Hz conditions. As concluded by Kwon and Turner (2001), masking in the modulation domain appears unable to fully account for the perception of temporally disrupted speech.

In summary, the present findings indicate that increasing the complexity of speech interruption by introducing a secondary gating rate results in highly nonmonotonic intelligibility patterns. Results obtained with application of fast secondary gating rates to speech already gated at a slower rate were consistent with the general premise of the glimpsing account: more frequent sampling of the speech cues over a wider temporal range within each primary gating cycle improved intelligibility. However, the magnitude of improvement in intelligibility with fast secondary gating varied with the primary gating rate and appeared to depend on the quantity and quality of the low-level speech cues needed for word identification. In the absence of sufficiently robust low-level speech cues with primary gating rates of 2–4 Hz, listeners could not benefit from higher order sentence-level cues, with performance remaining comparable to that of unrelated words. These results are consistent with the view that speech perception operates on different time scales, associated with different perceptual and cognitive processes. It appears that optimal perceptual reconstruction of information in interrupted speech involves the effective integration of high-order linguistic and low-level speech cues.

ACKNOWLEDGMENTS

We would like to thank Sejal Kuvadia for her assistance with subject testing and Dr. Arthur Boothroyd and Dr. Jeremy Fields for providing useful comments on an earlier draft of the manuscript. This work was partially supported by grants from NIH NIDCD (R03 DC008676 and R15 DC011916).

¹Although temporal integration of gated speech may differ in certain ways from that of speech interrupted by modulated noise (e.g., smoothing over short silent intervals cannot be based on auditory induction, and gating can introduce artificial onsets and offsets), interruption by modulated noise also introduces variables that have been known to affect speech intelligibility

such as forward and backward masking, modulation depth, and signal-to-noise ratio (Miller and Licklider, 1950; Dirks and Bower, 1970; Wilson and Carhart, 1969; Festen and Plomp, 1990; Nelson and Jin, 2003; George *et al.*, 2006; Füllgrabe *et al.*, 2006). However, there is likely considerable overlap between the two approaches to speech interruption as follows from the findings of Jin and Nelson (2010) who reported moderate-to-high correlations for the intelligibility of speech interrupted by either gating or modulated maskers.

²With the starting phase of both the primary and secondary rates 0 degrees, when the secondary rate was twice the primary rate, the result of dual-rate gating was equivalent to gating at the primary rate alone with a 25% duty cycle. For relationships between primary and secondary rates other than doubling, the result can no longer be described in terms of the primary rate alone. However, the total proportion of the original speech remaining after dual-rate gating was always equivalent to 25%, regardless of rate relationship.

³With concurrent gating at two rates, both rates are evident in the modulation spectrum with the difference and sum distortion components as the dominant intermodulation. Secondary gating at 24 Hz of primary rates of 2–4 Hz thus does not significantly alter the dominant region of the speech modulation spectrum.

- Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (2008). "Reverse hierarchies and sensory learning," *Philos. Trans. R. Soc. B* **364**, 285–299.
- 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., Meyers, M. D., and Brubaker, B. S. (1988). "Illusory continuity of interrupted speech: Speech rate determines durational limits," *J. Acoust. Soc. Am.* **84**, 1635–1638.
- Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeckowski, C. (1984). "Standardization of a test of speech perception in noise," *J. Speech Hear. Res.* **27**, 32–48.
- Boothroyd, A. (2010). "Adapting to changed hearing: the potential role of formal training," *J. Am. Acad. Audiol.* **21**, 601–11.
- Buss, E., Whittle, L. N., Grose, J. H., and Hall, J. W. (2009). "Masking release for words in amplitude-modulated noise as a function of modulation rate and task," *J. Acoust. Soc. Am.* **126**, 269–280.
- Cooke, M. (2006). "A glimpsing model of speech perception in noise," *J. Acoust. Soc. Am.* **119**, 1562–1573.
- Dirks, D. D., and Bower, D. R. (1970). "Effect of forward and backward masking on speech intelligibility," *J. Acoust. Soc. Am.* **47**, 1003–1008.
- Dirks, D. D., and Bower, D. R. (1971). "Influence of pulsed masking on spondee words," *J. Acoust. Soc. Am.* **50**, 1204–1207.
- 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.
- Drullman, R., Festen, J. M., and Plomp, R. (1994). "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Am.* **95**, 1053–1064.
- 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.
- Füllgrabe, C., Berthommier, F., and Lorenzi, C. (2006). "Masking release for consonant features in temporally fluctuating background noise," *Hear. Res.* **211**, 74–84.
- George, E. L. J., Festen, J. M., and Houtgast, T. (2006). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **120**, 2295–2311.
- Grant, K. W., and Seitz, P. F. (2000). "The recognition of isolated words and words in sentences: Individual variability in the use of sentence context," *J. Acoust. Soc. Am.* **107**, 1000–1011.
- Greenberg, S. (1996). "Understanding speech understanding," *Proceedings of the ESCA Tutorial and Advanced Research Workshop on the Auditory Basis of Speech Perception*, Keele, England.
- Gustafsson, H. A., and Arlinger, S. D. (1994). "Masking of speech by amplitude-modulated noise," *J. Acoust. Soc. Am.* **95**, 518–529.
- Houtgast, T., and Steeneken, H. (1985). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *J. Acoust. Soc. Am.* **77**, 1069–1077.
- Howard-Jones, P. A., and Rosen, S. (1993). "The perception of speech in fluctuating noise," *Acustica* **78**, 258–272.
- Huggins, A. W. (1975). "Temporally segmented speech," *Percept. Psychophys.* **18**, 149–157.
- Jenkins, J. J., Strange, W., and Edman, T. R. (1983). "Identification of vowels in 'vowelless' syllables," *Percept. Psychophys.* **34**, 441–450.

