

Interrupted speech perception: The effects of hearing sensitivity and frequency resolution

Su-Hyun Jin^{a)}

Division of Communication Sciences and Disorders, University of Texas, Austin, Texas 78712

Peggy B. Nelson

Department of Speech-Language Hearing Sciences, University of Minnesota, Minneapolis, Minnesota 55455

(Received 4 November 2008; revised 4 June 2010; accepted 7 June 2010)

Jin & Nelson (2006) found that although amplified speech recognition performance of hearing-impaired (HI) listeners was equal to that of normal-hearing (NH) listeners in quiet and in steady noise, nevertheless HI listeners' performance was significantly poorer in modulated noise. As a follow-up, the current study investigated whether three factors, auditory integration, low-mid frequency audibility and auditory filter bandwidths, might contribute to reduced sentence recognition of HI listeners in the presence of modulated interference. Three findings emerged. First, sentence recognition in modulated noise found in Jin & Nelson (2006) was highly correlated with perception of sentences interrupted by silent gaps. This suggests that understanding speech interrupted by either noise or silent gaps require similar perceptual integration of speech fragments available either in the dips of a gated noise or across silent gaps of an interrupted speech signal. Second, those listeners with greatest hearing losses in the low frequencies were poorest at understanding interrupted sentences. Third, low-to mid-frequency hearing thresholds accounted for most of the variability in Masking Release (MR) for HI listeners. As suggested by Oxenham and his colleagues (2003 and 2009), low-frequency information within speech plays an important role in the perceptual segregation of speech from competing background noise.

© 2010 Acoustical Society of America. [DOI: 10.1121/1.3458851]

PACS number(s): 43.71.Ky, 43.71.Hw, 43.66.Ed [AE: KWG]

Pages: 881–889

I. INTRODUCTION

Background noise is the primary complaint of listeners with sensorineural hearing loss (SNHL) and hearing aids are only partially successful in restoring satisfaction in background noise (Kochkin, 2007). Many natural background noises are temporally varying, such as clattering dishes or background conversations. Listeners with normal hearing (NH) sensitivity take advantage of gaps in these fluctuating or modulated maskers. They are able to “listen in the dips” of the modulated masker to extract information about the speech signal. This improvement in speech recognition observed with modulated maskers compared to steady-state maskers is referred to as a “masking release.” In a previous paper (Jin and Nelson, 2006) it was observed that NH listeners' performance could improve by as much as 80 percentage points when noise was gated versus steady, while performance of 9 young hearing-impaired (HI) listeners improved by, at best, only approximately half that amount. Even when HI listeners were presented amplified stimuli so that their performance in quiet and in steady noise was similar to NH listeners, the HI listeners experienced significantly reduced MR, with two of the 9 HI listeners experiencing little to no MR. Figures 1(a) and 1(b) show significantly lower speech recognition for HI listeners in fluctuating noise than for NH

listeners which resulted in reduced amount of MR at -5 and -10 dB SNR (adapted from Jin and Nelson, 2006).

Similar to “listening in the dips of fluctuating noise,” understanding interrupted speech is presumed to be a test of a listener's ability to combine speech information across separate segments presented over time, despite interruptions in the speech stream (Bashford *et al.*, 1988; Warren, 1984). In order to understand interrupted sentences, listeners must integrate the auditory information into a continuous speech signal. Reduced frequency resolution may also result in reduced integration of interrupted speech signals which are gated by periods of silence. Evidence from CI research suggests that when frequency resolution is reduced, listeners do not integrate or fuse interrupted speech (e.g., Nelson and Jin, 2004). In addition, Nelson and Jin (2004) showed that NH listeners presented with implant simulations are more successful at integrating interrupted sentences as the amount of spectral information increases. These data imply that detailed spectral information is important for combining information from segments of interrupted sentences. Furthermore, since there is no noise present, interruption of speech by silent gaps eliminates any effect of forward masking. Therefore, the current study examined the perception of interrupted speech for NH and HI listeners to investigate how well HI listeners are able to integrate fragments of speech without any possibility of forward masking. We further investigated how perception of speech interrupted by noise and by silent gaps are related to each other. For both sets of experiments, square-wave gating was used to interrupt the speech or the

