

Factors influencing recognition of interrupted speech

Xin Wang^{a)} and Larry E. Humes

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

(Received 24 November 2009; revised 29 July 2010; accepted 2 August 2010)

This study examined the effect of interruption parameters (e.g., interruption rate, on-duration and proportion), linguistic factors, and other general factors, on the recognition of interrupted consonant-vowel-consonant (CVC) words in quiet. Sixty-two young adults with normal-hearing were randomly assigned to one of three test groups, “male65,” “female65” and “male85,” that differed in talker (male/female) and presentation level (65/85 dB SPL), with about 20 subjects per group. A total of 13 stimulus conditions, representing different interruption patterns within the words (i.e., various combinations of three interruption parameters), in combination with two values (easy and hard) of lexical difficulty were examined (i.e., $13 \times 2 = 26$ test conditions) within each group. Results showed that, overall, the proportion of speech and lexical difficulty had major effects on the integration and recognition of interrupted CVC words, while the other variables had small effects. Interactions between interruption parameters and linguistic factors were observed: to reach the same degree of word-recognition performance, less acoustic information was required for lexically easy words than hard words. Implications of the findings of the current study for models of the temporal integration of speech are discussed.

© 2010 Acoustical Society of America. [DOI: 10.1121/1.3483733]

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

Pages: 2100–2111

I. INTRODUCTION

In daily life, speech communication frequently takes place in adverse listening environments (e.g., in a background of noise or with competing speech), which often causes discontinuities of the spectro-temporal information in the target speech signal. Regardless, recognition of such interrupted speech is often robust, not only because speech is inherently redundant (so listeners can afford the partial loss of information), but also because listeners can integrate fragments of contaminated or distorted speech over time and across frequency to restore the meaning. Various “multiple looks” models of temporal integration have been proposed, initially for temporal integration of signal energy at threshold (Viemeister and Wakefield, 1991) and, more recently, for the temporal and spectral integration of speech fragments or “glimpses” (e.g., Moore, 2003; Cooke, 2003). With regard to the temporal integration of speech fragments, most of the data modeled were for speech in a fluctuating background, such as competing speech; a background that would create a distribution of spectro-temporal speech fragments. However, the underlying mechanism of temporal integration for speech is still not fully understood, even for the simplest conditions: integration of the interrupted waveform of otherwise undistorted broad-band speech in quiet. This simpler case is the focus of this study.

In the context of speech, a “look” or “glimpse” is defined as “an arbitrary time-frequency region which contains a reasonably undistorted view of the target signal” (Cooke, 2003). The “looks” are sampled at the auditory periphery, stored, and processed “intelligently” in the form of spectral-

temporal excitation pattern (“STEP”) in the central auditory system (Moore, 2003). However, the theory of “multiple looks” needs to be further tested for speech perception with other factors that could potentially affect the integration process taken into account. First, at the auditory periphery, is it true that simply more “looks” produce better perception given the redundant nature of speech? Do characteristics of the “looks,” such as duration or rate (interruption frequency), affect the ultimate perception? Do duration and rate interact? Second, how do linguistic factors contribute to the “intelligent integration”? How do these linguistic factors interact with the interruption parameters to affect speech recognition?

Previous research on interrupted speech in quiet provides partial answers to these questions (e.g., Miller and Licklider, 1950; Dirks and Bower, 1970; Powers and Wilcox, 1977; Nelson and Jin, 2004). By arbitrarily turning the speech signal on and off multiple times over a given time period, or by multiplying speech by a square wave, these early researchers examined the effects of various interruption parameters, such as the interruption duty cycle (i.e., proportion of a given cycle during which speech is presented) and the frequency of interruption, on listeners’ ability to understand interrupted speech. Speech recognition performance was suggested to improve as the total proportion of the presented speech signal increased (e.g., Miller and Licklider, 1950). For instance, Miller and Licklider (1950) demonstrated a maximum improvement of 70% in word recognition scores when the duty cycle increased from 25% to 75%. In general, when the proportion of speech was $\geq 50\%$, speech became reasonably intelligible, regardless of the type of the speech material used (e.g., Miller and Licklider, 1950; Dirks and Bower, 1970; Powers and Wilcox, 1977; Nelson and Jin, 2004). This general finding seems to support the basic notion

^{a)}Author to whom correspondence should be addressed. Electronic mail: xw4@indiana.edu

of the “multiple looks” theory: more “looks” lead to better speech perception. However, most of the prior research kept the duty cycle fixed at 50% and varied the interruption frequency. In these cases, the duration of the glimpses co-varied with interruption frequency.

The effects of frequency of interruption on performance have yielded mixed results. For instance, using sentences interrupted by silence, Powers and Wilcox (1977) and Nelson and Jin (2004) concluded that, as interruption frequency increased, listeners’ performance increased. Miller and Licklider (1950), however, showed a more complex relationship between word recognition scores and interruption frequency. The inconsistent conclusions might result from the fact that Powers and Wilcox (1977) and Nelson and Jin (2004), as was typically the case, investigated the effect of a few slower interruption frequencies with only one duty cycle (e.g., 50%) while Miller and Licklider (1950) investigated multiple duty cycles and a relatively broader range of interruption frequencies. Although the pioneering study by Miller and Licklider (1950) provided results for a more extensive range of interruption parameters than most subsequent studies, the data are somewhat limited due to the use of just one talker and one presentation level, as well as a small sample of listeners. Recently, data from Li and Loizou (2007) showed that interrupted sentences with short-duration “looks” were more intelligible than the ones with looks of longer duration. However, their conclusion is based on a speech proportion (i.e., 0.33) that was lower than most of the earlier studies of interrupted speech. Therefore, given the few systematic studies of the effect of interruption frequency and the duration of each “look” on speech perception, especially for a sufficient sample size of listeners, such an investigation was the focus of this study.

Moreover, to confirm that the present results were not confined to one speech level or to one talker, additional presentation levels and talkers were sampled in this study. In most prior studies of interrupted speech, only one talker at a single presentation level, typically a moderate level of 60–70 dB SPL, was investigated. Here, the speech levels used were 65 and 85 dB SPL. The former was selected for comparability to most other studies and is representative of conversational level. The higher level was chosen in anticipation of eventual testing of older adults, many of whom will have hearing loss and will require higher presentation levels as a result. Also, it has been widely observed that there is a decrease in intelligibility when speech levels are increased above 80 dB SPL (e.g., Fletcher, 1922; French and Steinberg, 1947; Pollack and Pickett, 1958). Various causes have been proposed, but none is certain (e.g., Studebaker *et al.*, 1999; Liu, 2008) and their application to interrupted speech requires direct examination. With regard to talkers, only two were examined here, but two considerably different talkers: a male with a fundamental frequency of 110 Hz and a female with a fundamental frequency of 230 Hz. The nearly 2:1 ratio of fundamental frequencies, combined with several twofold changes in glimpse duration, also allowed for informal examination of the role played by the number of pitch pulses in the integration of interrupted speech. If talker fundamental frequency is the perceptual “glue” that holds to-

gether the speech fragments, the number of pitch pulses in a glimpse may prove to be a relevant acoustical factor.

Previous research has also shown that linguistic factors, such as context, syntax, intonation and prosody, can impact the temporal integration of speech fragments (e.g., Verschuur and Brocaar, 1983; Bashford and Warren, 1987; Bashford *et al.*, 1992). It is unclear, however, whether the benefit from linguistic factors differs for various combinations of interruption parameters. For example, will the benefit from linguistic factors differ as the proportion of the intact speech varies? There could potentially be interactions between the acoustic interruption parameters and top-down linguistic factors in interrupted speech based on the models of speech perception (e.g., Marslen-Wilson and Welsh, 1978; Gareth Gaskell William and Marslen-Wilson, 1997; Luce and Pisoni, 1998), yet these interactions have never been quantified.

