

Effects of age and hearing loss on the recognition of interrupted words in isolation and in sentences

Gary R. Kidd^{a)} and Larry E. Humes

Department of Speech and Hearing Sciences Indiana University Bloomington, Indiana 47405-7002

(Received 15 February 2011; revised 17 November 2011; accepted 5 December 2011)

The ability to recognize spoken words interrupted by silence was investigated with young normal-hearing listeners and older listeners with and without hearing impairment. Target words from the revised SPIN test by Bilger *et al.* [J. Speech Hear. Res. **27**(1), 32–48 (1984)] were presented in isolation and in the original sentence context using a range of interruption patterns in which portions of speech were replaced with silence. The number of auditory “glimpses” of speech and the glimpse proportion (total duration glimpsed/word duration) were varied using a subset of the SPIN target words that ranged in duration from 300 to 600 ms. The words were presented in isolation, in the context of low-predictability (LP) sentences, and in high-predictability (HP) sentences. The glimpse proportion was found to have a strong influence on word recognition, with relatively little influence of the number of glimpses, glimpse duration, or glimpse rate. Although older listeners tended to recognize fewer interrupted words, there was considerable overlap in recognition scores across listener groups in all conditions, and all groups were affected by interruption parameters and context in much the same way. © 2012 Acoustical Society of America. [DOI: 10.1121/1.3675975]

PACS number(s): 43.71.Es, 43.71.Gv, 43.71.An [MSS]

Pages: 1434–1448

I. INTRODUCTION

We often listen to speech under conditions that make it difficult to hear portions of the speech signal. Fluctuations in the level of the speech and background noise can render some of the spoken message inaudible or difficult to resolve. The ability to understand speech under conditions that obscure or distort parts of the speech signal is not well understood, but it is clear that there are large individual differences in this ability. For young normal-hearing listeners, it has been shown that this ability is largely independent of spectral and temporal resolving power as measured with simple or complex non-speech sounds (e.g., Kidd *et al.*, 2007; Surprenant and Watson, 2001). Furthermore, this appears to be a basic *familiar sound recognition* ability that is used in the recognition of both speech and non-speech sounds under difficult listening conditions (Kidd *et al.*, 2007). For many older adults, the ability to understand speech under difficult listening conditions is largely determined by a listener’s hearing sensitivity, as measured by the audiogram (e.g., see review by Humes and Dubno, 2010). However, it is clear that there are substantial individual differences in this ability that are not accounted for by the audiogram. These differences are most clearly observed under aided listening conditions which have restored the audibility of the speech signal (Humes and Dubno, 2010).

For older adults, the emerging evidence suggests that various cognitive abilities may be significant contributors to individual differences in speech-in-noise performance, especially when the “noise” is speech or speech-like (see review by Akeroyd, 2008). The exact nature of the cognitive processes underlying such individual differences in performance

have not been ascertained and likely vary with the stimulus details involved in a particular study. Among the many cognitive measures examined, working memory is perhaps the most frequently identified predictor of speech recognition abilities (Akeroyd, 2008). Other cognitive abilities, such as processing speed (Salthouse and Madden, 2008; Van Rooij *et al.*, 1989) and resistance to distraction (Darowski *et al.*, 2008; Hasher and Zacks, 1988), have also been associated with speech recognition abilities, but a clear picture of the relative importance of different cognitive abilities or the mechanisms by which they influence speech recognition has not emerged.

In general, however, for young and older adults alike, individual differences in the ability to recognize speech in the presence of potentially interfering noise appear to be due to some combination of at least two basic underlying abilities: (1) attentional focusing; defined as the ability to selectively attend to relevant acoustic information within a complex waveform that may include multiple signals, and (2) contextual inference; defined as the ability to generate good hypotheses regarding the likely contents of inaudible or distorted portions of an utterance, based on available information in the surrounding context. These two abilities may often work together: listeners who are better at making good guesses about missing speech information may also be better able to hear out the relevant acoustic cues in partially masked speech, because they adopt listening strategies that help them focus attention at the appropriate spectral and temporal locations. Thus, top-down processes utilizing partial information not only help to fill in missing information; they can also facilitate attentional focusing and bottom-up processing that provides crucial information about the speech signal.

There is a considerable body of literature relating to both of these abilities. Studies of attentional focusing as it

^{a)}Author to whom correspondence should be addressed. Electronic mail: kidd@indiana.edu

relates to speech recognition have examined the ability to perceptually segregate speech from background information (including other speech) as well as the ability to selectively attend to relevant information within the speech signal. (See Jones and Boltz (1989), Moore (2003), and Summerfield *et al.* (2006) for discussions of selective attention as it relates to speech perception.) Factors that interfere with attentional focusing are often lumped together and referred to as “informational masking” in the current literature. Although the term is too broad to be very useful, discussion and examination of the types or causes of informational masking has led to useful distinctions that are essentially different conditions or factors that make it more difficult to focus attention on a target stimulus in the presence of potentially interfering stimuli. For example, similar targets and distractors may perceptually fuse or mix in a way that makes target properties difficult to resolve (as discussed in work on auditory scene analysis [see, for example Bregman (1990); Cusack and Carlyon (2004)]. Distractors that can be more easily perceptually segregated from the target may still interfere by making it more difficult to “find” targets (i.e., allocate attention to the correct spectral and temporal locations) and ignore distractors within a complex stimulus, as shown in many investigations of stimulus uncertainty [see Watson (1987), Lutfi (1993), and Kidd *et al.* (2002), for examples and discussions of stimulus uncertainty effects]. Finally, distractors can attract attention and deplete processing resources, thus reducing the ability to attend to and process target information. [See Durlach *et al.* (2003), Watson (2005), Mattys *et al.* (2009), Cooke *et al.* (2008) for related discussions of informational masking.]

Studies of the use of context at many different levels (segmental, supra-segmental, lexical, semantic) have shown that listeners are often quite good at predicting or filling in information based on the surrounding context, (e.g., Elliott, 1995; George *et al.*, 2007; Grant and Seitz, 2000; Nittrouer and Boothroyd, 1990). Knowledge of the constraints on speech and language can help a listener make informed guesses about missing information, based on the context both before and after a masked portion of speech. This knowledge can also help to dynamically focus attention, which can facilitate the processing of relevant information at the moment it becomes available. The ability to effectively use context can be maintained with age, but increased processing demands associated with a degraded signal (due to hearing loss), cognitive aging, and a reduced ability to resist auditory distractions can make it difficult for older listeners to effectively use their accumulated linguistic knowledge (Darowski *et al.*, 2008; Mattys *et al.*, 2009; McCoy *et al.*, 2005).

The relative contribution of attentional focusing and contextual inference in facilitating speech recognition under difficult listening conditions likely varies across individuals and listening situations. However, little is known about the distribution of these abilities in listeners, the extent to which they may be correlated, or the amount of variance in speech recognition abilities they may account for. Although it may not be possible to isolate the two abilities entirely, it is possible to experimentally manipulate the degree to which each ability is called upon in a speech recognition task. The experiments described here are part of an effort to assess

these abilities and to determine how they may change with age and hearing loss.

One way to study the contextual inference ability is to examine the use of partial speech information in interrupted (or “glimpsed”) speech; that is, speech in which portions have been removed and replaced with silence. There have been many studies of the recognition of interrupted speech from a variety of perspectives. Listeners have been shown to be quite good at understanding interrupted speech over a wide range of interruption conditions (e.g., Miller and Licklider, 1950). This has been found for interrupted speech in quiet, in different types of noise, and for speech interrupted by or masked by modulated noise (Cooke, 2003; George *et al.*, 2006; George *et al.*, 2007; Iyer *et al.*, 2007; Jin and Nelson, 2010; Lee and Kewley-Port, 2009; Li and Loizou, 2007; Powers and Wilcox, 1977; Wang and Humes, 2010; Wilson *et al.*, 2010).

Each of these stimulus conditions tests different abilities to different degrees. Interrupted speech in quiet provides a measure of the ability to make use of partial information that is not confounded with the ability to hear out speech in noise. The recognition of speech interrupted by noise requires the use of partial information, but also introduces the possibility of masking of the speech (forward and backward) by the interleaved noise. Interrupted speech in continuous noise provides a test of both the ability to make use of partial information and the ability to hear out speech in noise. However, when favorable SNRs are used, this can provide a more natural listening condition with relatively little influence of the noise and minimization of the temporal-masking problems associated with modulated or interrupted noise. The more ecologically valid case of continuous speech in fluctuating noise also tests both abilities, but with less control of the partial information that is available to listeners (because of individual differences in the ability to take advantage of the dips in the noise and the difficulties in systematically manipulating the amount of speech available in the dips). Although some investigators have examined a variety of masking conditions designed to reveal different spectral or temporal portions of the speech waveform to the listener (e.g., Cooke, 2006), there is generally considerable uncertainty about the amount of masked speech information that is perceived (or glimpsed) by a given subject. The present study examines interrupted speech in quiet and in continuous noise, at a favorable SNR, to emphasize the ability to make use of partial information under conditions that provide good control of the available speech without strongly taxing the ability to isolate speech from background noise. Although the partial information provided by this type of stimulus interruption may be different from the information typically available in noisy listening conditions in which the SNR varies, the greater control over the available speech information and the abundant evidence that listeners are quite good at understanding interrupted speech justifies its use as a means of investigating the ability to make use of partial speech information. Furthermore, recognition performance with interrupted speech has been found to be highly correlated with the recognition of speech interrupted by noise and speech in modulated noise (Jin and Nelson, 2010), and models

of speech recognition based on the integration of brief glimpses accurately predict recognition performance in both stationary and fluctuating noise (Cooke, 2006). Thus, it appears that for a fairly wide range of interruption conditions, the ability to make use of glimpses of speech is largely independent of the way in which the glimpses are made available. Some exceptions are discussed below.