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

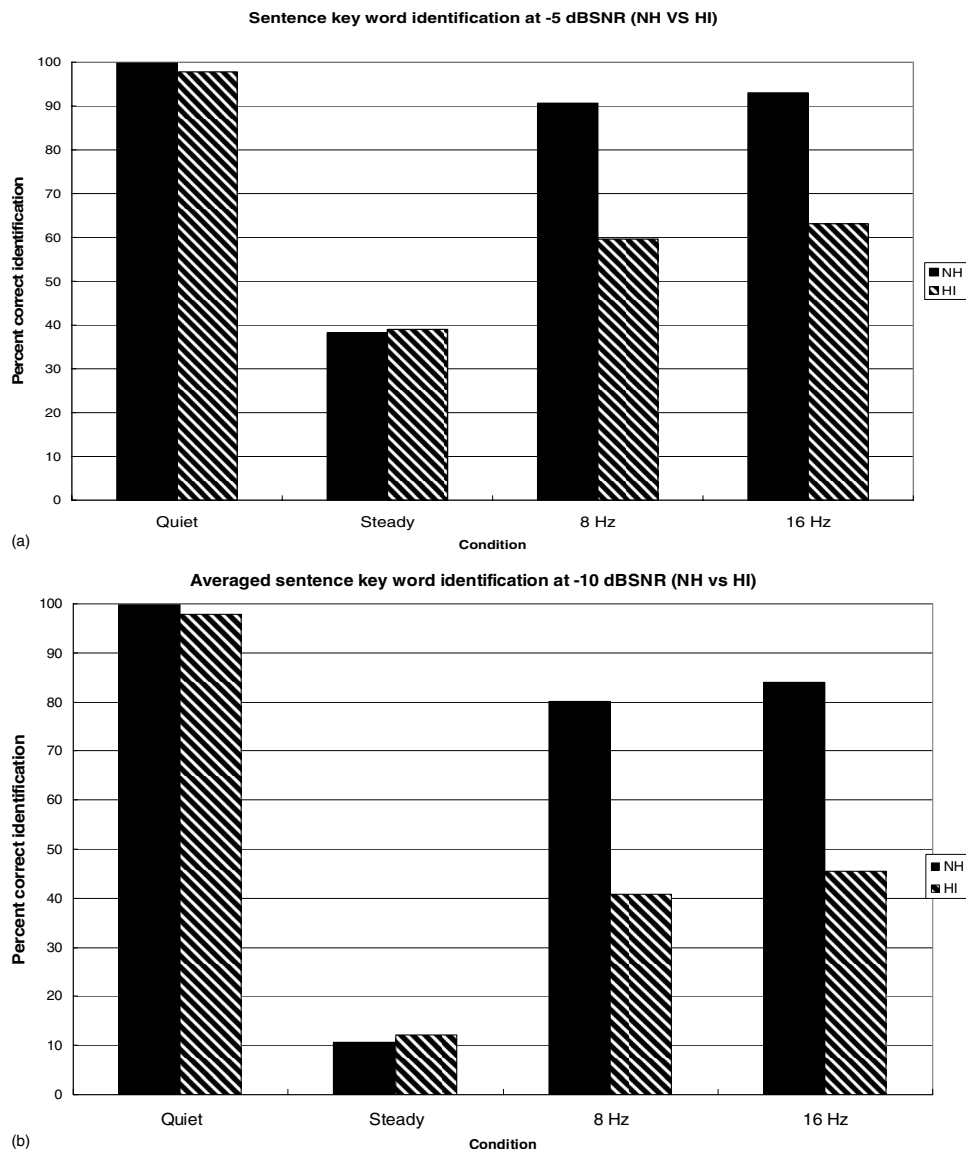


FIG. 1. [(a) and (b)] Summary of average percent correct keyword identification of HI and NH groups for sentence recognition at -5 and -10 dB SNR. No differences between performance of NH and HI listeners are seen for steady noise and quiet conditions, while significant differences are seen for the 8 and 16 Hz conditions. For more details, see Fig. 2 from Jin and Nelson (2006).

noise. Although this envelope is unlike most naturally occurring sounds, its use ensures that during the dips of the noise or the ‘on’ periods of the gated sentences, the signals are of nearly equal audibility [assuming based on Jin and Nelson (2006) that the effect of forward masking is minimal]. Thus we can compare across conditions.

Suprathreshold psychoacoustic abilities other than audibility, such as temporal masking and spectral resolution, have been proposed as possible contributors to understanding speech in fluctuating noise. For example, Peters *et al.* (1998) and Mackersie *et al.* (2001) suggested that HI listeners have reduced ability to use both spectral and temporal gaps in the background noise because of deficits in frequency selectivity as well as reduced audibility associated with hearing loss. Reduced MR is a decrease in the ability to listen in brief dips of noise, and it seems logical that this may be related to the temporal resolution abilities of HI listeners. Reduced MR could result from abnormal recovery from forward masking by the noise, thereby elevating thresholds for the signal in

the dips (Bacon *et al.*, 1998; Dubno *et al.*, 2002). Rhebergen *et al.* (2006), for example, included forward-masking thresholds in their extended speech-intelligibility index in order to account for speech performance in gated noise.

George *et al.* (2006) found that dip listening was related to reduced temporal processing, and not to spectral resolution abilities. They studied masking release in 29 listeners with sensorineural hearing loss, along with groups of NH and HI listeners listening through a simulated hearing loss. They tested listeners’

Speech Reception Threshold (SRT) for steady and fluctuating noises and measured masking release or benefit in terms of a decrease in SRT for the fluctuating noise. They tested spectral resolution abilities of listeners at 1 kHz (using the Speech Reception Bandwidth Threshold, SRBT) method of Noordhoek *et al.*, 2000). In addition they measured temporal resolution abilities using a novel Speech Reception Timewidth Threshold (SRTT) method, in which a threshold for temporal glimpse of speech is determined by adaptively

varying the duty cycle of the noise. They found that the benefit of fluctuating maskers was significantly related to the SRTT thresholds, but not to the SRBT measured at 1 k Hz. They concluded that masking release is related then to measures of temporal acuity and not to measures of frequency selectivity. This conclusion may be tempered somewhat by the finding that the SRBTs (spectral resolution thresholds) for all HI listeners fell within the 95% confidence intervals of the NH group, while the SRTTs (temporal resolution thresholds) of the HI listeners were significantly poorer than those of the NH group (shown in Fig. 6 of their paper). The SRTT task appears to be a measure of speech recognition in brief glimpses, rather than a pure measure of temporal resolution. Thus we believe that their overall finding (that MR is related to temporal, rather than spectral abilities) may not directly relate to other measures of spectral and temporal abilities.