The present study was carried out to investigate how different factors influence the temporal integration of otherwise undistorted speech by systematically removing portions of consonant-vowel-consonant (CVC) words in quiet listening conditions. Specifically, three interruption-related properties, interruption rate, on-duration, and proportion of speech presented, were varied systematically. Special lists of CVC words (Luce and Pisoni, 1998; Dirks *et al.*, 2001, Takayanagi *et al.*, 2002) were used to examine the effect of lexical properties (i.e., word frequency, neighborhood density and neighborhood frequency) on temporal integration. Talkers with different gender as well as different presentation levels will also be investigated to address more general questions related to word recognition. These variables are not being explored as additional independent variables in a systematic fashion. Rather, a range of speech levels and talker fundamental frequencies were sampled at somewhat extreme values in each case to explore the impact of these factors on the primary variables of interest: interruption-related factors and linguistic difficulty. We hypothesized that all three interruption-related factors would have an influence on the temporal integration of otherwise undistorted speech in quiet. We also expected to see further impact of lexical factors on the temporal integration process. It was further hypothesized that general factors that affect word recognition, such as talker gender and presentation level, will also have effects on interrupted words: the female talker will be more intelligible while the higher presentation level will reduce performance. However, we hypothesized that the relative effects of interruption parameters and linguistic difficulty would be similar across the range of speech levels and fundamental frequencies sampled.

II. METHODS

A. Subjects

A total of 62 young native English speakers with normal hearing participated in the study. The subjects included 30 men and 32 women who ranged in age from 18 to 30 years (mean=24 years). All participants had air-conduction thresholds ≤ 15 dB HL (ANSI, 1996) from 250 through 8000 Hz in both ears. They had no history of hearing loss or

recent middle-ear pathology and showed normal tympanometry in both ears. All subjects were recruited from the Indiana University campus community and were paid for their participation.

B. Stimuli and apparatus

1. Materials

The speech materials consisted of 300 digitally interrupted CVC words, which included two sets of 150 words produced by either a male or a female talker. The original digital recordings from Takayanagi *et al.* (2002) were obtained for use in this study. Each set of words included 75 lexically easy words (words with high word frequency, low neighborhood density, and low neighborhood frequency) and 75 lexically hard words (words with low word frequency, high neighborhood density, and high neighborhood frequency) according to the Neighborhood Activation Model (NAM) (Luce and Pisoni, 1998; Dirks *et al.*, 2001, Takayanagi *et al.*, 2002). The two talkers had roughly a one-octave difference in fundamental frequency (F_0) (F_0 of the male voice = 110 Hz; F_0 of the female voice = 230 Hz) and similar average durations (around 530 ms) for the CVC tokens they produced. Another 50 CVC words from a different lexical category (words with high word frequency, high neighborhood density, and high neighborhood frequency) recorded from a different male talker by Takayanagi *et al.* (2002) were used for practice.

2. Equalization

All 300 CVC words were edited using Adobe Audition (V 2.0) to remove periods of silence >2 ms at the beginning and end of the CVC stimulus files. Next, the words were either lengthened or shortened to be exactly 512 ms in length using the STRAIGHT program (Kawahara *et al.*, 1999) to avoid the potential impact of the word length and have better control on various stimulus conditions. All words were equalized to the same overall RMS level using a customized MATLAB program.

3. Speech interruption

The original speech was interrupted by silent gaps to create 13 stimulus conditions. A schematic illustration of the interruption parameters is shown in Fig. 1. The interrupted speech was generated using the following steps. First, two anchor points corresponding to the first and last 32 ms of the onset and offset of each signal were chosen. The two points were fixed across all but the two special stimulus conditions (i.e., 2g64v and 2g64c) to serve as the center points of the first and last interruptions. This was done to ensure that the signal contained at least one fragment of both the initial and final consonant so the phonetic difference across the “looks” will be minimized. Second, the space between the two anchor points was equally divided and the corresponding location points served as the center points for each speech fragment. The interruption rates (i.e., number of interruptions per word) of 2, 4, 6, 8, 12, 16, 24, 32, and 48 were used, with

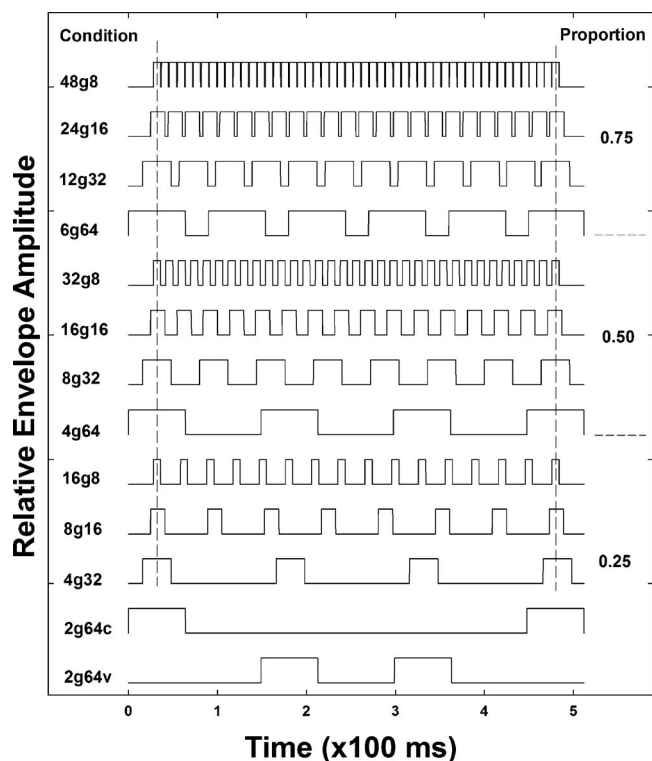


FIG. 1. Schematic illustration of the 13 stimulus conditions. The y-axis is the relative envelope amplitude and the x-axis is time in 100 ms scale, with a total duration of 512 ms. The 13 stimulus conditions are labeled on the left and the three total speech proportions are labeled on the right. The two vertical dashed lines correspond to the “anchor points” around which the first and last on-duration of the speech waveform was centered, respectively, for all stimulus conditions except 2g64c and 2g64v (stimulus conditions were labeled in “interruption rate-g-on duration” format).

on-durations (i.e., the durations that speech is gated on) of 8, 16, 32, and 64 ms. Various combinations of interruption rate and on-duration produced 11 stimulus conditions coded in “interruption rate-g-on-duration” format (i.e., 48g8, 24g16, 12g32, 6g64, 32g8, 16g16, 8g32, 4g64, 16g8, 8g16, 4g32). Two special conditions, 2g64c (two 64-ms on-durations in the consonant region only) and 2g64v (two 64-ms on-durations in the vowel region only), were used to represent the extreme conditions of missing phonemes. The 13 stimulus conditions resulted in three proportions of speech (i.e., the total proportion of speech that was gated on), which were 0.25, 0.5 and 0.75. Therefore, a total of three interruption parameters, interruption rate, on-duration and proportion of speech, were varied systematically. A 4-ms raised-cosine function was applied to the onset and offset of each speech fragment to reduce spectral splatter. The speech-interruption process was implemented using a custom MATLAB program. Five stimulus conditions (4g32, 8g16, 16g16, 32g8 and 6g64) determined to be of medium to easy difficulty in pilot testing were selected for use in practice (50 practice words, 10 per condition).

4. Calibration

A total of four white noises, two for each talker, were generated and spectrally shaped to match the long-term RMS amplitude spectrum of the original speech using Adobe Audition. Following spectral shaping, the total RMS levels of

two spectral-shaped noises were then equalized to the total RMS levels of the original concatenated speech, one for each talker; the total RMS level of another two spectral-shaped noises were equalized to that of the word with the highest amplitude among 150 words, one for each talker. A Larson-Davis Model 800 B sound level meter with Larson-Davis 2575 1-inch microphone was used for calibration. For RMS calibration, the noises were used to measure the levels of the center frequency of each 1/3-octave band from 80 to 8000 Hz and overall sound pressure level using an ER-3A insert earphone (Etymotic Research) in a 2-cm³ coupler (ANSI, 2004). For peak output calibration, the noises were used to measure the maximum output at center frequency of each 1/3-octave band from 1000 Hz and 4000 Hz with linear weighting. The output of the insert earphone at each of the computer stations was measured and the differences are all were within 1 dB range.