A previous study by Wang and Humes (2010) examined the effect of various interruption parameters (rate, glimpse duration, proportion of word glimpsed) and lexical difficulty on the recognition of isolated words in quiet by young, normal hearing listeners. In that study, the proportion of the speech that was glimpsed accounted for most of the variance in the recognition data, and the effects of the other interruption parameters were generally negligible and depended on the proportional duration. The relatively large effect of the total glimpsed proportion is consistent with earlier findings with speech in modulating noise (Cooke, 2006). The effects of interruption frequency and glimpse duration were small and also generally consistent with previous findings (e.g., Huggins, 1975; Li and Loizou, 2007; Miller and Licklider, 1950; Nelson and Jin, 2004). That is, performance tends to be lower at slow interruption rates (around 4 Hz) when listeners are required to “fill in” over longer stretches of missing speech. Slow interruption rates are less disruptive when larger proportions of the speech are preserved and less extrapolation through silence is required. An interaction between lexical difficulty (based on word frequency, neighborhood density, and neighborhood frequency) and the proportion of speech available (glimpsed) showed that the advantage of lexically easy words is greater when more speech information is available. This is consistent with the idea that top-down processes based on linguistic knowledge are more effective when the amount of acoustic (bottom-up) information is greater.

The present study extends this line of research to examine the effect of aging and hearing impairment on the ability to make use of the information available in interrupted speech. As in Wang and Humes (2010), the proportion of speech available and the number of glimpses were systematically varied to produce a range of interruption rates, glimpse durations, and silent (interruption) durations. This study, unlike Wang and Humes (2010), utilized stimuli from the revised SPIN test (Bilger *et al.*, 1984) to examine the recognition of isolated words, as well as words in sentence context with high and low predictability. The SPIN target words (the final word of each SPIN sentence) were presented in quiet and in full sentence context with speech-shaped noise (rather than the babble normally used with the SPIN materials). The identical interruption patterns were used for target words in isolation and in sentence context. In the sentence context, the interruption pattern was extended throughout the entire sentence to form a continuous interruption pattern. This set of conditions allows for an analysis of the effect of context on the use of partial speech information available in interrupted speech. The low probability sentences provide primarily non-linguistic (spectral and temporal) information about the speech and about the glimpse pattern that could facilitate recognition of the target word at the end of each sentence, while the high probability sentences also include

linguistic (semantic) information that significantly reduces uncertainty concerning the identity of the target word. The inclusion of young normal hearing (YNH), elderly normal hearing (ENH) and elderly hearing impaired (EHI) listeners allows for an examination of the influence of aging and hearing loss on the use of partial speech information in different contexts. This will allow us to determine whether context effects and age effects with interrupted speech differ from those obtained with other types of degraded speech (e.g., Pichora-Fuller *et al.*, 1995; Sheldon *et al.*, 2008).

It is expected that, as in previous work, the total glimpsed proportion of a word will be the primary determinant of recognition performance with little or no effect of the duration or number of glimpses, except in the most extreme cases. Because the total glimpsed proportion is independent of the other glimpse parameters in the present study, this provides a stronger test of the overriding influence of the total amount of speech information provided, regardless (to a great degree) of the number and duration of the individual glimpses. Violations of this proportional rule that may occur in particular glimpsing conditions (e.g., with particular glimpse durations or glimpse rates) or with increasing age or hearing loss would indicate a change in listening strategies or abilities. Examination of the pattern of sensitivity to the number and duration of glimpses (e.g., greater reliance on longer or more frequent glimpses) that may occur if the proportionality rule is violated will provide clues to the nature of any altered listening strategies.

II. EXPERIMENT 1: ISOLATED GLIMPSED WORDS

The first study examined the effect of interruption on the recognition of isolated words in quiet, using the target words from the revised SPIN test (R-SPIN; Bilger *et al.*, 1984). This experiment is similar to one reported by Wang and Humes (2010), but it utilizes a different set of stimuli, different interruption patterns, and it includes ENH and EHI subject groups, as well as a YNH group. This experiment also provides a stronger test of the effect of the total proportion of speech available to the listener (proportion glimpsed), because word duration is not held constant (by digital expansion or contraction), as it was in Wang and Humes (2010). Although the method (STRAIGHT; Kawahara *et al.*, 1999) used to normalize durations in that study produced very natural-sounding words, a stronger test of the effect of the proportion of speech available is provided when the total duration of the words is allowed to vary, so that the proportion glimpsed is not confounded with the total duration glimpsed. In the present study, word durations ranged from 300 ms to 600 ms, with speech proportions of 0.33 and 0.67 of the total word duration. The number of glimpses was also varied (4, 8, or 12 glimpses), with glimpse duration (and inter-glimpse pause duration) determined by the combination of the number of glimpses, the proportion of speech glimpsed, and word duration. Constant-duration glimpses were distributed equally throughout each word, with the first and last glimpses coinciding with the onset and offset of a word. This set of conditions also creates a wide range of interruption rates,

determined (like glimpse duration) by the two factorially combined glimpse parameters and word duration.

A. Methods

1. Subjects

Three groups of listeners participated in Experiment 1. Twenty-four young normal-hearing (YNH) listeners (ages 19–26 yrs, mean = 22.2 yrs), eight elderly normal-hearing (ENH) listeners (ages 63–72 yrs, mean = 68.1 yrs), and eight elderly hearing-impaired (EHI) listeners (ages 68–90 yrs, mean = 77.4 yrs) were paid for their participation in this experiment. The YNH subjects were tested first and a larger sample size was used here to provide a better estimate of individual differences in this population. When the two groups of older adults were tested, the variability was monitored as the data collection progressed and a sample size of eight provided sufficient power to detect reasonably small differences of practical significance (5–10 percentage points). The younger subjects were students at Indiana University and the older subjects were residents of the Bloomington, Indiana community who had served in a previous hearing study (which included screening to rule out serious cognitive or physical impairment) and had indicated a willingness to participate in other research studies. (The earlier study evaluated cognitive ability with the Mini-Mental State Exam (Folstein *et al.*, 1975) with a score of 26 or higher as the criterion for inclusion. That study also required that participants be ambulatory, have 20/40 corrected vision, and be able to complete multiple test sessions of 90–120 minutes each.) None of the subjects had prior exposure to the SPIN materials or to any earlier studies of interrupted speech. All subjects were native speakers of English.

All YNH listeners had pure tone thresholds ≤ 25 dB HL (ANSI, 2004) for all octave frequencies between 250 and 8000 Hz. ENH listeners were required to have a pure tone average (PTA_{500, 1000, 2000 Hz}) ≤ 15 dB HL and a high-frequency PTA (HFPTA_{1000, 2000, 4000 Hz}) ≤ 25 dB HL. All subjects had normal tympanograms and otoscopic findings. The thresholds of all ENH and EHI listeners are shown in Fig. 1. Although some ENH subjects have measurable threshold elevations that do not meet the definition of normal-hearing used for the younger group (primarily due to slightly elevated thresholds at 4 kHz and above), these subjects would generally not be considered hearing-impaired by most clinicians. Moreover, because levels are adjusted to ensure audibility for all subjects in this study, group effects are more likely to be associated with changes in listening strategies or the presence of cochlear pathology that may accompany more severe hearing loss, rather than with relatively small differences in thresholds. Correlations between PTA and performance (presented below) show little or no influence of hearing sensitivity in these experiments.

2. Stimuli

The stimuli were monosyllabic words spoken by a male talker. The words were taken from the sentences included in the R-SPIN test, revised by Bilger *et al.* (1984) and

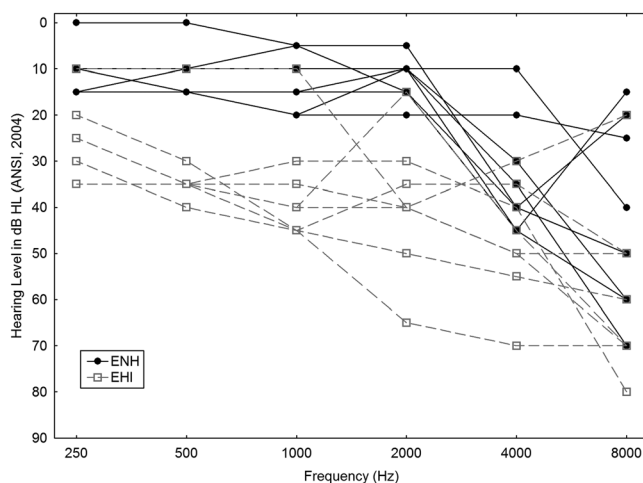


FIG. 1. Thresholds for ENH and EHI listeners in Experiment 1.

developed originally by Kalikow *et al.* (1977). The 400 R-SPIN sentences without background noise were digitized and the target word (always the last word in a sentence) was copied from each sentence. The boundary between the target word and the preceding word was selected to preserve the intelligibility of the target word while minimizing any audible trace of the preceding word. All edits of the word stimulus file were made at zero crossings of the waveform to minimize transients. The onset and offset of each digitally copied word was carefully examined to eliminate leading and trailing information (silence or other words) without removing any of the target word. Amplitudes were adjusted to achieve the same RMS level for all words. The intelligibility of the full set of 400 stimuli (200 words, each spoken in a high predictability (HP) and a low predictability (LP) context) was assessed in an open-set word-identification test in which each stimulus was presented once. Based on pilot testing with another group of 24 YNH subjects, a subset of 250 stimuli (125 words spoken in both a HP and a LP context) was selected for this study by including only words, with durations between 300 ms and 600 ms, which were correctly identified by at least 21 of the 24 subjects. The distribution of word durations is shown in Fig. 2.

Interrupted versions of the 250 tokens were created by digitally replacing portions of each word with silence. The onset and offset of each speech fragment or “glimpse” was smoothed with a 4-ms raised-cosine function to minimize spectral artifacts. Six glimpsing patterns were generated by creating 4, 8, or 12 equal-duration glimpses that combined equaled 1/3 (0.33) or 2/3 (0.67) of the total word duration. The first and last glimpses were always aligned with the beginning and ending of a word, with the other glimpses equally spaced with a constant pause duration, as illustrated in Fig. 3. Because of the variation in word duration, this set of conditions yielded a range of glimpse durations, pause durations, and interruption rates within each of the six glimpsing patterns shown in Fig. 3. The combination of six glimpsing conditions (three numbers of glimpses combined with two glimpsing proportions) and 250 words (two tokens of each of the 125 words, one spoken in HP context and one in LP context) resulted in 1500 stimuli.