In fact, [Jin and Nelson \(2006\)](#) noted weak correlation between abnormal forward masking and the amount of MR from HI listeners listening to sentence materials. Reduced temporal resolution seemed to be an incomplete explanation of poor sentence recognition of HI listeners in gated noise. As a result, there is a need to investigate the role of spectral processing further. Sensorineural hearing loss (SNHL) has been known to be associated with poorer-than-normal spectral resolution. Measures of auditory-filter bandwidths in HI listeners show considerable scatter and are only weakly related to the degree of hearing loss ([Glasberg and Moore, 1989](#); [Moore et al., 1999](#)). Some previous studies have found significant relationships between reduced frequency resolution and speech perception. For example, [van Schijndel et al. \(2001\)](#) and [Noordhoek et al. \(2000\)](#) observed that reduced frequency resolution and distorted coding of spectral information resulted in reduced speech perception in noise for listeners.

The effects of reduced spectral resolution on speech perception can be inferred from studies of cochlear-implant (CI) listeners and NH listeners presented with vocoder-processed stimuli that simulate cochlear implant processing (e.g., [Nelson et al., 2003](#)). Similar to HI listeners, CI users often experience difficulty understanding speech in noise. Even those users who can understand speech in quiet are adversely affected by noise at favorable signal-to-noise ratio (SNR) and have greatly reduced MR. It has been speculated that the limited spectral resolution of CI users might be responsible for this impairment, and this has been shown in NH listeners presented with processed signals so as to simulate speech processed through a cochlear implant. For example, [Qin and Oxenham \(2003\)](#) reported that implant simulation listeners have significantly less masking release when presented with speech in modulated versus steady noise. They compared normal-hearing listeners' speech recognition in steady speech-shaped noise, modulated speech-shaped noise, and single-talker speech interference under conditions of 4-, 8-, and 24-band implant simulations. They found that increasing the number of spectral channels significantly improved listeners' performance in modulated noise. Both [Qin and Ox-](#)

[enham \(2003\)](#) and [Nelson and Jin \(2004\)](#) show that increased numbers of vocoded spectral bands are critical for segregation of speech from complex noise.

Overall, then, there is convincing evidence that HI listeners experience reduced MR compared to NH listeners, but the factors that may influence this reduction are not fully understood. The current study investigates further spectral resolution factors that may affect the recognition of interrupted speech. It is hypothesized that audibility and temporal resolution are incomplete explanations for reduced masking release in HI listeners, and that the spectral resolution abilities of listeners with cochlear damage, like those with cochlear implants, are significantly related to speech recognition in interrupted sentences. To investigate this issue, the perception of gated speech and frequency resolution of HI listeners was examined and compared to NH listeners. It is hypothesized that reduced spectral selectivity associated with hearing loss might explain a significant proportion of the variance for sentence recognition in interrupted conditions.

II. METHODS

A. Participants

The same groups of NH and HI listeners from [Jin and Nelson \(2006\)](#) participated in the current study. The NH group consisted of eight listeners (20–25 years old) with thresholds no more than 20 dB HL across audiometric frequencies (250–8000 Hz). The HI group consisted of nine individuals (20–52 years old) with sensorineural hearing loss who had thresholds greater than 20 dB HL but no more than 70 dB HL between 250 and 4000 Hz. The individual thresholds for the HI group and the average thresholds for the NH group in terms of dB SPL are shown in Table I ([American National Standards Institute, 1996](#)). All of the participants were native speakers of English.

B. Stimulus preparation and procedures

1. Interrupted speech recognition

We defined the interrupted speech in this study as speech stimuli gated (or interrupted) with silent gaps rather than noise. For interrupted speech recognition, IEEE sentences spoken by five male and five female native speakers of English were used. Blocks of 10 sentences were presented with each block containing one sentence spoken by each talker. Sentences were presented in random order. Each sentence consisted of five key words and performance of each listener was calculated in terms of percent correct identification (PC). To minimize learning of the sentences, the lists of IEEE sentences that were used in the speech-recognition-in-noise task from [Jin and Nelson \(2006\)](#) study were excluded. The sentences in quiet were modified by using square gating with 4-ms \cos^2 ramp so that the speech sounds were on and off using a 50% gating cycle. Gate frequencies were 1, 2, 4, 8 and 16 Hz, resulting in regular bursts of speech that ranged in duration from approximately 31 ms (16 Hz gate frequency) to 500 ms (1 Hz gate frequency). No noise was present for this experiment. Depending on the gate frequency, whole or only parts of key words were available to listeners. For speech stimuli, we used the same method ap-

TABLE I. Audiometric hearing thresholds for the better ear for HI listeners and the right ear for NH in dB SPL, and overall speech and noise level (at -5 dB SNR) as a function of frequencies in Hz.