C. Procedures

Recognition of interrupted speech was assessed in two sound-treated booths that complied with ANSI (1999) guidelines. Subjects were randomly assigned to one of the three groups, which differed in terms of the talker and presentation level (i.e., male talker at 65 and 85 dB SPL, and female talker at 65 dB SPL), with roughly 20 subjects per group. Each subject only participated in one of the three groups. The three-group design facilitated the examination of the influence of two general factors: comparing performance for the male talker at 65 and 85 dB SPL will reveal any effects of presentation level, while comparing performance for the male and female talkers at 65 dB SPL will reveal any effects of talker gender. The male talker was randomly chosen to be presented at the high level. Subjects were seated in different cubicles and faced the computer monitors in one of the two booths. The right ear was designated as the test ear. A familiarization session was conducted prior to data collection. During familiarization, subjects heard 50 interrupted CVC words from a different male talker. Subjects were instructed to respond in an open-set response format (i.e., typing CVC words that matched what they heard). Visual feedback was provided through flipping cards printed with the target words. During data collection, all 13 stimulus conditions were presented in a proportion-blocked manner: stimuli were presented from low to high speech proportion (i.e., 0.25, 0.5 and 0.75) in order to minimize learning effects. Within each proportion, the order of stimulus conditions was randomized. Furthermore, within each talker group, all 150 CVC words were presented in random order for each of the 13 stimulus conditions. The test was conducted using the same open-set response format as in the familiarization session, but no feedback was provided. Testing was self-paced, so that subjects could take as much time as needed to respond after they heard a word.

Subjects' responses were collected via a MATLAB program and scored by a Perl program. The Perl program scored subjects' responses using a database, a combination of Celex (Baayen *et al.*, 1995) and a series of word lists for misspellings, homographs and nonwords, and then marking the an-

swers either as "correct," "incorrect," or "other" (e.g., misspelling, homograph and nonwords not previously identified). After subjects' responses were processed by the Perl program, a manual check was performed for items in the "other" category in order to find words that could be taken as correct responses according to their pronunciation. Two criteria were used to adjust for spelling errors. First, if a misspelling (two letters transposed) resulted in a new word (e.g., grid and gird), then the entry was counted as wrong. However, if a misspelling resulted in a nonword (e.g., "frim" for "firm"), the entry was counted as correct even though the two words would not have the same pronunciation. Second, any answer pronounced like the target word would be counted as correct (e.g., beak, beek, beeck). After the manual check, the newly defined accepted misspelled words were counted as "correct" and then percent-correct scores were re-calculated. Overall, about 0.3% of the total responses that were misspelled were subsequently accepted as correct responses.

III. RESULTS

First, the raw data from three subject groups were plotted to illustrate the effect of interruption parameters and then re-plotted to illustrate the effect of talker gender and presentation level. Second, two General Linear Model (GLM) repeated-measures (i.e., repeated-measures ANOVA) analyses were conducted to analyze the between-group factors of talker gender and presentation level, the within-subject factors of stimulus conditions and lexical properties. Finally, post-hoc comparisons were conducted as follow-up to the GLM analyses. In all cases, the percent-correct scores were transformed into rationalized arcsine units (RAU; Studebaker, 1985) before data analysis to stabilize the error variance. A set of GLM repeated-measures analyses was performed, before any other analyses, to examine the potential learning effect. The results suggested that no systematic effect of presentation order (learning) was observed for any speech proportion or subject group.

A. Effect of various factors on temporal integration

Figures 2–4 display transformed percent-correct scores (in RAU) as a function of on-duration, with symbols representing number of speech fragments and lines connecting results for the same on-proportion. In each figure, the top panel shows results for lexically easy words and the bottom panel for lexically hard words. Figures 2 and 3 depict the scores for the male talker at 65 and 85 dB SPL, respectively, and Fig. 4 depicts the scores for the female talker at 65 dB SPL. Figure 5 re-plots the transformed percent-correct scores as a function of stimulus condition, with talker gender and lexical properties as parameters. Similarly, Fig. 6 re-plots the transformed percent-correct scores as a function of stimulus condition, with level and lexical properties as parameters. The vertical dashed lines in Figs. 5 and 6 demark the stimulus conditions corresponding to the speech on-proportions of 0.25, 0.5 and 0.75. All told, the data in Figs. 2–6 suggest that the proportion of speech is the main factor that determines

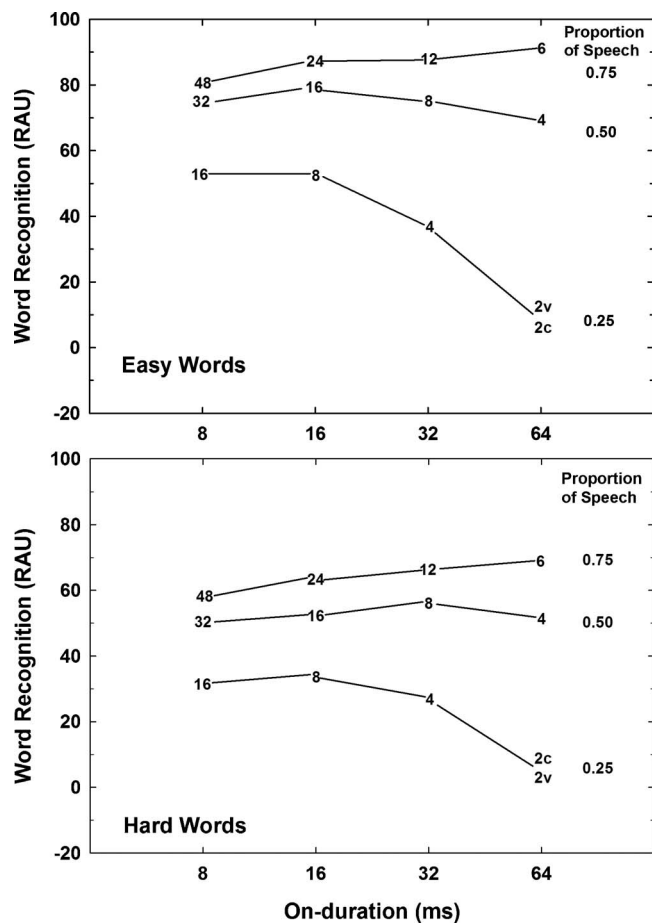


FIG. 2. The effect of three interruption parameters on recognition of interrupted words. Scores (in RAU) are plotted as a function of on-duration (ms) with proportion of speech as the parameter for the male talker group at 65 dB SPL. The numbers used as symbols represent the number of on-durations per stimulus. The top and bottom panels present results for easy words and hard words, respectively.

the intelligibility of words interrupted by silence. In general, the intelligibility of the interrupted CVC words increases as the proportion of speech increases. For a given proportion of speech, however, performance still varied for different stimulus conditions, with the largest variations seen for a proportion of 0.25.