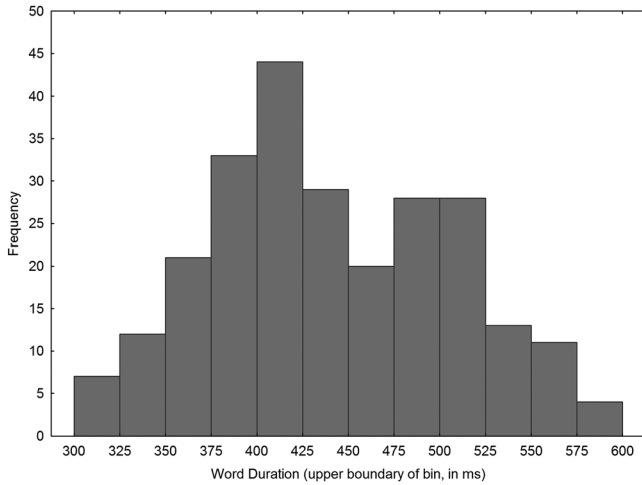


FIG. 2. The distribution of durations for SPIN target words selected for use in both experiments.

3. Presentation levels

Presentation levels were adjusted to ensure that speech information was audible for the HI listeners and to provide comparable presentation levels for all listeners. For the HI listeners, the long-term spectrum of the full set of words was measured and a filter was applied to shape the spectrum according to each listener's audiogram. The shaping was applied with a 68 dB SPL overall speech level as the starting point, and gain was applied as necessary at each 1/3 octave band to produce speech presentation levels at least 13 dB

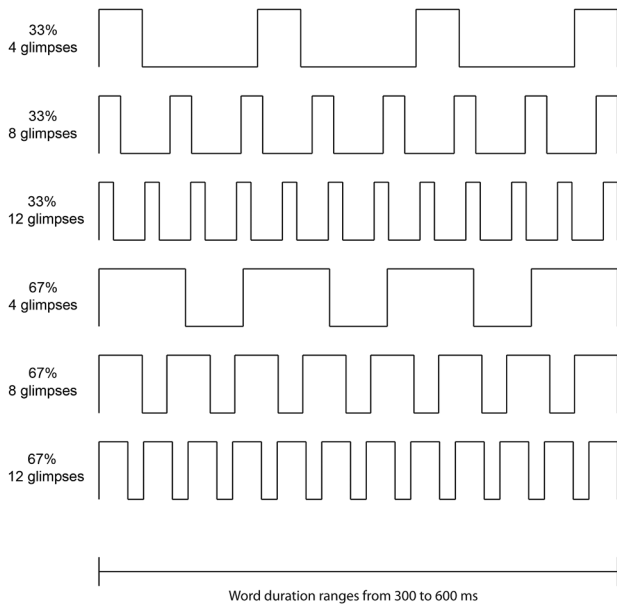


FIG. 3. Schematic representation of the interruption patterns used in this study. Relative on and off times for portions of a word are indicated by the high (on) and low (off) portions of each line pattern. The glimpse proportion and number of glimpses are indicated to the left of each corresponding line pattern. Initial and final glimpses were always aligned with word onset and offset, respectively. In Experiment 2, only the eight-glimpse patterns were used. When a word appeared in sentence context (always the final word), the interruption patterns were extended continuously through the entire sentence, maintaining the same target word alignment as in the no-context conditions.

above threshold from 125 Hz to 4000 Hz. Because this often resulted in relatively high presentation levels, a presentation level of 85 dB SPL (without any spectral shaping) was used for the YNH and ENH listeners to avoid level-based differences in performance between groups. Previous work has shown that presentation levels above 80 dB SPL generally lead to somewhat poorer intelligibility, for both uninterrupted speech (e.g., Dubno *et al.*, 2003; Fletcher, 1922; French and Steinberg, 1947; Pollack and Pickett, 1958; Studebaker *et al.*, 1999) and interrupted speech (Wang and Humes, 2010).

4. Procedure

All testing was done in a single-walled sound-treated booth that met or exceeded ANSI guidelines for permissible ambient noise for earphone testing (American National Standards Institute, 1999). Stimuli were presented to the right ear, using an Etymotic Research ER-3 A insert earphone. A disconnected earphone was inserted in the left ear to block extraneous sounds. Stimuli were presented by computer using Tucker Davis Technologies System 3 hardware (RP2 16-bit D/A converter, HB6 headphone buffer). Each listener was seated in front of a touchscreen monitor, keyboard, and mouse. On each trial, the word "LISTEN" was presented on the monitor, followed by the presentation of a word 500 ms later. The subject's task was to type the word they just heard using the computer keyboard. Subjects were instructed to make their best guess if they were unsure. The next trial was initiated by either clicking on (with the mouse) or touching a box on the monitor labeled "NEXT." In addition to responses that were spelled correctly, homophones and phonetic spellings were scored as correct responses, using a computer program for scoring. Several typographical errors, identified by visual examination of a large portion of the responses, were also added to the scoring database and counted as correct. (We find this procedure more efficient and at least as reliable as collecting spoken responses. Our subjects had little or no difficulty typing a single word on a computer keyboard.)

Testing consisted of five trial blocks of 300 trials (with short breaks after every 100 trials) in 1 or 2 sessions of no longer than 90 minutes each. Each of the 1500 stimuli (250 speech tokens \times 6 glimpse conditions) was presented once. A different word was selected randomly (from the full set of 250 tokens) on each trial and the sequence of trials cycled through a random permutation of the six glimpse conditions in each consecutive set of six trials. Prior to data collection, all subjects were presented with 24 practice trials, which included four examples of each of the six glimpse conditions using words that were not used in the main experiment.

B. Results

Percent-correct scores were computed for each subject in each condition and converted to rationalized arcsine units (RAU; Studebaker, 1985) to stabilize error variance while preserving a range of values similar to percent-correct scores. These scores were used in all analyses summarized below. Analysis of variance was used to examine main effects and

interactions using a 3 (listener groups) \times 3 (number of glimpses) \times 2 (total proportion glimpsed) factorial design with one between-subject factor (group) and two within-subject factors.

1. Age and hearing loss

Group differences were first examined to determine whether the ability to understand the interrupted words varied with age or hearing loss. Overall, there was a wide range of individual differences on this task, with considerable overlap in performance levels across the three groups. Differences between groups were not large, but there was a significant effect of group ($F(2,37) = 3.76, p < 0.05$). *Post hoc* analysis (Tukey HSD; $p < 0.05$) revealed that the only significant difference was between the YNH group (mean = 66.6 RAU) and the ENH group (56.2 RAU). The EHI group performed at a level in between these two groups (mean = 62.0 RAU). Although all listeners in the ENH group met the criteria used for normal hearing in this study, six of them had mild to moderate hearing loss at 4000 Hz (see Fig. 1). Thresholds for all ENH listeners met the 13-dB SL criterion up to 4000 Hz, but thresholds at 6000 fell slightly below 13-dB SL for five of these listeners. However, the same number of subjects in the EHI group fell below this criterion at 6000 Hz (due to gain limitations), and all but one subject fell below at 8000 Hz in each group. Thus, audibility alone does not appear to account for the lower performance of the ENH group.

2. Total proportion glimpsed

The performance of each listener group in each of the six conditions is shown in Fig. 4. This figure shows a clear advantage for conditions in which 2/3 of the total word duration was available for all groups, regardless of the number of glimpses. This was reflected in a significant main effect of the total proportion of glimpsed speech ($F(1, 37) = 383.95; p < 0.001$) and the lack of an interaction between that variable and the listener group ($F(2, 37) = 0.93; p > 0.05$). In general, the difference in speech-recognition performance

for the two proportions examined here is large and amounts to about 25–30 RAU across groups and number of glimpses.

3. Number of glimpses

The number of glimpses also influenced performance ($F(2, 74) = 8.78; p < 0.001$), but this variable interacted with listener group ($F(4, 74) = 22.05; p < 0.001$). The normal-hearing listeners (young and elderly) performed slightly worse as the number of glimpses increased, while EHI listeners performed worst with the fewest (four) glimpses. Performance with four glimpses for the EHI group was approximately the same as for the ENH group, while performance of the EHI group with 8 or 12 glimpses was close to that for the YNH group. In general, across groups and speech proportions, the effect of number of glimpses is small and amounts to about 5–7 RAU. A significant three-way interaction between group, number of glimpses, and the proportion of glimpsed speech ($F(4, 74) = 6.15; p < 0.001$) is primarily due to the fact that the differential effect of the number of glimpses on performance for the three groups is more pronounced at the smaller proportion of available speech.

4. Glimpse duration and interruption rate

Although glimpse duration and interruption rate were not directly manipulated as factors in the design, the combination of word duration, proportion of glimpsed speech, and number of glimpses resulted in a wide range of glimpse durations and interruption rates. Because these variables took on a large number of values over a wide range, it is possible to examine average performance for small ranges (bins) of glimpse rates or glimpse durations. Figure 5(a) shows the percentage of correctly recognized words as a function of glimpse duration (in 5-ms bins) for the YNH, ENH, and EHI groups. When the proportion of speech available is 0.67, the functions are fairly flat and similar across all three groups, but there is a rapid decrease for the largest one or two durations for the elderly groups, especially for the EHI group. For the lower proportion of glimpsed speech, the

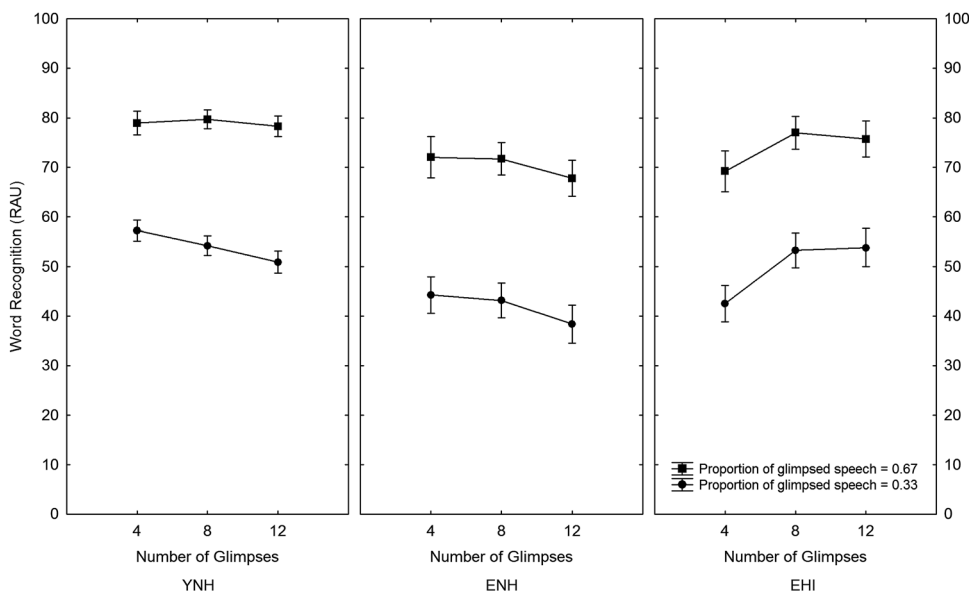


FIG. 4. Mean word recognition performance (in RAU) in each condition of Experiment 1. Performance is shown for each listener group as a function of number of glimpses, with glimpse proportion as the parameter. Error bars indicate standard errors.