	500 (dB SPL)	1000 (dB SPL)	2000 (dB SPL)	3000 (dB SPL)	4000 (dB SPL)	6000 (dB SPL)	8000 (dB SPL)	Speech (Leq)
AVG								
NH	27	18	22	20	23	28	27.5	65.7
HI1	41	51.5	58.5	62.5	64	60	53	82
HI2	66	76.5	63.5	52.5	59	75	68	85
HI3	26	21.5	28.5	52.5	70	70	68	74.5
HI4	16	21.5	53.5	52.5	60	75	68	74
HI5	36	31.5	33.5	42.5	54	65	68	76
HI6	31	46.5	48.5	52.5	59	70	73	77
HI7	46	61.5	68.5	72.5	79	90	88	84
HI8	31	31.5	33.5	57.5	49	65	58	73.5
HI9	26	21.5	23.5	42.5	59	65	68	69.7

plied from previous study (Jin and Nelson, 2006) to compensate for the hearing loss for HI listeners. That is, for each HI listener, the overall level of the speech was adjusted based on his/her hearing sensitivity. For example, from Table I, HI1 has a mild-to-moderate hearing loss at mid to high frequencies for the tested ear. To amplify the level of speech and noise at the frequency range where HI1 has the most hearing loss but to maintain the level at other frequencies, speech and noise were passed through a Rane GE 60 graphic equalizer shaped to approximate the half-gain rule (Dillon, 2001) for each individual HI listeners. Then, the overall range of gated sentences were presented at approximately 40 dB SL *re* individual listener's pure tone average (except H2 and H8 due to the limitation of the equipment) with two blocks of IEEE sentences per gate frequency, for a total of 8 blocks that were presented in random order. The presentation level of the sentences was consistent at 65 dB SPL for all NH listeners. Table I also shows the overall speech level (Leq) for individual HI listeners and NH listener.

2. Frequency selectivity

Frequency selectivity was measured for individual listeners at 2000 Hz and 4000 Hz using the notched-noise method (using a symmetric notch with fixed-noise-level paradigm) of estimating auditory filter bandwidths and slopes (Moore and Glasberg, 1983; Patterson *et al.*, 1982; Stone *et al.*, 1992). Bands of noise were digitally generated with a mathematical software package (Matlab 6.0), mixed with signal tone (either 2000 or 4000 Hz) and played through Tucker-Davis Technology (TDT) hardware. All stimuli (both notched noise and signal tones) were 500 ms in duration. The level of the noise was 45 dB SL *re* individual's pure tone average at the signal tone for HI listeners, and 70 and 85 dB SPL for the NH group, approximately 45 and 60 dB SL, respectively. A three-interval forced-choice procedure described above was used to obtain thresholds for each condition. All of the intervals in any given trial contained samples of the band-stop noise. The cutoff frequencies were f_{lo} and f_{hi} for the low-pass (below the signal frequency) and high-pass noise (above the signal frequency) respectively. Two of the intervals contained the notched-noise only (standard intervals) and the other interval contained both the noise and a

2-kHz tone (or 4-kHz tone) centered in the spectral gap of the noise (target interval). The listener's task was to determine which of the intervals had the tone. Correct answer feedback was presented following the listener's response on each trial. Each subject had at least one hour of practice on the task before actual data collection began. For a given f_{lo} and f_{hi} , three blocks were presented to a listener to obtain three thresholds. If the standard deviation of the thresholds for a particular condition was greater than 2 dB, the same block was repeated until variance among any set of three thresholds was within 2 dB. Four sets of cutoff frequencies were selected, deriving g -values of 0.0, 0.1, 0.2, and 0.4 from the following formula.

$$g = |f - f_0|/f_0 = \Delta f/f$$

where f_0 is the center frequency and f is either the lower or upper cutoff frequencies. For instance, for $g=0.2$ and $f_0 = 2.0$ kHz, $f_{lo}=1.6$ kHz and $f_{hi}=2.4$ kHz

The g values used in the current experiment were suggested by Stone *et al.* (1992) who proposed simplified methods to measure auditory filter shapes and bandwidth. Thresholds for each tone measured with four different notched noises were then entered into a FORTRAN computer program which derives the rounded top exponential (roex) filter shape by using the following equation (Patterson *et al.*, 1982):

$$W(g) = (1 - r)(1 - pg)^* \exp(-pg) + r$$

where p is a parameter that determines both the bandwidth and the slope of the filter ends and r is a dynamic range limiter. This two-parameter filter allows good filter approximations to be made in a relatively short period of time. The roex (p) filter shape allowed us to estimate equivalent rectangular bandwidths (ERBs) and slopes of the auditory filters for 2-kHz and 4-kHz by using a formula suggested by Stone *et al.* (1992).

$$ERB = (2fc/p_u + 2fc/p_l)$$

where fc is the center frequency (either 2-kHz or 4-kHz) and p_u and p_l are upper and lower p values.