Two GLM repeated-measures analyses were performed, one with talker gender and another with presentation level as between-subject factors, and both with stimulus conditions as well as lexical properties as within-subjects factors. The first GLM repeated-measures analysis revealed a significant ($p < 0.05$) main effect of talker gender [$F(1, 39) = 17.41$], whereby scores were significantly higher for the female talker than for the male talker. There were significant ($p < 0.05$) main effects observed for within-subject factors: stimulus condition [$F(12, 468) = 841.53$] and lexical properties [$F(1, 39) = 374.13$]. In addition, there were significant ($p < 0.05$) interactions between the within-subject and between-subjects factors: stimulus condition by talker [$F(12, 468) = 5.68$], stimulus condition by lexical properties [$F(12, 468) = 32.02$] and stimulus condition by lexical properties by talker [$F(12, 468) = 10$]. The second GLM repeated-measures analysis revealed a significant ($p < 0.05$) main ef-

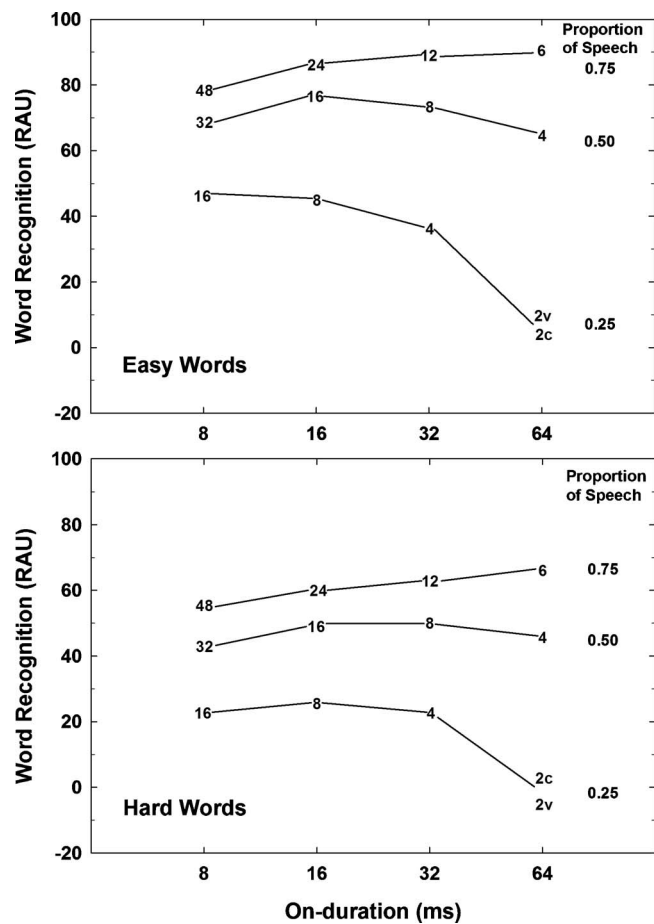


FIG. 3. Same as Fig. 2, but for the male talker at 85 dB SPL.

fect of presentation level [$F(1, 40) = 5.15$], whereby scores were significantly higher for the lower presentation level than for the higher presentation level. There were significant ($p < 0.05$) main effects observed for within-subject factors: stimulus condition [$F(12, 480) = 1183.02$] and lexical properties [$F(12, 480) = 2.45$]. There were also significant interactions between the within-subject factors and between-subjects factors: stimulus condition by level [$F(1, 40) = 438.79$] and stimulus condition by lexical properties [$F(12, 480) = 56.52$].

Post-hoc pair-wise comparisons (Bonferroni-adjusted t-tests) were conducted following the GLM analyses. A total of six sets (i.e., 2 levels of lexical difficulty \times 3 subject groups = 6: "Male65Easy," "Male65Hard," "Male85Easy," "Male85Hard," "Female65Easy," "Female65Hard") of pair-wise comparisons were conducted for stimulus conditions comprised of each of the three proportions of speech (i.e., 0.25, 0.5 and 0.75). A few trends that were consistent across all or most of the six sets of post-hoc pair-wise comparisons are discussed in the ensuing paragraphs (please see the Appendix for full details). First, for all five stimulus conditions with a speech proportion of 0.25, scores for the 2g64v and 2g64c conditions were significantly lower than those for the other stimulus conditions, whereas those for the 8g16 condition were not significantly different from those for the 16g8 condition for all six sets of pair-wise comparisons. In addition, for four of the six sets of pair-wise comparisons, scores

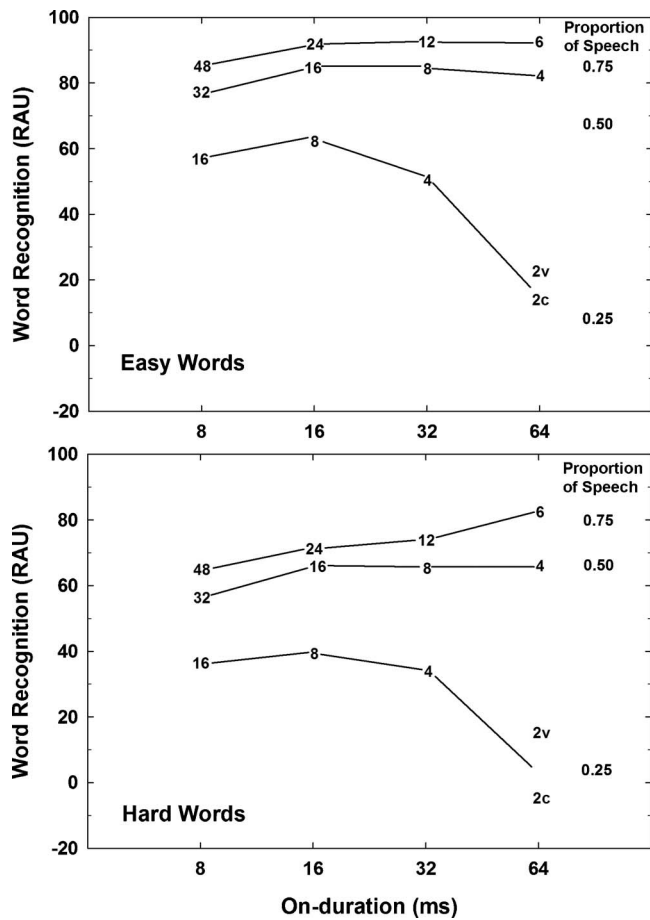


FIG. 4. Same as Fig. 2, but for the female talker at 65 dB SPL.

for the 4g32 condition were significantly lower than those for the 8g16 condition. Second, for all four stimulus conditions with a speech proportion of 0.75, scores for the 48g8 condition were always significantly lower than those in the other

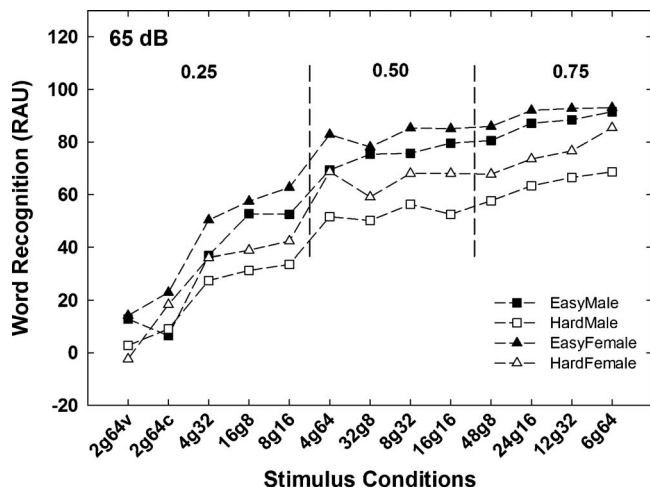


FIG. 5. Effect of talker gender on recognition of interrupted words at 65 dB SPL. Scores (in RAU) were plotted for 13 stimulus conditions, with talker gender and lexical difficulty as parameters. Conditions were roughly arranged in ascending order of performance from left to right, grouped according to the proportion of speech presented (indicated by vertical dashed lines and text labels at the top). The filled and open symbols represent easy and hard words, respectively. Square and triangle symbols represent male and female talkers, respectively.