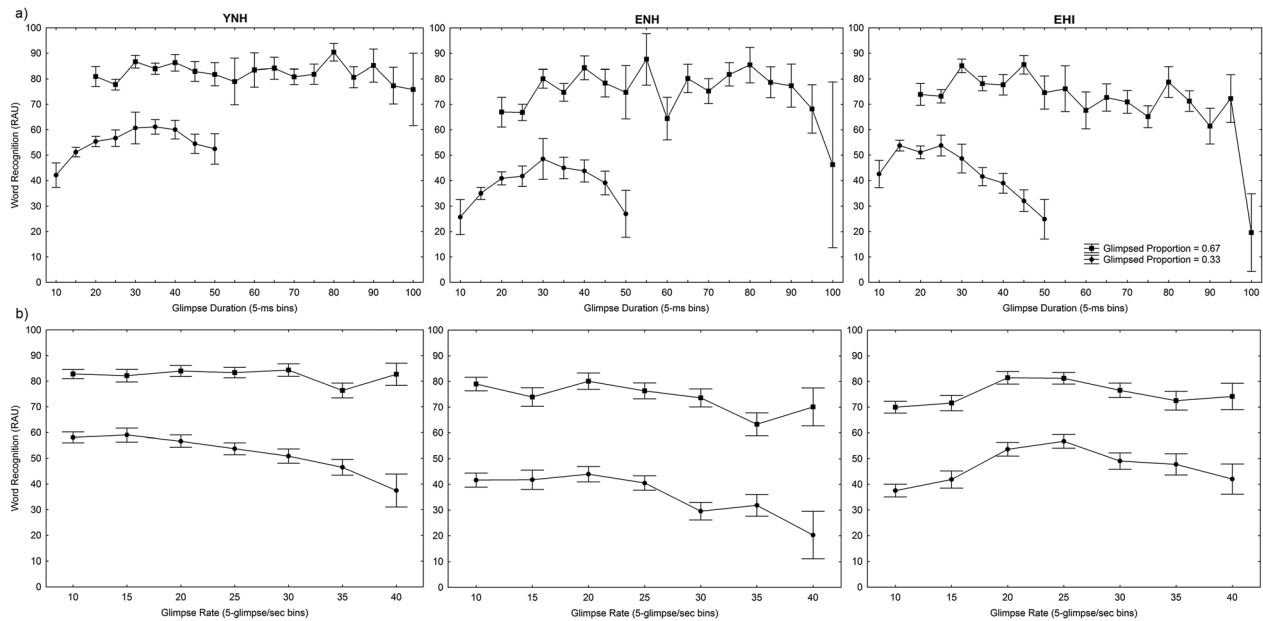


FIG. 5. Word recognition performance (in RAU) as a function of glimpse duration (a) and glimpse rate (b). Data for each listener group are shown in separate plots, with the glimpse proportion as the parameter. Error bars indicate standard errors.

functions are similar for all three groups and have an inverted U shape, with a peak near glimpse durations of 35 ms for the normal-hearing groups, but a slightly lower peak and a larger decrease with longer durations for the EHI group. This value is close to the minimum useable full-bandwidth glimpse of 25 ms estimated by [Cooke \(2006\)](#) and by [Vestergaard et al. \(2011\)](#) using different techniques. Although it is unclear why performance deteriorates with longer glimpses at the smaller glimpse proportion, this may be due to the longer silent (unglimpsed) portions of speech under those conditions (roughly 70 to 130 ms for the 0.33 proportion and 15 to 65 ms for the 0.67 proportion, for glimpse durations greater than 30 ms).

Figure 5(b) shows the percentage of correctly recognized words as a function of interruption rate (glimpses/sec in 5-glimpse/sec bins) for the YNH, ENH, and EHI groups. At the larger proportion of glimpsed speech, the functions are fairly flat and similar for all three groups. At the smaller proportion of glimpsed speech, performance decreases slightly with increasing glimpse rate, with the EHI group showing a slight decrease for the slowest glimpse rates as well (with a peak at 20–25 glimpses/sec).

Overall, the data in Fig. 5 show more similarities than differences across groups in terms of the influence of the total proportion of glimpsed speech, the glimpse duration, and the glimpse rate. When differences do appear among the groups, it is typically the EHI group that differs from the other two groups and only for a small subset of conditions, mainly when the proportion of speech available is 0.33. At this proportion, EHI listeners were adversely affected by relatively long glimpse durations to a greater extent than the other listeners. That the detrimental effect of glimpse duration was apparent at much shorter durations for the smaller proportion of available speech indicates that the duration of the inter-glimpse pause (which is longer at the smaller proportion for a given glimpse duration) is the crucial variable here.

Effects of glimpse rate are also more prominent at the smaller proportion of available speech, but the functions are remarkably flat over a wide range. The tendency for performance to decrease with increasing rate, although weak and only at the lower proportion of available speech, is somewhat inconsistent with earlier work ([Miller and Licklider, 1950](#); [Wang and Humes, 2010](#)). This pattern was observed by Wang and Humes at the higher proportions of available speech (0.50 and 0.75), but not at the lowest value (0.25). This difference may be due, at least in part, to the range of values investigated and to the fact that the highest rates in the present design occur with the shortest words as well as the briefest glimpses.

5. Correlations across conditions

Examination of the relative performance of subjects in the different conditions provides an indication of the degree to which subjects may be differentially affected by the different stimulus conditions. Although all conditions measure the ability to understand speech based on incomplete information (at least in part), some listeners may be more or less affected by particular glimpsing parameters (e.g., number of glimpses or glimpse durations) than others. While common method variance (i.e., shared variance due to common task and procedures) will contribute to higher correlations, lower correlations can still occur if subjects are differentially affected by the different stimulus conditions. This is especially true in the older group, because performance in some conditions may be differentially affected by aging and hearing loss.

For both younger and older listeners, correlations between all pairs of conditions were generally quite high. Because the numbers of subjects is relatively small (24 in the younger group and 16 in the combined older group), all scatterplots were examined for outliers or other anomalies

that might alter the interpretation of the correlation coefficients and none were found. The younger listeners were somewhat less consistent across conditions than the older listeners, mainly due to the performance of two listeners. Two of the 24 younger listeners showed unusually small improvements as the proportion of glimpsed speech increased, leading to lower correlations ($0.4 < r < 0.6$) when comparing the two glimpse proportions. However, with those two subjects removed, all correlations for the YNH listeners were greater than 0.67 and significant ($p < 0.001$), with an average of 0.85 (range; 0.67–0.95).

Performance was more consistent across conditions for the older listeners. For this group (ENH and EHI subjects combined), the average correlation was 0.84 (range; 0.61–0.98), with all correlations significant ($p < 0.05$).

6. Correlations with hearing sensitivity

With this degree of consistency across subjects, it is worthwhile to examine the extent to which age or hearing loss can account for word-recognition performance among the elderly listeners. The average word recognition score (in RAU) was computed across all conditions for each ENH and EHI subject, and correlations between that variable and both age and hearing loss were computed. Both a three-tone pure-tone average (PTA; mean of 500, 1000, and 2000 Hz) and a three-tone high-frequency pure-tone average (HFPTA; mean of 1000, 2000, and 4000 Hz) were computed. These correlations were all small and non-significant. The correlation between age and word recognition was -0.21 . Correlations between word recognition and both PTA and HFPTA were less than 0.1 (in absolute value). Thus, for this group of 63–90 year-old listeners, neither age nor hearing loss had a substantial influence on the ability to make use of partial information in audible speech.

III. EXPERIMENT 2: INTERRUPTED WORDS AND SENTENCES IN CONTINUOUS SPEECH-SHAPED NOISE

Experiment 1 replicated and extended the findings of Wang and Humes (2010) with a new set of words that can now be used to examine the ability to make use of partial speech information in the context of a full sentence. In Experiment 2, these same words were presented both in isolation and in the context of a full sentence. The sentences were the original 250 R-SPIN sentences from which the words used in Experiment 1 were extracted. This allowed for the examination of the effect of a larger speech context when the target word is highly predictable (e.g., “The watchdog gave a warning growl.”) and when the context provides very little information that can be used to predict the target word (e.g., “She wants to talk about the crew.”). This experiment employed a subset of the interruption patterns used in Experiment 1 (excluding the four- and twelve-glimpse patterns), applied to the entire sentence context. These interruption patterns provided the identical glimpses (locations and durations) of target words as in Experiment 1, but each interruption pattern was present throughout the preceding sentence context. The stimuli also differed from those in

Experiment 1 in that they were presented in speech-shaped noise. A signal-to-noise ratio (SNR) of +10 dB was used to avoid ceiling effects with the HP sentences while still allowing acceptable recognition scores in the most difficult conditions (based on pilot studies). The addition of noise also created more natural listening condition in which relatively low-level background noise was present during both audible and inaudible portions of the speech.

Previous work with the SPIN materials has generally shown a large effect of predictability on the recognition of the target words (e.g., Bilger *et al.*, 1984; Elliott, 1995; Humes *et al.*, 2007). A similar effect is expected here, but it is not clear how the effect may vary with different proportions of available speech. Linguistic properties of the target speech signal, such as lexical difficulty of the target words, have been demonstrated to impact speech-recognition performance for interrupted speech (Wang and Humes, 2010). Perhaps other top-down linguistic properties of the speech signal, such as the semantic richness of the content, will also impact the effects of interruption parameters on speech-recognition performance. For example, it is possible that the relative performance in LP and HP contexts might depend on the proportion of available speech or other interruption parameters. Performance in the HP context will likely depend on the percentage of non-target words identified, but even the context provided by the LP sentences may influence performance differentially in the various interruption conditions. LP-sentence contexts, for example, may facilitate the recognition of the target words due to co-articulatory cues, speaker and rate normalization, and the greater spectral-temporal predictability (prosodic patterns) of the final target words provided by the context. However, in the present case, the LP context might also provide some benefit not found in other experiments because listeners may be able to take advantage of the regular interruption pattern in the sentence to direct their attention more effectively within the target word. The inclusion of YNH, ENH, and EHI listeners in this experiment will allow for an evaluation of the effect of age and hearing loss on the use of context for different amounts and patterns of speech interruption.