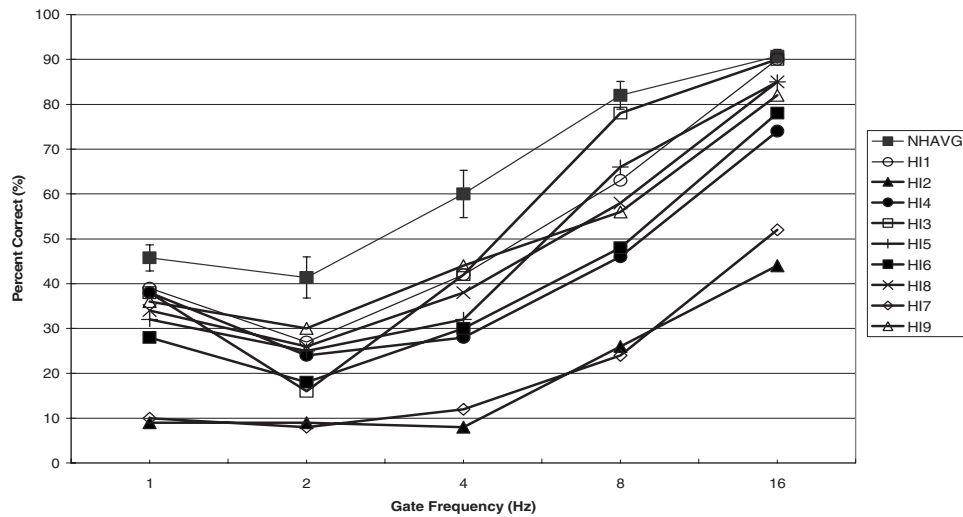


FIG. 2. The percent correct identification of individual HI listeners and NH group for recognition of interrupted sentence by silent gaps.

3. Stimulus presentation

All the stimuli were presented monaurally through an earphone (Telephonics, TDH-49P). For HI listeners, the test ear with thresholds between 40–60 dB HL for 2-kHz and/or 4-kHz was selected (except for HI3 and HI7 at 4-kHz). For NH listeners, the right ear was used as the test ear.

C. Analysis

Results of each experiment were entered into Analysis of Variance (ANOVA) with repeated measures to examine differences between groups (NH vs. HI) and main effects within a group. A stepwise linear regression was conducted to find which factors most contributed to the variance in performance for interrupted speech recognition. Interrupted-speech recognition scores in percent correct (PC) were correlated with ERBs and slopes of the auditory filters for 2 kHz and 4 kHz, and hearing sensitivity of individual participants. To compare speech intelligibility performance by individual listeners for speech interrupted by silent gaps and for speech in gated noise, the same analysis was applied to data from the previous study (Jin and Nelson, 2006) examining speech perception in gated noise.

III. RESULTS

A. Interrupted sentence recognition

Figure 2 displays the PC keyword identification of each HI listener along with average scores of the NH group. Both NH and HI listeners were able to identify fewer key words when sentences were gated at the slowest rates (1 and 2 Hz). For most of listeners (except HI2 and HI7), performance was poorest for 2-Hz gating. As the gating rate became faster (4 Hz or higher), the identification scores improved dramatically. The average scores of the NH group at each gate frequency were higher than the scores of HI listeners. At the fastest rate (16 Hz), the PC for four HI listeners (HI1, HI5, HI8 and HI9) was quite similar to that of NH listeners. A one-way ANOVA with repeated measures indicated that differences between groups were significant [$F(1,15) = 25.768, p = 0.000$] and gate frequency had significant effect

on performance [$F(4,60) = 139.398, p = 0.000$]. The interaction between group and gate frequency was also significant [$F(4,60) = 3.701, p = 0.009$].

To examine the relationship between sentence recognition in gated noise (Fig. 1) and interrupted sentence recognition the PC for both NH and HI listeners from each task in 8 Hz and 16 Hz were correlated. A strong correlation across listeners was found between the PC for these two recognition tasks ($0.8 < r < 0.9$) which was statistically significant ($p < 0.001$). When the keyword identification of sentences in gated noise and gated sentences was compared within the HI group only, the range of correlation coefficients was also high ($0.8 < r < 0.93$) and statistically significant ($p < 0.001$). Figure 3 shows the average percent correct identification for NH and HI listener groups for the different speech tasks, including sentence recognition in gated noise at -5 and -10 dB SNR as well as interrupted sentence recognition (speech interrupted by silent gaps).

B. Spectral resolution (tone detection with notched noise)

Results of tone detection in notched noise for NH and HI listeners are summarized in Tables II and III, respectively. The auditory filter characteristics, equivalent rectangular bandwidth (ERB) and slope of the filter (p), were obtained by using formulae suggested by Patterson *et al.* (1982). Several investigators suggested that the wide ERB of HI listeners might result from higher presentation levels compared to NH listeners (Dubno and Dirks, 1989). To examine the characteristics of auditory filters for both NH and HI groups at comparable noise presentation levels, the NH listeners were tested at two noise levels, 70 and 85 dB SPL. The latter level, especially, is close to the presentation level in dB SPL for some HI listeners (refer to Table II). Within the NH group, as the notched-noise level increased, the ERB appeared wider and the slope got shallower. A one-way ANOVA of NH data with the noise level as a within subject factor indicated that the presentation level of the noise had a significant effect on slope [$F(1,7) = 27.086, p = 0.000$] but not on ERB [$F(1,7) = 2.46, p = 0.127$].