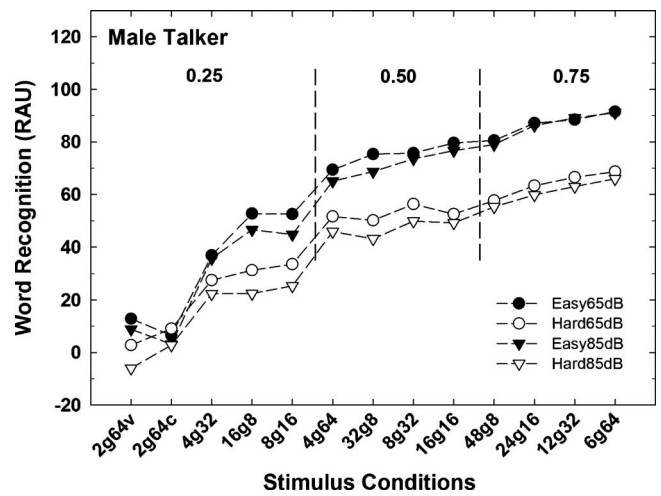


FIG. 6. Effect of presentation level on recognition of interrupted words for the male talker. Scores (in RAU) were plotted for 13 stimulus conditions, with presentation level and lexical difficulty as parameters. Conditions were roughly arranged in ascending order of performance from left to right, grouped according to the proportion of speech presented (indicated by vertical dashed lines and text labels at the top). The filled and open symbols represent easy and hard words, respectively. Circle and down triangle symbols represent 65 dB SPL and 85 dB SPL, respectively.

three test conditions for all six sets of pair-wise comparisons. Scores for the 6g64 condition were not significantly different from those in the 12g32 condition for all but one set of pair-wise comparison. In addition, scores for the 24g16 condition were significantly lower than those for the 6g64 condition for four of the six sets of pair-wise comparisons. Third, for the four stimulus conditions with a speech proportion of 0.5, scores for the 8g32 condition were not significantly different from those in the 16g16 condition for all six sets of pair-wise comparisons. Scores for the 32g8 condition, in the meantime, were significantly lower than those in the 16g16 condition for five of the six sets of pair-wise comparisons; and lower than those in the 8g32 condition for four of the six pair-wise comparisons. Results of the post-hoc pair-wise comparisons confirmed the secondary role played by interruption rate and on-duration in the integration process. That is, even when the proportion of speech was fixed, as the other two parameters changed, the intelligibility of the interrupted words changed significantly, often in a consistent fashion across all or most of the six sets of post-hoc pair-wise comparisons. In general, this secondary influence was primarily observed for a speech proportion of 0.25: interrupted words with fast interruption rates and short on-durations tended to be more intelligible than words with slow interruption rates and long on-durations. The opposite trend was suggested for the speech proportion of 0.75. There was no such obvious trend observed for the speech proportion of 0.5.

B. Correlations among stimulus conditions and individual differences

To examine the consistency of individual differences in recognition of interrupted words, word recognition scores were averaged across different test conditions and lexical

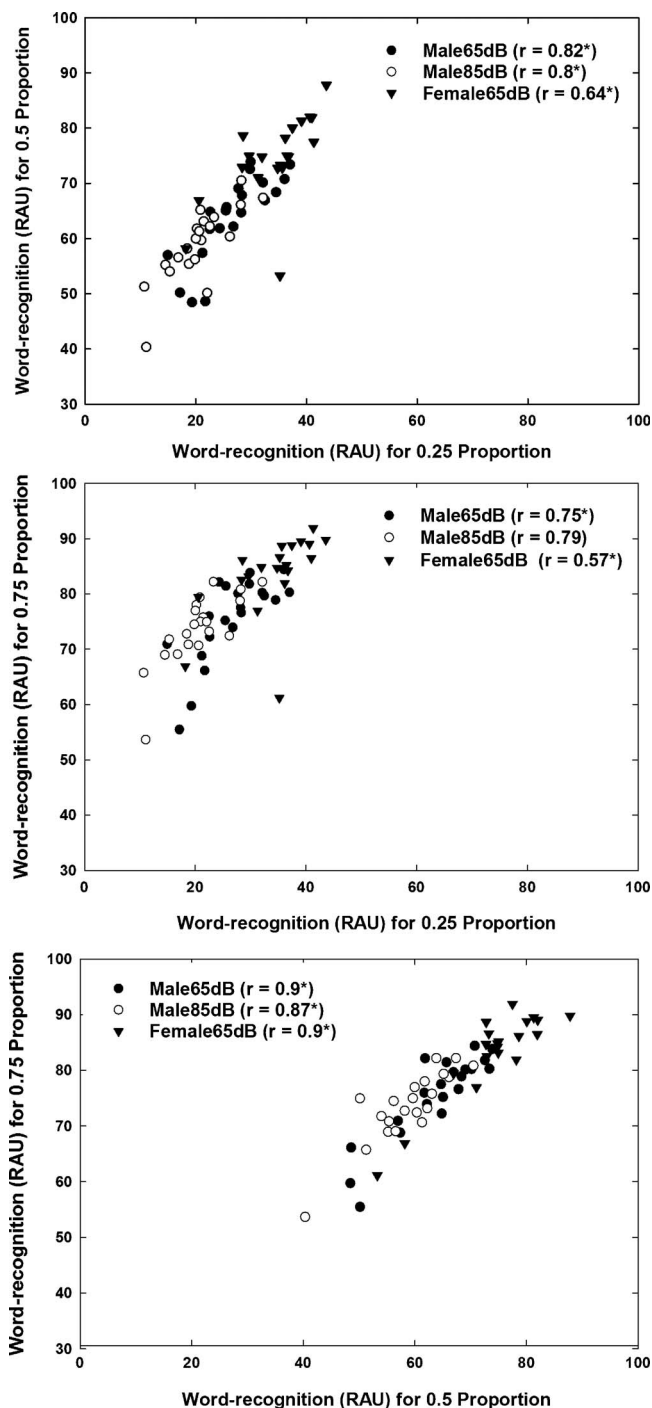


FIG. 7. Scatter plots of word recognition scores for three proportions. The top, middle and bottom panels represent relationships among word recognition scores between 0.25 and 0.5, 0.25 and 0.75, 0.5 and 0.75 proportions, respectively. The “Male65dB” (words recorded by the male talker and presented at 65 dB SPL) group is labeled as filled circles, “Male85dB” (words recorded by the male talker and presented at 85 dB SPL) group is labeled as unfilled circles, and “Female65dB” is labeled as filled triangles. Correlations between average scores for each proportion for each of the three subject groups are shown in the parentheses within the figure legends. Asterisks indicate statistically significant correlations at $p \leq 0.05$.

difficulties within each proportion for each subject group and then subjected to a series of correlational analyses. Results showed strong positive correlations ($r=0.57$ to 0.9 , $p < 0.05$) among scores for all proportions, as shown in Fig. 7. In

general, results suggest consistency in performance: subjects who scored highest or lowest in a given condition did likewise across the other conditions. A substantial range of individual differences was observed. The inherent differences in recognition of interrupted words among individuals, as well as differences in terms of talker and presentation level across the three subject groups, may have contributed to the large individual differences. In general, however, the individual who is good at piecing together the speech fragments in one condition remains among the better performing subjects in other conditions.

IV. DISCUSSION

The current study was designed to investigate the effects of various factors on the temporal integration of interrupted speech. In particular, the effects of interruption parameters (e.g., interruption rate, on-duration and proportion of speech) on the temporal integration process were examined. In addition, the effects of linguistic factors (e.g., lexical properties) as well as the interaction between the interruption parameters and linguistic factors were examined. Other general factors, such as talker gender and presentation level, were also investigated.

A. Effects of interruption parameters

It was demonstrated that all three interruption parameters had an influence on the temporal integration of speech glimpses. The influence is determined by a complex nonlinear relationship among three parameters: performance is determined primarily by the proportion of speech listeners “glimpsed,” but modified by both the interruption rate and duration of the “glimpses” or “looks.”