A. Methods

1. Subjects

Experiment 2 employed the same three categories of listeners tested in Experiment 1. There were 12 listeners in each group: YNH (ages 20–23 yrs, mean = 21.8 yrs), ENH (ages 65–84 yrs, mean = 72.2 yrs), and EHI (ages 66–88 yrs, mean = 77.0 yrs). Recruitment, payment, and exclusion criteria were the same as for Experiment 1. Thresholds for the EHI and ENH listeners are shown in Fig. 6. All ENH subjects met the 13-dB SL criterion up to 4000 Hz, as did the EHI subjects, after amplification. As in Experiment 1, many subjects were somewhat below 13-dB SL at 6000 Hz (5 ENH and 10 EHI) and all but one ENH subject fell below this level at 8000 Hz.

The subjects in Experiment 2 had no prior experience with the SPIN stimuli, with one exception. One of the EHI listeners who had participated in Experiment 1 was inadvertently

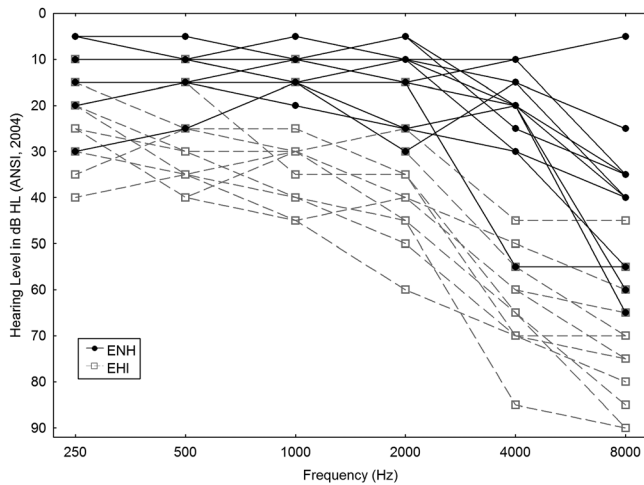


FIG. 6. Thresholds for ENH and EHI listeners in Experiment 2.

included in Experiment 2 (after a delay of nearly six months and a change in research assistants). When this was discovered, the subject's data were examined for evidence of any distinguishing performance characteristics. Because the subject was among the poorest performers in both experiments and was affected by the different conditions in the same way as other listeners, it was decided that familiarity with the stimuli was not an issue in this case and the subject was not replaced. Analysis of the data with this subject excluded produced essentially the same pattern of results as reported below.

2. Stimuli

The set of words and interruption patterns were identical to those used in Experiment 1, except that the number of glimpses of each target word was limited to the intermediate value (eight). In this experiment, each of the words was presented in isolation and in the context of the full sentence from the R-SPIN materials (Bilger *et al.*, 1984; Kalikow *et al.*, 1977). The individual words were the same ones used in Experiment 1, and all interruption patterns were applied to the words in the same manner. To avoid a level mismatch between the preceding sentence and the final word, the same RMS adjustment used to equate the final words for Experiment 1 was applied to the entire sentence. However, rather than using the same isolated word twice (i.e., the tokens from both HP and LP contexts), based on the equivalent overall performance observed in Experiment 1, only one token of each word (randomly chosen from either the HP or LP context) was used in this experiment in the isolated-word conditions. (Recognition of final-word tokens extracted from HP and LP sentences was nearly identical for YNH listeners in Experiment 1; mean percent correct for the two types was 67.0% vs 67.3%.) Thus, there were 125 words with two proportions of glimpsed speech (0.33 and 0.67) for a total of 250 isolated-word stimuli.

The sentences had durations that ranged from 1.32 to 2.40 sec, with an average of 1.72 sec. Each of the 125 target words was presented in both HP and LP sentence contexts in the set of 250 unique sentences. Two proportions of

glimpsed speech (0.33 and 0.67) were applied to the set of sentences to create a set of 500 sentence stimuli (250 HP and 250 LP) with interruption patterns matching those used with isolated words. (See the patterns for 8 glimpses in Fig. 3.) The interruption patterns were applied so that a consistent interruption pattern was maintained throughout the sentence, with target words keeping the same location and duration of glimpses as used in the isolated word context (and as used in Experiment 1). Because of variation in sentence duration, sentences often began with a single glimpse or silence duration that was shorter than the value used in the rest of the sentence.

Speech-shaped noise was added to each sentence using broadband noise shaped to match the long-term spectrum of the full set of target words. A different randomly chosen section of a 10-sec sample of noise was used for each sentence. The duration of the noise sample was matched to each sentence, with 250 ms of leading and trailing noise. The speech and noise were mixed at +10 dB SNR measured at the target word. Presentation levels were adjusted as in Experiment 1, with a relatively high level for the normal-hearing listeners (85 dB SPL) and shaped level adjustments for the hearing-impaired listeners to ensure audibility.

3. Procedure

Experiment 2 was conducted with the same equipment and testing conditions used in Experiment 1. Except for the arrangement of trial blocks, all procedures were also the same as in Experiment 1. Because listeners heard each of the 125 target words six times (two glimpse proportions in each of three contexts: isolated, in an HP sentence, and in an LP sentence), it was necessary to counterbalance the order of presentation to control for any effect of word or sentence familiarity. Six counterbalance orders were created in which every sequence of 125 trials was divided into three trial blocks (consisting of 41 or 42 trials each), one for each of the three context conditions. No word or sentence was repeated within each of these 125-stimuli sequences. The glimpsed proportion (0.33 or 0.67) was randomly chosen on each trial, with the constraint that each word was presented at each glimpsed proportion once in each context condition. Thus, in the first 125 trials, all three context conditions and both glimpse proportions were presented without repeating any stimuli. Each of the following series of 125 trials was arranged in the same manner. The order of the three context conditions within each consecutive series of 125 trials was counterbalanced, using all six permutations. Two of the 12 listeners in each group were randomly assigned to each of the six counterbalance orders.

The task was the same as in Experiment 1. In the sentence conditions, subjects were asked to enter the final word of the sentence only. Scoring of responses was the same as in Experiment 1. There was no direct measure of a subject's recognition of other words in the sentences. The focus in this study was placed on the influence of the preceding sentence information on the recognition of the target (final) words, not the perception of the entire sentence. Of course, to the extent that differences in performance are observed between the LP and HP sentences, one can safely assume that these

portions of the sentence were heard and recognized, at least partially. Testing took place over two sessions of no longer than 90 mins each. Prior to data collection, all subjects were presented with 24 practice trials, which included four examples of each of the two glimpse conditions using words that were not used in the main experiment.

B. Results

As in Experiment 1, percent correct scores were computed for each subject in each condition and converted to rationalized arcsine units (RAU; Studebaker, 1985) for the data analysis. To examine the effect of counterbalance order, the data were first analyzed using a 6 (counterbalance order) \times 3 (groups) \times 3 (context) \times 2 (total proportion glimpsed) factorial design. There was no main effect of counterbalance order ($F(5,18) = 2.27, p > 0.05$) and only listener group interacted with counterbalance order ($F(10,18) = 2.74, p < 0.05$). This interaction reflects marginally significant variation in the difference between groups across the six counterbalance orders. However, because these differences are small and follow no interpretable pattern, counterbalance order was removed from the design in the following analyses.

Analysis of variance of the resulting $3 \times 3 \times 2$ design revealed significant main effects for all three variables (listener group: $F(2,33) = 7.22, p < .005$; context: $F(2,66) = 749.86, p < 0.001$; and total proportion of glimpsed speech: $F(1,33) = 2292.96, p < 0.001$). Overall performance was slightly lower than in Experiment 1, as expected given the addition of background noise in this experiment, but the relative performance of the three groups was somewhat similar. *Post hoc* analyses (Tukey HSD; $p < 0.05$) revealed that YNH listeners (mean = 56.4 RAU) performed significantly better than ENH (40.0 RAU) and EHI (45.0 RAU) listeners, while the difference between the latter two groups was not significant.

1. Total proportion glimpsed

Figure 7 shows performance in all conditions for both groups of listeners. The substantial improvement in performance as the total proportion of speech is increased from 0.33 to 0.67 can be seen for both groups and in all contexts. Overall, performance was approximately 38 percentage points and 40 RAU better with the larger proportion of available speech. The effect of glimpse proportion increased with additional context, increasing from 28 percentage points (30 RAU) with isolated words, to 38 percentage points (38 RAU) for LP sentences, and to 47 percentage points (53 RAU) for HP sentences. This was reflected in the significant glimpse proportion by context interaction ($F(2,66) = 136.15, p < 0.001$).

A significant interaction between listener group and the glimpsed proportion of speech ($F(2,33) = 18.08, p < 0.001$) reflected a somewhat greater performance difference between the two speech proportions (by roughly 10 percentage points) for the ENH listeners compared to the other two groups of listeners. This was primarily due to poorer performance at the smaller speech proportion for the ENH group. A significant three-way interaction between group, proportion, and context ($F(4,66) = 5.09, p < 0.005$) was due primarily to group differences in the amount of improvement across context conditions as the proportion of available speech increased. The ENH group showed greater relative improvements for the sentence conditions (especially the HP sentences) compared to the other two listener groups when the proportion of glimpsed speech was increased.