- Jin, S. H., and Nelson, P. B. (2010). "Interrupted speech perception: The effects of hearing sensitivity and frequency resolution," *J. Acoust. Soc. Am.* **128**, 881–889.
- Kwon, B. J., and Turner, C. W. (2001). "Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?," *J. Acoust. Soc. Am.* **110**, 1130–1140.
- Mattys, S. L. (1997). "The use of time during lexical processing and segmentation: A review," *Psychonom. Bull. Rev.* **4**, 310–329.
- Miller, G. A., and Licklider, J. C. R. (1950). "The intelligibility of interrupted speech," *J. Acoust. Soc. Am.* **22**, 167–173.
- Moore, B. (2003). "Temporal integration and context effects in hearing," *J. Phonetics* **31**, 563–574.
- Nahum, M., Nelken, I., and Ahissar, M. (2008). "Low-level information and high-level perception: The case of speech in noise," *PLoS Biol.* **6**, 978–991.
- 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.
- Nelson, P. B., and Jin, S. (2004). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **115**, 2286–2294.
- 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**, 1085–99.
- Peterson, G., and Lehiste, I. (1962). "Revised CNC lists for auditory tests," *J. Speech Hear. Disord.* **27**, 62–70.
- Pollack, I. (1975). "Auditory informational masking," *J. Acoust. Soc. Am.* **57**, S5.
- Poeppe, D., Idsardi, W. J., and Van Wassenhove, V. (2008). "Speech perception at the interface of neurobiology and linguistics," *Philos. Trans. R. Soc. B* **363**, 1071–1086.
- 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.
- Rosen, S. (1992). "Temporal information in speech: Acoustic, auditory and linguistic aspects," *Philos. Trans. R. Soc. London B* **336**, 367–373.
- Shafiro, V., and Raphael, L. J. (2007). "Phonetic interpretation of white noise in stop and fricative contexts," *J. Psycholinguist. Res.* **36**(6), 457–467.
- Shafiro, V., Sheft, S., and Risley, R. (2011). "Perception of interrupted speech: Cross-rate variation in the intelligibility of gated and concatenated sentences," *J. Acoust. Soc. Am. Express Lett.* **130**, EL108–EL114.
- Spitzer, S., Liss, J., Spahr, T., Dorman, M., and Lansford, K. (2009). "The use of fundamental frequency for lexical segmentation in listeners with cochlear implants," *J. Acoust. Soc. Am. Express Lett.* **125**, EL236–EL241.
- Strange, W., Jenkins, J. J., and Johnson, T. L. (1983). "Dynamic specification of coarticulated vowels," *J. Acoust. Soc. Am.* **74**, 695–705.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech. Hear. Res.* **28**, 455–462.
- TigerSpeech Technology, Innovative Speech Software. (<http://www.tigerspeech.com>) last retrieved 10/20/2010.
- Yost, W. A., and Sheft, S. (1989). "Across critical band processing of amplitude modulated tones," *J. Acoust. Soc. Am.* **85**, 848–857.
- Yost, W. A., and Sheft, S. (1994). "Modulation detection interference: Across frequency processing and auditory grouping," *Hear. Res.* **79**, 48–58.
- 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.
- Warren, R. M. (1984). "Perceptual restoration of obliterated sounds," *Psychol. Bull.* **96**, 371–383.
- Watson, C. S. (2005). "Some comments on informational masking," *Acta Acust. United Acust.* **91**, 501–512.
- Wilson, R., and Carhart, R. (1969). "Influence of pulsed masking on the threshold for spondees," *J. Acoust. Soc. Am.* **46**, 998–1010.
- Wingfield, A., and Tun, P. A. (2007). "Cognitive supports and cognitive constraints on comprehension of spoken language," *J. Am. Acad. Audiol.* **18**, 548–558.