1. Proportion of speech

In general, recognition of interrupted words increases as the proportion of speech presented increases, as shown in Figs. 2–6. This finding is consistent with previous findings in quiet and noise (e.g., Miller and Licklider, 1950; Cooke, 2006). More specifically, the current study discovered that the growth of intelligibility with increasing speech proportion is not linear. The biggest increment of intelligibility (about 30 to 50 RAUs) occurred when the speech proportion increased from 0.25 to 0.5. Less improvement (about 5 to 20 RAUs) is observed for the same amount of increment when speech proportion increases from 0.5 to 0.75. Although the effect of the speech proportion on performance was dominant among the interruption parameters studied, there is still substantial variation in performance for different stimulus conditions having the same speech proportion, particularly for speech proportion of 0.25, which indicates the influence of other interruption parameters.

2. Interruption rate

Findings of the current study show that the effect of interruption rate varies depending on the proportion of speech presented. For all three subject groups, when the available speech proportion is 0.25, word recognition tended

to increase as interruption rate increased. For stimuli with 0.5 proportion of speech, intelligibility remained generally constant as interruption rate increased. When the proportion of speech increased to 0.75, word recognition slightly decreased with increasing interruption rate. The dynamic change in speech recognition with interruption rate at different speech proportions is, in general, consistent with the trends reported by Miller and Licklider (1950) for a comparable range of interruption frequencies (i.e., 4 to 96 Hz), although the latter interpolated the trend through relatively sparse data points. For instance, both Miller and Licklider (1950) and the current study demonstrated an increase in speech recognition as the frequency of interruption increased above 4 Hz. Particularly, both studies showed that the lowest scores occurred near a modulation frequency of 4 Hz. When this interruption rate was used, entire phonemes were eliminated, which in turn caused a large decrease in intelligibility. In addition to the effect of missing phonemes, the disruptions in temporal envelope, an important cue for speech perception (e.g., Van Tasell *et al.*, 1987; Drullman, 1995; Shannon *et al.*, 1995), might also account for the dramatically decreased performance. Previous research has suggested that extraneous modulation from the interruptions may interfere with processing of the inherent envelope rate of speech itself, especially at low modulation rates (e.g., Gilbert *et al.*, 2007). In analogous psychoacoustic contexts, this interference is referred to as modulation detection interference (MDI) (e.g., Sheft and Yost, 2006; Kwon and Turner, 2001; Gilbert *et al.*, 2007). As Drullman *et al.* (1994a, 1994b) suggested, modulation frequencies between 4 and 16 Hz are important for speech recognition. When the modulations in this frequency range were intact, filtering out very slow or fast envelope modulations in the waveform did not significantly degrade speech intelligibility. In the current study, the interruption rates ranged from about 4 to 96 Hz, which means that an MDI-like process could potentially play a detrimental role. Regardless of the similar general trend observed in current study and Miller and Licklider (1950), there is still some difference. For instance, both studies observed a roll-off in speech recognition as the frequency of interruption reached a certain rate. However, in the current study, speech recognition started to decrease around an interruption rate of 64 Hz for a speech proportion of 0.5 and around a rate of 24 Hz for speech proportion of 0.75. While in the Miller and Licklider (1950) study, the roll-off occurred when the frequency of interruption was higher than 100 Hz, based on the relatively sparse data provided in that study. The difference could result from different experimental paradigms used in both studies. For instance, Miller and Licklider (1950) used lists of phonetically balanced words interrupted by an electronic switch, while the current study used lexically easy and hard words digitally interrupted and windowed. For the comparable range of interruption frequency and speech proportion, the current study also agrees with previous research that examined the effect of interruption frequency on speech recognition for one particular speech proportion (e.g., 0.50; Nelson and Jin, 2004). Findings from current study suggest that the mixed results regarding the effect of interruption fre-

quency on speech recognition from previous research may result from an interaction between interruption frequency and the proportion of speech.

3. Duration

On-duration is the interruption parameter that has received the least amount of attention in previous research. The current study found that the effect of signal duration on the recognition of interrupted words varies depending on the proportion of the speech presented. When only a limited proportion of speech is available (i.e., 0.25), CVC words with shorter speech on-durations are more intelligible than those with sparse, long on-durations. This finding supports the previous discovery by Li and Loizou (2007), which found that interrupted IEEE sentences with frequent short (20 ms) on-durations were more intelligible than the ones with sparse, long (400–800 ms) on-durations when the proportion of speech was 0.33. When more speech information is available (i.e., speech proportion ≥ 0.5), however, CVC words with longer on-duration have a slight advantage. The disadvantage of longer duration at low speech proportions (i.e., 0.25 or 0.33) is likely caused, in part, by the number of phonemes sampled. That is, at the proportion of 0.25, the longest on-duration was 64 ms, which only provided either a portion of the center vowel or the initial/final consonants of the CVC words used in this study. Comparatively, decreasing the on-duration for this proportion resulted in multiple samples of each phoneme comprising the CVC word, which provided a better picture of the whole utterance than the longer on-duration. In addition, the duration of the silence between each speech on-duration is also likely to play a role in the decrease of speech recognition. In the current study, the silent interval is longest overall when the speech on-duration is 64 ms at a speech proportion of 0.25 (e.g., silence interval is 384 ms for 2g64c condition). Huggins (1975) examined the influence of speech intervals as well as silence intervals using temporally segmented speech created by inserting silence into continuous speech. Results suggested that durations of both speech and silence intervals determine the intelligibility of temporally segmented speech. Huggins (1975) suggested that, when the silent interval was long, each speech on-duration tended to be processed as isolated fragments of speech. When the speech interval was short, however, it was easier for the ear to “bridge the gap” and combine related speech segments before and after the silence. This suggests that in the temporal integration process for interrupted speech, the duration of the speech segments and silence should both be considered.

B. Effect of linguistic factors

The findings from the current study show that, on average, lexically easy words were more intelligible than lexically hard words for most interruption conditions. When limited acoustic information was available (i.e., 0.25), on average, lexically easy words result in a ~ 14 -RAU advantage compared to lexically hard words. Particularly, the more difficult interruption conditions improve less (e.g., 1 RAU

for 2g64c condition) while the less difficult interruption conditions improve more (e.g., 20 RAUs for 8g16 condition). When enough acoustic information is available (i.e., speech proportion ≥ 0.5), the advantage of lexically easy words is about 20 RAUs, with some minor variations among different interruption conditions. This suggests an interaction between the acoustic interruption-related factors and the top-down linguistic factors in temporal integration: interruption-related acoustic factors dominate the perception when the total amount of acoustic information is low; as the amount of acoustic information increases, top-down linguistic factors start to play a role and emerge as significant factors.

The interaction observed between acoustic interruption factors and top-down processing in the current study was also suggested by other studies. For instance, [Wingfield *et al.* \(1991\)](#) examined the ability of young and elderly listeners to recognize words increased in word-onset duration (“gated” words). They found that both groups need 50% to 60% of the word-onset information to recognize the “gated” words without context, but only 20% to 30% when the words are embedded in sentence context. In the current study, to reach the same degree of speech recognition, less acoustic information was required for lexically easy words than hard words. In addition, the finding that consistent lexical advantage occurs only after the speech proportion is above 0.5 suggests that a certain amount of acoustic information is required before listeners can make full use of their knowledge of the lexicon to fill in the blanks in the interrupted speech.

C. Effects of other general factors

Although the interruption factors, especially the proportion of speech, as well as linguistic factors appear to be the main determining factors for recognition of interrupted speech, other general factors, such as talker and presentation level, also played a role in recognizing words interrupted with silence.