2. Effect of context

The addition of the sentence context clearly facilitated target word recognition for all groups of listeners at both proportions of available speech. *Post hoc* comparisons revealed significant differences between all context conditions (Tukey

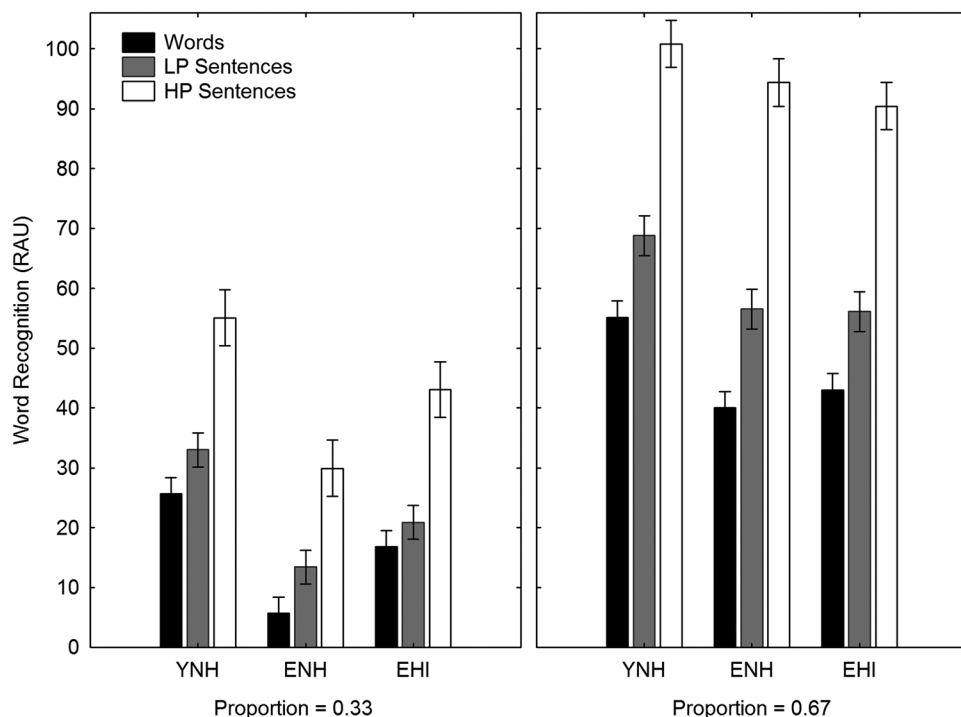


FIG. 7. Word recognition performance (in RAU) in each condition of Experiment 2. Performance in each of the three context conditions (Words, LP sentences, HP sentences) are shown for each listener group at the smaller glimpse proportion (left panel) and at the larger glimpse proportion (right panel).

HSD, $p < 0.05$). Performance with the LP sentences was somewhat better than with isolated words, averaging 10 percentage points (10 RAU) across groups and proportions, and the HP sentences resulted in even greater improvements, averaging 24 percentage points (27 RAU) across groups and proportions (relative to performance in the LP context). Not surprisingly, context effects were greater at the larger glimpsed proportion when more of the context speech was available, as indicated by a significant context by proportion interaction (discussed above). The context effect was about 9 percentage points (11 RAU) greater at the higher proportion. (Note that the near-ceiling performance of the normal-hearing subjects at the higher proportion suggests that the benefit of HP sentences may be somewhat underestimated for the normal-hearing groups in this condition.) There was no interaction between listener group and context ($F(4,66) = 0.56$, $p > 0.05$) and the three-way interaction (as noted above) was associated with differential improvement for the ENH listeners across the three context conditions with the increase in the glimpsed proportion.

3. Correlations across conditions

Examination of the relative performance of subjects across conditions revealed results similar to those in Experiment 1. The average correlation between conditions in Experiment 2 for the YNH group was 0.71 (range; 0.54–0.85), with all but the lowest correlation (comparing isolated words at the larger glimpsed proportion with HP sentences at the smaller glimpse proportion) significant ($p < 0.05$). The one relatively low correlation was due to two subjects performing unusually well with the HP sentences, relative to their performance with isolated words. For the older group of listeners in Experiment 2 (ENH and EHI combined), the average correlation was 0.82 (range: 0.58–0.92), with all correlations significant ($p < 0.01$).

4. Correlations with hearing sensitivity

As in Experiment 1, correlations between age and word recognition scores (in RAU) were fairly low, although the correlation obtained in Experiment 2 ($r = -0.34$) was slightly higher than in Experiment 1 ($r = -0.21$). The magnitude of all correlations between word recognition and hearing loss (both PTA and HFPTA) was less than 0.1 in both experiments. Thus, hearing loss does not appear to influence the ability to make use of partial information in audible speech while aging has a small negative effect that is significant only when comparing younger (less than 30 yrs) to older (greater than 60 yrs) listeners, but not within the 63–90 year age range of elderly subjects in the present study.

III. GENERAL DISCUSSION

These experiments show that the proportion of speech available to a listener is a strong predictor of the recognition of words in interrupted speech. Young and older normal-hearing listeners, as well as older listeners with hearing loss (with amplification to ensure audibility), show large improvements in word recognition as the available proportion of speech increases from 0.33 to 0.67 of the total speech

duration. This is true for single interrupted words in quiet (Experiment 1) and for interrupted words and sentences in noise (Experiment 2). The pattern of results in Experiment 1 is similar to that obtained in an earlier experiment (Wang and Humes, 2010) with a different set of stimuli and a different selection of interruption parameters. Together, these experiments show that the proportion of available speech is generally a better predictor of word recognition with interrupted speech than other interruption parameters (such as number of glimpses, glimpse duration, and interruption rate).

Although the finding that the perception of glimpsed speech depends strongly on the proportion of available speech is not new, these experiments provide new information about the magnitude of the proportional duration effect, the consistency of the effect across listener groups, and the relative lack of effects of other interruption parameters. For relatively long glimpse and inter-glimpse durations, as discussed below, and perhaps for slower interruption rates than those examined in the present study, the timing and location of glimpses can have a greater impact on performance. Furthermore, averaging across glimpse parameters and the various placements of glimpses within each of the target words when examining the effect of proportional duration (as done here) likely masks some significant effects of glimpse duration and placement, such as those observed when glimpsing vowels vs. consonants (e.g., Fogerty and Kewley-Port, 2009; Kewley-Port *et al.*, 2007). Thus, the observed proportional duration effect does not indicate that information is uniformly distributed within words and sentences; it merely indicates that when glimpse placement is not systematically controlled in temporally distributed glimpsing patterns, performance is largely determined by the available proportion of speech, with little influence of glimpse duration or rate over a wide range.

In Experiment 1, the number of glimpses also had an effect on word recognition performance, but the effect was relatively small and differed for the three listener groups. In Experiment 1, performance tended to worsen with increases in the number of glimpses for the YNH and ENH groups, but for the EHI group, performance was worse with four glimpses than with eight or twelve glimpses. Wang and Humes (2010) found that performance tended to improve with increases in the number of glimpses at the lowest proportion of available speech (0.25), but the opposite trend was observed (weakly) at the largest proportion of available speech (0.75). Because Wang and Humes (2010) held word duration constant (at 512 ms), the number of glimpses was confounded with glimpse duration. This was not true in the current study, in which target word duration varied naturally over a 2:1 ratio (from 300 to 600 ms). Thus, a given number of glimpses in Experiment 1 included many glimpse durations that were shorter (and some that were longer) than those in Wang and Humes (2010) at a similar proportion of available speech. This might account, in part, for the differences in the influence of the number of glimpses in the two studies. That is, the decrease in performance with fewer glimpses observed at the smaller proportions of speech in Wang and Humes (2010) may have been partly due to the associated increase in the duration of each glimpse. Because of the way interruption patterns are manipulated in these

studies, when glimpse durations increase, the pauses between glimpses also increase and listeners are required to “fill in” longer stretches of missing speech. This can lead to worse recognition performance, as has been seen in other studies using longer, sparsely distributed glimpses (e.g., Li and Loizou, 2007). It is not clear why this effect of fewer glimpses was only seen with EHI listeners. However, it appears that while normal-hearing listeners tend to do slightly better when a given proportion of available speech is made available in fewer glimpses (within the present range of glimpse parameters), EHI listeners are more likely to have difficulty combining information across glimpses when longer glimpses are separated by relatively long silences. Because EHI listeners performed about as well as YNH listeners with 8 and 12 glimpses, but close to the poorer performance of ENH listeners with only 4 glimpses, it may be that the EHI listeners have developed listening strategies (due to listening with impaired hearing) that work well only when glimpses and silent gaps are shorter and better distributed over time.

Except for the difference noted above, both experiments reported here generally show that older listeners, even with hearing impairment (with audibility restored), recognize interrupted speech in essentially the same way as young normal-hearing listeners. Although the older listeners tend to perform at somewhat lower levels, they are affected by the proportion of speech and by context in the same way as the younger listeners. The largest differences between groups were seen when the smaller proportion of speech was available. In Experiment 2, ENH subjects performed worse than EHI subjects in this condition but performed as well as EHI listeners with the larger proportion of available speech. A similar, but weaker, tendency was seen in Experiment 1. It is not clear whether the performance differences observed when less speech is available are due to the EHI listeners having greater experience listening to partly inaudible speech, or to somewhat lower sensation levels for some ENH listeners (due to slightly elevated thresholds and no corrective amplification). However, the rather small disadvantage that some ENH listeners had at higher frequencies suggests that this performance difference is not entirely accounted for by differences in sensation level.

A. Effect of context

The consistent effect of context in Experiment 2 shows that the presence of a sentence context has a substantial beneficial effect on the recognition of an interrupted word at the end of an interrupted sentence. Despite very low levels of isolated word recognition (especially for the ENH listeners at the smaller proportion of glimpsed speech) the context of a low predictability sentence leads to significant improvements in word recognition, and a high-predictability sentence leads to relatively large improvements. Part of the facilitation associated with the LP context (apart from that due to prosodic or coarticulatory information) may be due to the greater predictability of the target-word glimpse pattern when it is presented throughout the preceding context. Listeners may be able to get more information from each glimpse when they can anticipate the timing of the glimpses. But the additional

facilitation with the HP context, even in the most difficult conditions, shows that listeners are recognizing enough of the words to take advantage of the predictability of the target words. Thus, even under interruption conditions that might be expected to provide less context information (based on the poor recognition of isolated words in those conditions), the context effect is robust. This indicates that the combination of information across glimpses and the filling in of missing information can occur even when information for individual words is quite limited. The finding that all groups show similar context facilitation despite differences in overall performance indicates that the ability to use context to facilitate word recognition is not diminished by aging or hearing loss, and that this ability can be used effectively even under conditions that lead to poor recognition of individual words.

In some previous investigations with the SPIN materials, older subjects were found to benefit more from the context provided by HP sentences than were younger listeners (e.g., Pichora-Fuller *et al.*, 1995; Sheldon *et al.*, 2008). Whether difficulty was manipulated by varying the SNR (Pichora-Fuller *et al.*, 1995) or by the number of vocoder bands (Sheldon *et al.*, 2008), older listeners tended to show a greater HP – LP performance difference than did younger listeners at intermediate levels of difficulty (i.e., when performance with LP and HP sentences were significantly above the floor and below the ceiling). This HP – LP advantage was due to a greater difference between younger and older listeners with LP sentences than with HP sentences at a given SNR, with older listeners performing worse in both conditions. The lack of a significant interaction between context and listener group in the present study suggests that the underlying psychometric functions for high and low context conditions may have more similar shapes for young and old listeners with interrupted speech than in the cited studies using noise-masked or vocoded speech.