Average PC identification for the perception of speech interrupted by either noise (at -5 & -10 dB SNR) or silent gaps

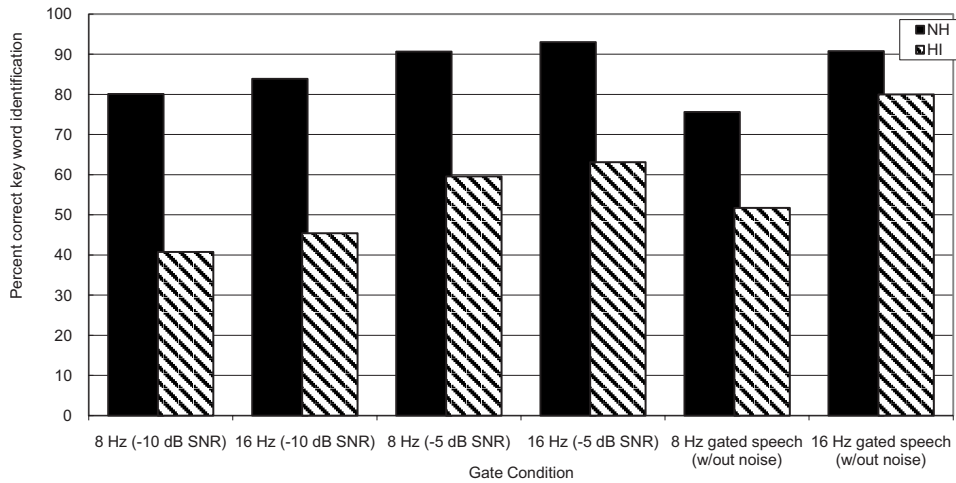


FIG. 3. The average percent correct identification of HI and NH groups for recognition of interrupted sentence by gated noise (-5 and -10 dB SNR) and silent gaps.

Compared to the results of the NH group when the noise was presented at 70 dB SPL, the critical bandwidths for HI listeners were greater and the slopes were shallower for both 2-kHz and 4-kHz tones. Within the HI group, those who had close to normal hearing sensitivity at the frequency of the signal tone (HI8 at 2-kHz, for example) showed relatively smaller ERB and higher p values than other HI listeners but not quite as good as NH listeners. When compared to performance of NH listeners at the 85 dB SPL noise level, the results for a few HI listeners (HI8 and HI9 at 2 kHz and HI5 at 4 kHz) were quite similar. However, statistical analysis (ANOVA) indicated that the auditory filter characteristics of NH listeners were significantly narrower with steeper slopes than HI listeners both for 2 kHz and 4 kHz even when the absolute levels of noise were similar for both groups. The average ERB values for HI listeners (for both 2 kHz and 4 kHz) were approximately 2.7 times to those for NH listeners when the level of noise was 70 dB SPL. When the noise level for NH listeners was 85 dB SPL, the ratio of the average ERB for HI listeners to that for NH listeners was 2.2. Since the auditory filter characteristics at different frequen-

cies were expected to be different and ERB and p values represent different aspects of auditory filters, group differences for each parameter were analyzed by using four separate simple ANOVAs. The results show that the differences in ERB and p values between NH (at 85 dB SPL) and HI listeners were significant for both ERB [$F(1, 15)=9.205, p=0.008$] and slope [$F(1, 15)=9.2394, p=0.008$] at 2-kHz signal tones. The group differences were also significant at 4-kHz for both ERB [$F(1, 15)=17.905, p=0.000$] and slope [$F(1, 15)=15.564, p=0.001$].

C. Statistical analysis

One of the main purposes of the current investigation was to determine the possible contributions of hearing sensitivity (audiometric pure tone thresholds) and suprathreshold psychoacoustic abilities such as spectral resolution (characteristics of the auditory filter) to perception of interrupted speech either by noise or silent gaps. Therefore, the results of spectral resolution and hearing sensitivity were correlated with both the performance of interrupted speech recognition

TABLE II. ERB and p values for 2 and 4 kHz tones from NH listeners.

	2 kHz				4 kHz			
	70 dB SPL		85 dB SPL		70 dB SPL		85 dB SPL	
	ERB	p	ERB	p	ERB	p	ERB	p
NH1	235.8	34.0	385.8	20.7	518.7	30.9	707.7	22.6
NH2	296.3	27.3	350.5	20.8	638.0	25.3	773.7	20.8
NH3	244.7	32.7	346.2	23.1	679.7	23.5	801.1	20.0
NH4	286.0	28.0	411.9	19.4	459.9	36.5	676.1	23.1
NH5	293.7	27.4	398.3	20.1	658.4	24.4	799.4	20.0
NH6	317.6	25.2	317.3	25.2	574.2	28.0	650.4	24.7
NH7	321.2	24.9	372.5	21.5	666.1	24.0	678.0	23.6
NH8	235.2	34.0	293.1	27.3	627.8	25.5	758.7	21.1
Avg.	284.96	28.50	355.69	22.49	614.87	26.74	733.91	21.90
SD	33.37	3.53	42.36	2.89	76.48	4.54	63.83	1.88

TABLE III. ERB and p values for 2 and 4 kHz tones from HI listeners.