1. Effect of talker gender

The current study found that interrupted words with silent gaps produced by the female talker were significantly more intelligible than those produced by the male talker (by about 5 RAUs for lexically easy and 10 RAUs for lexically hard words, see Fig. 5). This is consistent with prior research on the recognition of “clear speech” (e.g., [Bradlow *et al.*, 1996](#); [Bradlow and Bent, 2002](#); [Bradlow *et al.*, 2003](#); [Ferguson, 2004](#)). It has been suggested that a key contributor to this gender difference resides in the inherent differences in fundamental frequency between males and females (e.g., [Assmann, 1999](#)). In our study, the fundamental frequency (F_0) of the female talker was about one octave higher than the F_0 of the male talker (230 Hz vs. 110 Hz). For a given set of interruption parameters, this difference corresponds to twice as many pitch periods in the interrupted speech produced by the female than the male, which might facilitate integration of speech across speech fragments. That is, with more pitch periods in a given glimpse, a more reliable estimate of fundamental frequency may be obtained from moment to moment during the word and this may facilitate in-

tegration of glimpses over time. Inspection of the data, however, does not support this as a significant factor in the integration of speech information. Given the nearly octave separation in F_0 for the two talkers in this study and the use of many twofold differences in on-duration, it is possible to examine many pairs of stimulus conditions for which the number of pitch periods in a glimpse would be about the same for the male and female talker in this study. When doing so, scores for the male talker remain slightly lower than those for the female talker, suggesting that other differences between the male and female voices are responsible for the superior intelligibility of the female talker, rather than difference in the number of pitch periods per glimpse. Importantly, however, the overall trends in the data regarding the effects of interruption-related and linguistic factors were the same for both the male and female talker (see Figs. 2 and 4).

2. Effect of presentation level

Interrupted words presented at a higher intensity (85 dB SPL) were significantly less intelligible (difference generally < 5 RAUs) than those presented at a lower intensity (65 dB SPL). The decline occurred mainly for conditions with a low to medium proportion of speech (0.25 to 0.5). The difference in performance between the two presentation levels was statistically significant, although the effects of presentation level were small compared to other factors. It is a widely observed phenomenon that there is a decrease in intelligibility when speech levels increased above 80 dB SPL (e.g., [Fletcher, 1922](#); [French and Steinberg, 1947](#); [Pollack and Pickett, 1958](#); [Studebaker *et al.*, 1999](#)). Such research has been conducted exclusively with uninterrupted speech, either in quiet or in a background of steady-state noise. Therefore, despite acoustic differences between interrupted and continuous speech, both are likely subject to the same fundamental, and yet to be identified, mechanism which causes a decline in recognition at higher levels. Importantly, however, the overall trends in the data regarding the effects of interruption-related and linguistic factors were the same for both presentation levels (see Figs. 2 and 3).

D. Implications for theories of temporal integration of speech

1. More than “multiple looks”

The current study supports the general notion of “multiple looks.” However, it is not just the number of looks or their duration, but the combination of these two factors, the proportion of speech remaining, which is most critical. For example, from Figs. 5 and 6, speech-recognition performance for 6 evenly spaced 64-ms “looks” at CVC words is always superior, by about 10–15 RAU, than 32 8-ms or 16 16-ms looks at the same CVC words. The total proportion of speech integrated is the primary factor determining the recognition of interrupted words. This finding is consistent with earlier research regarding the temporal integration process in speech using different experimental paradigms. For instance,

in a syllable-discrimination task using a change/no-change procedure, Holt and Carney (2005) discovered that, as number of repetitions of standard and comparison /CV/ syllables increased (i.e., more chances for “looks”), discrimination threshold improved. Similarly, Cooke (2006) found that listeners performed better in a consonant-identification task when they could access and integrate more target speech in which the local SNR exceeded 3 dB (i.e., regions referred to as “glimpses”). The same integration was also observed in a word-recognition task by Miller and Licklider (1950).

On the other hand, however, the current study suggests that when the speech proportion is fixed, it is not necessarily true that “more looks” always produce better recognition. For instance, when speech proportion was 0.75, more “looks” actually resulted in reduced recognition (e.g., 48g8 < 24g16 < 12g32 < 6g64). Thus, it is not solely the proportion of speech or the number of looks that impact the perception of interrupted speech. When a low proportion of speech was available (i.e., speech proportion=0.25), CVC words with frequent and short speech on-durations tended to be more intelligible. In contrast, when the proportion of speech was high (i.e., speech proportion ≥ 0.5), CVC words with sparse and long on-durations were more intelligible. This suggests that speech recognition varies according to the way the available speech proportion is sampled.

2. Intelligent temporal integration

It has been suggested that temporal integration is not a simple summation or accumulation process; instead, it is an “intelligent” integration process. For instance, Moore (2003) proposed that, in the temporal integration process, the internal representation of a signal, which is calculated from the STEP derived at the auditory periphery, is compared with a template in long-term memory. In addition, a certain degree of “intelligent” warping or shifting of the internal representation would be executed to adapt to speech- or talker-related characteristics. The current findings provide further support for the concept of “intelligent temporal integration.” The fact that, for lexically easy and hard words that were interrupted by the same interruption parameters, lexically easy words were more intelligible than lexically hard words suggests that temporal integration process intelligently integrates the linguistic information to facilitate the word recognition. An additional interesting finding is that linguistic factors provided more benefit to perception when the amount of acoustic information available was low; and less benefit when the available acoustic information was high. The dynamics between the interruption parameters and linguistic factors suggests that the temporal integration process may intelligently place different weights on acoustically driven bottom-up and linguistically driven top-down information to maximize the overall benefit.

3. Effects from other general factors

In addition to the interruption and linguistic factors, other acoustically related factors, such as talker gender and presentation level, also shaped the recognition of interrupted

speech, even when the patterns of “looks” were the same. Therefore, these data suggests that the factors which play roles in the recognition of continuous speech, such as talker gender and presentation level, also influence the recognition of interrupted speech. Models of the temporal integration of speech should also take these factors into consideration.

V. CONCLUSIONS

In summary, findings from the current study suggest that both interruption parameters and linguistic factors affect the temporal integration and recognition of interrupted words. Other general factors such as talker gender and presentation level also have an impact. Among these factors, the proportion of speech and lexical difficulty produced primary and large effects (>20 RAUs) while on-duration, interruption rate, talker gender, and presentation level produced secondary and small effects (about 5–10 RAUs). Findings from the current study support the traditional theory of “multiple looks” in temporal integration, but suggest that the process is not driven simply by the number of looks or their duration. Rather, it is the combination of these factors and their contribution to the total proportion of the speech signal preserved that is critical. These results suggest an intelligent temporal integration process in which listeners might modify their integration strategies based on information from both acoustically driven bottom-up and linguistically driven top-down processing stages in order to maximize performance.

ACKNOWLEDGMENTS

This work was submitted by the first author in partial fulfillment of the requirements for the Ph.D. degree at Indiana University. This work was supported by research Grant R01 AG008293 from the National Institute on Aging (NIA) awarded to the second author. We thank Sumiko Takayanagi for providing original digital recordings of the CVC words. Judy R. Dubno, Jayne B. Ahlstrom and Amy R. Horwitz at the Medical University of South Carolina provided valuable comments on this manuscript.

APPENDIX: POST-HOC PAIR-WISE COMPARISONS

The results of the post-hoc pair-wise comparisons for six sets of the test conditions-“Male65Easy,” “Male65Hard,” “Male85Easy,” “Male85Hard,” “Female65Easy,” “Female65Hard”-are presented in the following table. The three separate panels in the table represent each proportion of speech tested (i.e., 0.25, 0.5 and 0.75). Within each panel, the rows and columns contain the conditions tested for each speech proportion and the cells indicate the differences (in RAU) between the two conditions. Asterisks indicate statistically significant differences at $p \leq 0.05$.