In a summary of several studies using the SPIN materials with babble or noise background, Humes *et al.* (2007) found that average HP-LP differences were generally similar for YNH and elderly listeners. Difference scores were generally around 40 percentage points (with a range from about 35 to 50 points) under conditions in which performance with LP sentences was between 20% and 60%. In the present study, performance with LP sentences was generally in this same range (with means a bit outside of this range in two of the six conditions), but the size of the context effect (LP vs HP) was consistently lower than 40 percentage points (range = 15 to 33 points). Table I shows the context gains (both HP-LP and LP-Words), measured in percentage point differences, in all conditions of Experiment 2. The older listeners did tend to have somewhat greater context gains than the YNH group at the larger glimpse proportion for HP-LP (primarily due to lower LP scores), but not for LP-Words, and not at the smaller glimpse proportion. In some cases, the relatively low context gains in this study may be due in part to the selection of glimpse conditions that put LP performance at levels closer to the upper and lower extremes of the psychometric functions, where gains tend to be lower. However, gains were relatively low even for comparisons with mid-range performance levels. It may be that glimpsing

TABLE I. Context gains in Experiment 2 (differences in percentage points).

	33% glimpsed		67% glimpsed		Overall	
	LP-Words	HP-LP	LP-Words	HP-LP	LP-Words	HP-LP
YNH	0.07	0.23	0.14	0.24	0.11	0.24
ENH	0.06	0.15	0.17	0.33	0.11	0.24
EHI	0.03	0.22	0.14	0.29	0.08	0.26

during the sentence leading up to the final target word somewhat reduces the context information available to the listener relative to other types of stimulus degradation.

Overall, it appears that older listeners tend to have more difficulty than younger listeners in making use of reduced speech information whether it is due to noise masking, babble masking, reduction of fine structure (vocoded speech), or interruption (as in the present study). The addition of sentence context, even in the absence of useful semantic

predictability, is helpful to both younger and older listeners and can sometimes reduce the older listeners' disadvantage.

Comparison of performance levels in the current study with those obtained by Dirks *et al.* (1986), who reported performance for LP and HP sentences in the SPIN babble as a function of the articulation index (AI; ANSI S3.5-1969, 1/3-octave band method) reveals a good correspondence for the relative values of LP and HP scores. Percent-correct scores in the present study with a total glimpsed proportion of 0.33 are roughly equal to scores with an AI of 0.2 – 0.25, while scores in the 0.67 glimpse condition are similar to those observed with an AI in the 0.4 – 0.5 range. Although the specific AI values may not be particularly meaningful, the comparison suggests similar performance gains as a function of both AI and the proportion of speech glimpsed with the SPIN materials. (More recent work with the speech intelligibility index, modified to account for speech recognition in fluctuating noise (Rhebergen and Versfeld, 2005; Rhebergen *et al.*, 2006), may provide more useful comparisons.) This, along

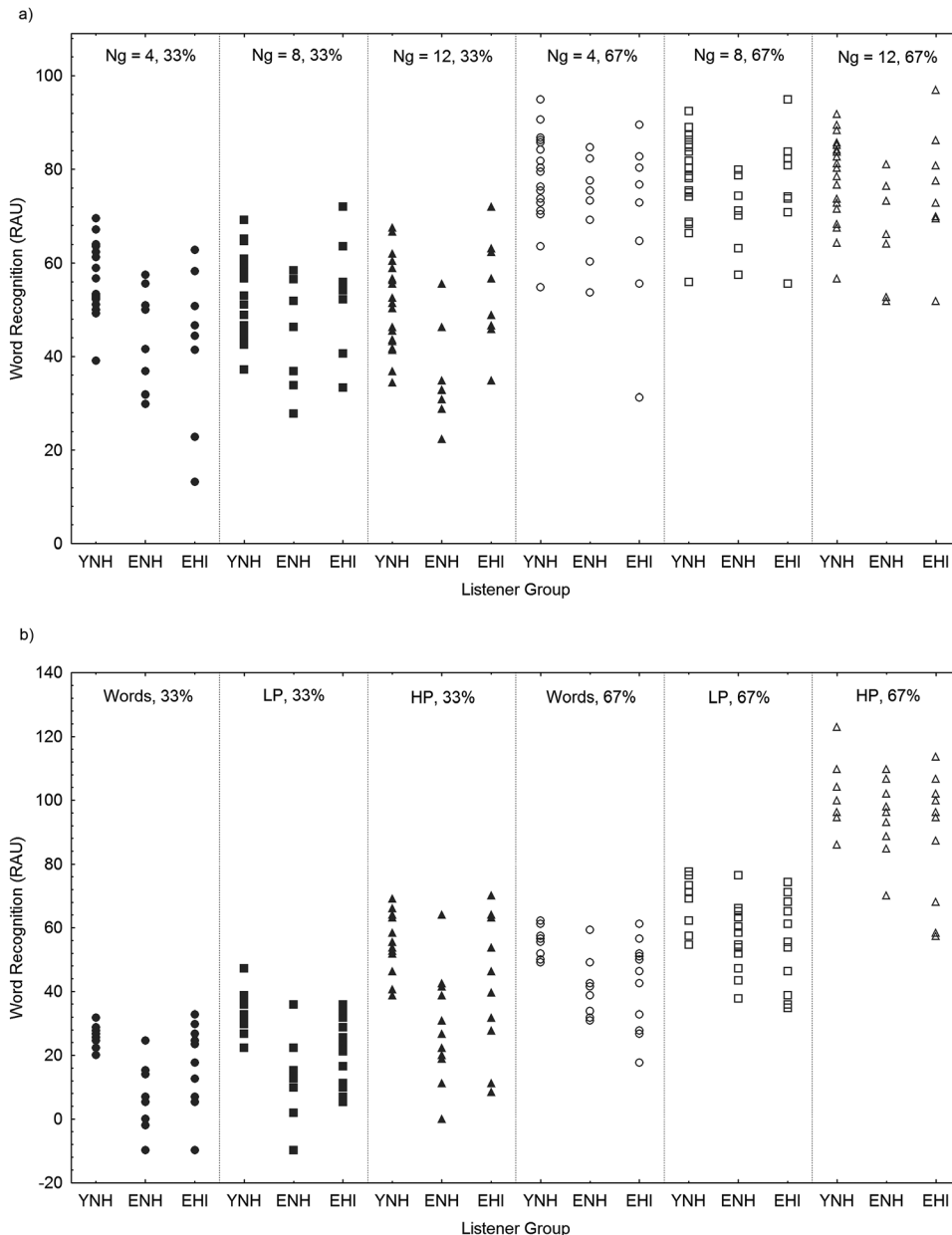


FIG. 8. Individual differences in word recognition performance (in RAU) in Experiment 1 (a) and Experiment 2 (b). Each plot is divided into six sections, one for each experimental condition, in which data points for all subjects in each listener group are shown (Ng = number of glimpses; percentages indicate the percent of target words glimpsed; Words, LP, and HP indicate the context conditions). Considerable overlap in performance across listener groups is seen in all conditions and within-group individual differences are generally much greater than the differences between group means.

with the similarities noted above with noise-masked, babble-masked, and vocoded SPIN materials, indicates that the perception of interrupted speech is generally similar to the perception of other forms of disrupted speech that reduce the available speech information. These similarities suggest a common underlying ability to fill in missing speech information (i.e., contextual inference) when only partial information is available. Although an independent attentional-focusing ability may also play a role, the similarity of the pattern of data across comparable conditions using continuous noise masking vs glimpsed speech (in quiet or with a favorable SNR) does not provide any support for the hypothesis of a greater reliance on an independent attentional focusing ability when speech is mixed with noise than when portions of speech are replaced with silence. Research with a larger number of subjects performing a variety speech and non-speech tasks with different types of disruption or interference will be needed to provide more information about the different underlying abilities that are used to understand speech under difficult listening conditions.

B. Individual differences

Although the number of subjects in this study is too small to provide a clear indication of the range of abilities on this task in the general population, the range of performance within each of the groups deserves some consideration. All groups showed a wide range of performance in the recognition of interrupted words in isolation and in sentence context. In both experiments, the difference between subjects within each group was substantially larger than the difference between group means. There was considerable overlap between groups despite few outliers in any of the groups. In Experiment 1, the average range in word recognition scores across the six conditions was 34 percentage points (35 RAU) for the YNH group and 29 percentage points (29 RAU) for the ENH group, while the EHI group had an average range of 44 percentage points (45 RAU). The EHI group also had the greatest average range in Experiment 2 (40 percentage points; 46 RAU), while the YNH and ENH groups had ranges of 20 percentage points (23 RAU) and 35 percentage points (42 RAU), respectively. Although the EHI group included some of the poorest performing listeners, other listeners in this group performed as well as or better than the best listeners in the YNH group. Figure 8 shows the scores, in RAU, for each subject in each condition of Experiment 1 [Fig. 8(a)] and Experiment 2 [Fig. 8(b)]. It is clear from these data (and from the correlational data reported above) that the ability to make use of partial information in interrupted speech varies widely, with only a small portion of the variation in that ability accounted for by age or hearing loss. YNH do tend to perform better than ENH and EHI listeners, but many elderly listeners are as good as or better than the average YNH listener at understanding interrupted speech.

IV. CONCLUSIONS

- (1) When listening to interrupted speech, the proportion of an utterance that is available (or “glimpsed”) is the major

determinant of speech recognition performance. Changes in glimpse duration, glimpse rate, and the number of glimpses have relatively little effect on performance, except at extreme values.

- (2) Older listeners, with or without hearing impairment, tend to have somewhat greater difficulty recognizing interrupted speech than younger listeners, but many older listeners perform as well as or better than younger listeners. Older listeners are affected by changes in the proportion of glimpsed speech, and other interruption parameters, in the same way as younger listeners.
- (3) The beneficial effect of adding a sentence context is essentially the same for YNH, ENH, and EHI listeners. An interrupted word that appears at the end of an interrupted sentence is recognized significantly better than the same interrupted word presented in isolation. This is true even for sentences that provide little or no information about the identity of the target word (i.e., the LP sentences from the SPIN test) and even under interruption conditions that render most words unintelligible when presented in isolation.
- (4) The facilitation of word recognition performance by the addition of predictability in interrupted sentences (comparing LP and HP sentences) is similar to that observed with other methods of reducing the available speech information (such as noise masking, babble masking, and vocoding). The perception of interrupted speech appears to have much in common with the perception of other forms of degraded speech.