	2 kHz		4 kHz	
	ERB	p	ERB	p
HI1	872.0	13.0	2113.2	7.8
HI2	1583.3	4.9	2898.8	5.4
HI3	568.3	14.1	1967.5	10.2
HI4	1330.0	9.5	1740.5	15.0
HI5	526.2	16.1	853.4	18.9
HI6	761.0	10.9	937.9	17.8
HI7	691.5	26.4	1496.5	11.7
HI8	382.2	21.5	1268.9	23.7
HI9	475.7	16.8	2037.1	7.8
Avg.	789.8	15.0	1650.1	13.8
SD	433.5	6.8	668.6	6.2

and that of speech perception in gated noise for HI listeners obtained from [Jin and Nelson \(2006\)](#). The reason why the results for NH listeners were excluded was because from the previous study, we found that the correlation was overwhelmingly affected by hearing sensitivity when the results of the NH group were included. Compared to the NH group whose hearing sensitivity and performance in different tasks were very uniform, HI listeners in the current study had various hearing thresholds across the audiometric frequencies and showed different performance levels. Therefore, if only the results for HI listeners were analyzed, it might be possible to observe the differential contribution of hearing sensitivity and suprathreshold abilities that affect the ability of HI listeners to understand interrupted speech.

Tables IV and V show the relationship between the PC scores for interrupted sentences, hearing sensitivity, and the performance in spectral resolution (ERB and slope, p) for the HI group. As shown in Table IV, the results from the recognition of speech interrupted by silent gaps at different frequencies were highly correlated with hearing thresholds at 0.5, 1 and 2 kHz but not with 4 kHz.

With the auditory filter shape, only the ERB at 2 kHz showed strong correlation with the perception of interrupted speech at higher rates (4, 8 and 16 Hz). As reported in the previous paper, the PC of speech in gated noise showed a relatively strong correlation with hearing sensitivity at 0.5, 1 kHz and 2 kHz regardless of gate frequency or SNR (Table

TABLE V. Correlation coefficient between the percent correct identification (PC) of gated sentence in noise at -5 and -10 dB SNR and Psychoacoustical measures for HI group

Correlation coefficient	-5 dB SNR		-10 dB SNR	
	8 Hz	16 Hz	8 Hz	16 Hz
Hearing sensitivity (0.5 kHz)	-0.6	-0.87	-0.72	-0.84
Hearing sensitivity (1 kHz)	-0.67	-0.6	-0.84	-0.88
Hearing sensitivity (2 kHz)	-0.6	-0.6	-0.62	-0.67
Hearing sensitivity (4 kHz)	-0.01	-0.1	-0.23	-0.25
ERB at 2 kHz	-0.5	-0.6	-0.53	-0.51
ERB at 4 kHz	-0.07	-0.62	-0.3	-0.4
Auditory filter slope (P) at 2 kHz	0.24	0.25	0.23	0.11
Auditory filter slope (P) at 4 kHz	0.01	0.5	0.37	0.41

V). The new analysis showed that ERBs at 2 and 4 kHz were substantially associated with the results from speech perception in gated noise at 16 Hz in -5 dB SNR.

The results of stepwise regression analysis conducted for PC identification scores of HI listeners for interrupted by either noise or silent gaps at 8 or 16 Hz are summarized in Table VI. The first column shows the speech tasks and the second column shows the predictor that was present in the regression equation of the stepwise analysis along with the variance (R^2) accounted for by the predictor. The last column displays their significance levels. The results of the stepwise regression showed that hearing threshold at 0.5 and 1 kHz were the strongest predictors of performance. In other words, hearing sensitivity at low-to-mid frequencies accounts for a substantial amount of the variance in the PC of interrupted speech. Once the hearing threshold factor had been removed, the auditory filter shape did not seem to play a role in interrupted sentence recognition even though there was a significant correlation between ERB and PC of interrupted sentence recognition (as shown Table IV and V).

IV. DISCUSSION

The previous report of [Jin and Nelson \(2006\)](#) showed that in this group of HI listeners, forward masking explained some of the variance in masking release (MR) performance, but only for CV identification. A significant amount of the variance in performance for sentence recognition in gated noise was unexplained by temporal masking abilities. In the current study we attempted to explain the remaining variance

TABLE IV. Correlation coefficient between the percent correct identification (PC) of interrupted sentence recognition and psychoacoustic measures for HI group.

Correlation coefficient	1 Hz	2 Hz	4 Hz	8 Hz	16 Hz
Hearing sensitivity (0.5 kHz)	-0.8	-0.62	-0.7	-0.6	-0.7
Hearing sensitivity (1 kHz)	-0.8	-0.7	-0.77	-0.73	-0.76
Hearing sensitivity (2 kHz)	-0.65	-0.6	-0.77	-0.7	-0.71
Hearing sensitivity (4 kHz)	-0.26	-0.5	-0.25	-0.2	-0.27
ERB at 2 kHz	-0.4	-0.4	-0.6	-0.6	-0.63
ERB at 4 kHz	-0.24	-0.26	0.2	-0.26	-0.4
Auditory filter slope (P) at 2 KHz	0.06	0.03	0.1	0.43	0.1
Auditory filter slope (P) at 4 KHz	0.25	0.29	0.18	0.2	0.35

TABLE VI. Summary of stepwise regression analysis for interrupted speech perception scores for HI listeners.