Group		Test conditions (0.25)					Test conditions (0.5)				Test conditions (0.75)		
		2g64c	4g32	8g16	16g8		8g32	16g16	32g8		12g32	24g16	48g8
Male65Easy	2g64v	6.358*	-24.038*	-39.681*	-39.899*	4g64	-6.307*	-10.197*	-5.930*	6g64	3.114	4.364	10.961*
	2g64c		-30.396*	-46.039*	-46.257*	8g32		-3.890	0.377	12g32		1.249	7.846*
	4g32			-15.643*	-15.861*	16g16			4.267*	24g16			6.597*
	8g16				-0.219								
Male65Hard	2g64v	-6.246	-24.682*	-30.768*	-28.503*	4g64	-4.707*	-0.892	1.446	6g64	2.123	5.299*	10.986*
	2g64c		-18.436*	-24.522*	-22.257*	8g32		3.815	6.153*	12g32		3.176*	8.862*
	4g32			-6.086*	-3.821	16g16			2.338	24g16			5.686*
	8g16				2.265								
Male85Easy	2g64v	5.684*	-26.894*	-35.917*	-37.771*	4g64	-8.492*	-11.620*	-3.647	6g64	2.123	5.299*	10.986*
	2g64c		-32.579*	-41.601*	-43.455*	8g32		-3.128	4.845	12g32		3.176*	8.862*
	4g32			-9.022*	-10.877*	16g16			7.973*	24g16			5.686*
	8g16				-1.854								
Male85Hard	2g64v	-8.771*	-28.308*	-31.324*	-28.399*	4g64	-3.992	-3.292	2.732	6g64	2.981	6.130*	10.688*
	2g64c		-19.537*	-22.553*	-19.628*	8g32		0.700	6.724*	12g32		3.149*	7.707*
	4g32			-3.016	-0.091	16g16			6.024*	24g16			4.558*
	8g16				2.925								
Female65Easy	2g64v	-8.812*	-36.213*	-48.623*	-43.348*	4g64	-2.462	-2.208	4.782*	6g64	0.295	1.025	7.118*
	2g64c		-27.401*	-39.812*	-34.536*	8g32		0.253	7.243*	12g32		0.730	6.824*
	4g32			-12.410*	-7.135	16g16			6.990*	24g16			6.093*
	8g16				5.275	32g8				48g8			
Female65Hard	2g64v	-20.642*	-38.493*	-44.703*	-41.246*	4g64	0.685	0.695	9.642*	6g64	8.815*	11.899*	17.692*
	2g64c		-17.851*	-24.061*	-20.604*	8g32		0.010	8.957*	12g32		3.083	8.877*
	4g32			-6.210	-2.752	16g16			8.947*	24g16			5.793*
	8g16				3.457								

ANSI (1996). "Specifications for audiometers," ANSI S3.6-1996 (American National Standards Inst., New York).

ANSI (1999). "Maximum permissible ambient levels for audiometric test rooms," ANSI S3.1-1999 (American National Standards Inst., New York).

ANSI (2004). "Specification for audiometers," ANSI S3.6-2004 (American National Standards Inst., New York).

Assmann, P. F. (1999). "Fundamental frequency and the intelligibility of competing voices," in Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, CA, 1-7 August, 179-182.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM)* (University of Pennsylvania, Philadelphia).

Bashford, J., Jr., and Warren, R. (1987). "Effects of spectral alternation on the intelligibility of words and sentences," *Percept. Psychophys.* **42**, 431-438.

Bashford, J. A., Riener, K. R., and Warren, R. M. (1992). "Increasing the intelligibility of speech through multiple phonemic restorations," *Percept. Psychophys.* **51**, 211-217.

Bradlow, A. R., and Bent, T. (2002). "The clear speech effect for non-native listeners," *J. Acoust. Soc. Am.* **112**, 272-284.

Bradlow, A. R., Kraus, N., and Hayes, E. (2003). "Speaking clearly for children with learning disabilities: Sentence perception in noise," *J. Speech Lang. Hear. Res.* **46**, 80-97.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (1996). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," *Speech Commun.* **20**, 255-272.

Cooke, M. (2003). "Glimpsing speech," *J. Phonetics* **31**, 579-584.

Cooke, M. (2006). "A glimpsing model of speech perception in noise," *J. Acoust. Soc. Am.* **119**, 1562-1573.

Dirks, D. D., and Bower, D. (1970). "Effect of forward and backward masking on speech intelligibility," *J. Acoust. Soc. Am.* **47**, 1003-1008.

Dirks, D. D., Takayanagi, S., and Moshfegh, A. (2001). "Effects of lexical factors on word recognition among normal-hearing and hearing-impaired listeners," *J. Am. Acad. Audiol.* **12**, 233-244.

Drullman, R. (1995). "Temporal envelope and fine structure cues for speech intelligibility," *J. Acoust. Soc. Am.* **97**, 585-592.

Drullman, R., Festen, J. M., and Plomp, R. (1994a). "Effect of temporal

envelope smearing on speech perception," *J. Acoust. Soc. Am.* **95**, 1053-1064.

Drullman, R., Festen, J. M., and Plomp, R. (1994b). "Effects of reducing slow temporal modulations on speech reception," *J. Acoust. Soc. Am.* **95**, 2670-2680.

Ferguson, S. H. (2004). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," *J. Acoust. Soc. Am.* **116**, 2365-2373.

Fletcher, H. (1922). "The nature of speech and its interpretation," *J. Franklin Inst.* **193**, 729-747.

French, N. R., and Steinberg, J. C. (1947). "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Am.* **19**, 90-119.

Gareth Gaskell William, M. G., and Marslen-Wilson, W. D. (1997). "Integrating form and meaning: A distributed model of speech perception," *Lang. Cognit. Processes* **12**, 613-656.

Gilbert, G., Bergeras, I., Voillery, D., and Lorenzi, C. (2007). "Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues," *J. Acoust. Soc. Am.* **122**, 1336-1339.

Holt, R. F., and Carney, A. E. (2005). "Multiple looks in speech sound discrimination in adults," *J. Speech Lang. Hear. Res.* **48**, 922-943.

Huggins, A. W. F. (1975). "Temporally segmented speech," *Percept. Psychophys.* **18**, 149-157.

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.

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.

Li, N., and Loizou, P. C. (2007). "Factors influencing glimpsing of speech in noise," *J. Acoust. Soc. Am.* **122**, 1165-1172.

Liu, C. (2008). "Rollover effect of signal level on vowel formant discrimination," *J. Acoust. Soc. Am.* **123**, EL52-EL58.

Luce, P. A., and Pisoni, D. B. (1998). "Recognizing spoken words: The neighborhood activation model," *Ear Hear.* **19**, 1-36.

Marslen-Wilson, W. D., and Welsh, A. (1978). "Processing interactions and

- lexical access during word recognition in continuous speech," *Cogn. Psychol.* **10**, 29–63.
- Miller, G., and Licklider, J. (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.
- Nelson, P., 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.
- 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**, 195–199.
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Sheft, S., and Yost, W. A. (2006) "Modulation detection interference as informational masking," in *Hearing: From Basic Research to Applications*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, S. Uppenkamp, and J. Verhey (Springer Verlag, New York).
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Lang. Hear. Res.* **28**, 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**, 2431–2444.
- Takayanagi, S., Dirks, D. D., and Moshfegh, A. (2002). "Lexical and talker effects on word recognition among native and non-native listeners with normal and impaired hearing," *J. Speech Lang. Hear. Res.* **45**, 585–597.
- Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (1987). "Speech waveform envelope cues for consonant recognition," *J. Acoust. Soc. Am.* **82**, 1152–1161.
- Verschuure, J., and Brocaar, M. P. (1983). "Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise," *Percept. Psychophys.* **33**, 232–240.
- Viemeister, N. F., and Wakefield, G. H. (1991). "Temporal integration and multiple looks," *J. Acoust. Soc. Am.* **90**, 858–865.
- Wingfield, A., Aberdeen, J. S., and Stine, E. A. L. (1991). "Word onset gating and linguistic context in spoken word recognition by young and elderly adults," *J. Gerontol.* **46**, 127–129.