ACKNOWLEDGMENTS

This work was supported by research Grant No. R01 AG008293 from the National Institute on Aging (NIA) awarded to the second author. We thank graduate research assistants Sara Brown, Tera Quigley and Hannah Fehlberg for their assistance with data collection and reduction.

- ANSI (1969). S3.5-1969, *American National Standard Method for the Calculation of the Articulation Index* (American National Standards Institute, New York).
- ANSI (1999). S3.1-1999, *Maximum Permissible Ambient Levels for Audiometric Test Rooms* (American National Standards Institute, New York).
- ANSI (2004). S3.6-2004, *Specification for Audiometers* (American National Standards Institute, New York).
- Akeroyd, M. A. (2008). “Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults,” *Int. J. Audiol.* **47**, 53–71.
- Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (1984). “Standardization of a test of speech perception in noise,” *J. Speech Hear. Res.* **27**(1), 32–48.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).
- Cooke, M. (2003). “Glimpsing speech,” *J. Phonetics*, **31**(3/4), 579–584.
- Cooke, M. (2006). “A glimpsing model of speech perception in noise,” *J. Acoust. Soc. Am.* **119**(3), 1562–1573.
- Cooke, M., Garcia Lecumberri, M. L., and Barker, J. (2008). “The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception,” *J. Acoust. Soc. Am.* **123**(1), 414–427.
- Cusack, R., and Carlyon, R. P. (2004). “Auditory perceptual organization inside and outside the laboratory,” in *Ecological Psychoacoustics*, edited by J. G. Neuhoff (Elsevier, New York), pp. 15–18.

- Darowski, E. S., Helder, E., Zacks, R. T., Hasher, L., and Hambrick, D. Z. (2008). "Age-related differences in cognition: the role of distraction control," *Neuropsychology* 22(5), 638–644.
- Dirks, D. D., Bell, T. S., Rossman, R. N., and Kincaid, G. E. (1986) "Articulation index predictions of contextually dependent words," *J. Acoust. Soc. Am.* 80(1), 82–92.
- 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(4), 2084–2094.
- Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003). "Note on informational masking," *J. Acoust. Soc. Am.* 113(6), 2984–2987.
- Elliott, L. L. (1995), "Verbal auditory closure and the speech perception in noise (SPIN) Test," *J. Speech Hear. Res.* 38(6), 1363–1376.
- Fletcher, H. (1922). "The nature of speech and its interpretation," *J. Franklin Inst.* 193, 729–747.
- Fogerty, D., and Kewley-Port, D. (2009). "Perceptual contributions of the consonant-vowel boundary to sentence intelligibility," *J. Acoust. Soc. Am.* 126(2), 847–857.
- Folstein, M. F., Folstein, S. E., and McHugh, P. R. "Mini-mental state. A practical method for grading the cognitive state of patients for the clinician," *J. Psychiatr. Res.* 12(3), 189–198.
- French, N. R., and Steinberg, J. C. (1947) "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Am.* 19, 19–119.
- George, E. L., 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(4), 2295–2311.
- George, E. L., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., and Houtgast, T. (2007). "Auditory and nonauditory factors affecting speech reception in noise by older listeners." *J. Acoust. Soc. Am.* 121(4), 2362–2375.
- 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(2), 1000–1011.
- Hasher, L., and Zacks, R. T. (1988). "Working memory, comprehension, and aging: A review and a new view," *The Psychology of Learning and Motivation: Advances in Research and Theory*, 22, 193–225.
- Huggins, A. W. (1975). "Temporally segmented speech," *Percept. Psychophys.* 18(2), 149–157.
- Humes, L. E., Burk, M. H., Coughlin, M. P., Busey, T. A., and Strauser, L. E. (2007). "Auditory speech recognition and visual text recognition in younger and older adults: similarities and differences between modalities and the effects of presentation rate," *J. Speech Lang. Hear. Res.* 50(2), 283–303.
- Humes, L. E., and Dubno, J. R. (2010). "Factors affecting speech understanding in older adults. In S. Gordon-Salant, R. D. Frisina," in *The Aging Auditory System: Perceptual Characterization and Neural Bases of Presbycusis*, edited by A. N. Popper and R. R. Fay (Springer, New York), pp. 211–257.
- Iyer, N., Brungart, D. S., and Simpson, B. D. (2007). "Effects of periodic masker interruption on the intelligibility of interrupted speech," *J. Acoust. Soc. Am.* 122(3), 1693–1701.
- Jin, S. H., and Nelson, P. B. (2010). "Interrupted speech perception: the effects of hearing sensitivity and frequency resolution," *J. Acoust. Soc. Am.* 128(2), 881–889.
- Jones, M. R., and Boltz, M. (1989). "Dynamic attending and responses to time," *Psychol. Rev.* 96, 459–491.
- Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (1977). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *J. Acoust. Soc. Am.* 61(5), 1337–1351.
- Kawahara, H., Masuda-Katsuse, I., and de Cheveign, A. (1999). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," *Speech Commun.* 27, 187–207.
- Kewley-Port, D., Burk, T. Z., and Lee, J. H. (2007). "Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners," *J. Acoust. Soc. Am.* 122(4), 2365–2375.
- Kidd, G., Jr., Mason, C. R., and Arbogast, T. L. (2002). "Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns," *J. Acoust. Soc. Am.* 111(3), 1367–1376.
- Kidd, G. R., Watson, C. S., and Gygi, B. (2007). "Individual differences in auditory abilities," *J. Acoust. Soc. Am.* 122(1), 418–435.
- Lee, J. H., and Kewley-Port, D. (2009). "Intelligibility of interrupted sentences at subsegmental levels in young normal-hearing and elderly hearing-impaired listeners," *J. Acoust. Soc. Am.* 125(2), 1153–1163.
- Li, N., and Loizou, P. C. (2007). "Factors influencing glimpsing of speech in noise," *J. Acoust. Soc. Am.* 122(2), 1165–1172.
- Lutfi, R. A. (1993). "A model of auditory pattern analysis based on component-relative-entropy," *J. Acoust. Soc. Am.* 94(2 Pt. 1), 748–758.
- Mattys, S. L., Brooks, J., and Cooke, M. (2009). "Recognizing speech under a processing load: Dissociating energetic from informational factors," *Cogn. Psychol.* 59(3), 203–243.
- McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., and Wingfield, A. (2005). "Hearing loss and perceptual effort: downstream effects on older adults' memory for speech," *Q. J. Exp. Psychol. A* 58(1), 22–33.
- Miller, G. A., and Licklider, J. C. R. (1950). "The intelligibility of interrupted speech," *J. Acoust. Soc. Am.* 22, 167–173.
- Moore, B. C. J. (2003). "Temporal integration and context effects in hearing," *J. Phonetics* 31(3/4), 563–574.
- 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(5), 2286–2294.
- Nittrouer, S., and Boothroyd, A. (1990). "Context effects in phoneme and word recognition by young children and older adults," *J. Acoust. Soc. Am.* 87(6), 2705–2715.
- Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (1995) "How young and old adults listen to and remember speech in noise," *J. Acoust. Soc. Am.* 97(1), 593–608.
- Pollack, I., and Pickett, J. M. (1958). "Masking of speech by noise at high sound levels," *J. Acoust. Soc. Am.* 30, 127–130.
- Powers, G. L., and Wilcox, J. C. (1977). "Intelligibility of temporally interrupted speech with and without intervening noise," *J. Acoust. Soc. Am.* 61(1), 195–199.
- Rhebergen, K. S., and Versfeld, N. J. (2005). "A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," *J. Acoust. Soc. Am.* 117(4), 2181–2192.
- Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (2006). "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," *J. Acoust. Soc. Am.* 120(6), 3988–3997.
- Salhouse, T. A., and Madden, D. J. (2008). "Information processing speed and aging," in *Information Processing Speed in Clinical Populations*, edited by J. K. DeLuca and Jessica H. (Taylor & Francis, London), pp. 221–241.
- Sheldon, S., Pichora-Fuller, M. K., and Schneider, B. A. (2008). "Priming and sentence context support listening to noise-vocoded speech by younger and older adults," *J. Acoust. Soc. Am.* 123(1), 489–499.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* 28(3), 455–462.
- Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (1999). "Monosyllabic word recognition at higher-than-normal speech and noise levels," *J. Acoust. Soc. Am.* 105(4), 2431–2444.
- Summerfield, A. Q., Culling, J. F., and Assmann, P. F. (2006). "The Perception of Speech Under Adverse Conditions: Contributions of Spectro-Temporal Peaks, Periodicity, and Interaural Timing to Perceptual Robustness," in *Listening to Speech: An Auditory Perspective*, edited by S. Greenberg and W. A. Ainsworth (Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey), pp. 223–235.
- Surprenant, A. M., and Watson, C. S. (2001). "Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners," *J. Acoust. Soc. Am.* 110(4), 2085–2095.
- Van Rooij, J. C., Plomp, R., and Orlebeke, J. F. (1989). "Auditive and cognitive factors in speech perception by elderly listeners: I. Development of test battery," *J. Acoust. Soc. Am.* 86(4), 1294–1309.
- Vestergaard, M. D., Fyson, N. R., and Patterson, R. D. (2011). "The mutual roles of temporal glimpsing and vocal characteristics in cocktail-party listening," *J. Acoust. Soc. Am.* 130(1), 429–439.
- Wang, X., and Humes, L. E. (2010). "Factors influencing recognition of interrupted speech," *J. Acoust. Soc. Am.* 128(4), 2100–2111.
- Watson, C. S. (1987). "Uncertainty, informational masking and the capacity of immediate auditory memory," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum Associates, Hillsdale, NJ), pp. 267–277.
- Watson, C. S. (2005). "Some comments on informational masking," *Acta Acust. Acust.* 91(3), 502–512 (2005).
- Wilson, R. H., McArdle, R., Betancourt, M. B., Herring, K., Lipton, T., and Chisolm, T. H. (2010). "Word-Recognition Performance in Interrupted Noise by Young Listeners with Normal Hearing and Older Listeners with Hearing Loss," *J. Am. Acad. Audiol.* 21(2), 90–109.