Speech task	Predictor	Significance
8 Hz at -5 dB SNR	HS 1000 Hz (0.58)	$p < 0.05$
16 Hz at -5 dB SNR	HS 500 Hz (0.75)	$p < 0.01$
8 Hz at -10 dB SNR	HS 1000 Hz (0.72)	$p < 0.01$
16 Hz at -10 dB SNR	HS 1000 Hz (0.78)	$p < 0.01$
8 Hz w/out noise	HS 1000 Hz (0.56)	$p < 0.05$
16 Hz w/out noise	HS 1000 Hz (0.58)	$p < 0.05$

by measuring spectral resolution abilities and the ability of listeners to perceive speech interrupted by silence, where forward masking is not an issue. Results from the current study indicate that much of the variance in MR for sentences was accounted for by low-mid frequency sensitivity of the HI listeners. Even though the auditory filter shape (ERBs at 2 and 4 kHz) showed strong correlations with the PC of interrupted speech recognition, once the contribution of hearing threshold was removed, ERB did not contribute significantly to the variance in speech recognition. It is possible that because hearing thresholds and spectral resolution are strongly associated with each other, when one factor was removed, another would not have significance in speech recognition. Another possibility is that since the auditory filter shapes were measured only at higher frequencies and that low-frequency thresholds were correlated with performance, those high-frequency ERBs were not a significant factor. The relationship between the auditory filter shape and interrupted speech recognition needs to be investigated more thoroughly in the future.

Previously a similar analysis of the contribution of forward masking suggested that temporal masking only accounted for 19% of the variance in performance (see Table IV from Jin and Nelson, 2006). Thus it appears that both temporal and spectral resolution are less significant factors than hearing threshold in explaining the lack of MR for sentence perception in HI listeners. In addition, listeners' identification of gated speech was significantly related to MR. It was hypothesized that MR might be related to auditory integration of speech information over gaps in the ongoing signal. Based on the current study, HI listeners must have audibility, access, and use of detailed spectral information in brief glimpses of a temporally varying signal. Jin and Nelson (2006) hypothesized that abnormal forward masking may affect the audibility of the initial segments of CV signals in the dips of the noise, but abnormal forward masking did not explain the majority of MR reduction experienced by HI listeners. Instead, HI listeners, in addition to having reduced audibility of signals in the dips of the noise, also have impoverished spectrotemporal detailed representation of speech in the dips of the noise, which is typically associated with hearing loss. That is, due to hearing loss, the amount of information in the dips of the noise is reduced, and HI listeners then need more glimpses of the reduced information to achieve comparable performance as NH listeners.

This is evident in the results of the investigation of HI listeners' recognition of gated speech. In that experiment, there is no masking noise present, so there is no recovery

from prior stimulation. Still, HI listeners generally showed reduced speech recognition relative to NH listeners. The performance of a few listeners' (HI 1, 3, 5, 8 and 9) at fast gate frequency (16 Hz) was close to that for NH listeners. For the other listeners, and for slower gate frequencies, however, performance of the HI group was significantly poorer than the NH group. This suggests that each glimpse of speech was impoverished and less informative to the HI listeners. One consequence of this proposed reduced information in the dips is that listeners do not integrate the speech information over time into a coherent speech stream. Reduced information in each dip results in less well established vowel formants, formant glides indicating consonants, and the like. This hypothesis is supported by the fact that at the faster gate frequencies, several HI listeners improved in performance and started to approach normal performance.

It might also be argued that the gated speech experiment presents a kind of modulation masking for the HI listeners. That is, the natural envelope of the speech material is disrupted by the gating. For example, Hedrick and Carney (1997) and Hedrick and Younger (2007) have shown that some HI listeners rely on envelope cues for speech recognition more than NH listeners do, presumably because of reduced spectral resolution. In addition, Lorenzi *et al.* (1997) studied three listeners with SNHL for their modulation masking, and found that these listeners had abnormally broad modulation filters. Either because of increased modulation masking, or simply because of increased reliance on speech envelopes, HI listeners might be expected to do poorly in gated speech because of envelope disruptions. This hypothesis remains to be tested.

It could also be argued that the HI listeners did not have fully restored audibility of the signal in the dips of the noise. The listeners were amplified to approximate a half-gain rule, and their recognition of quiet speech approached 100%. In addition, it is important to note that in this group of amplified listeners, their performance in steady noise at both -5 and -10 dB SNR was the same as NH listeners. This suggests even if full audibility was not achieved, these HI listeners were able to hear out speech information in steady noise as much as NH listener could. Otherwise greater SNR loss would be seen for the HI listeners. It also suggests that for young HI listeners with mild to moderate losses, improving audibility reduces SNR loss for steady noise, but does not restore normal MR. While we assume that the audibility of speech was not equivalent for the two groups of listeners, we have provided a test of MR using amplified stimuli that approach optimal amplification conditions. Even when amplification is functionally 'ideal' such that performance is matched to that of NH listeners for quiet and steady noise conditions, reduced MR persists. Further systematic evaluation of the role of audibility is warranted.

In summary, results from the current study suggest that HI listeners not only have reduced audibility of speech in the dips of modulated noise, but they also have reduced access to the detailed information in the dips of the noise. The fact that spectral resolution seems to contribute little beyond the contribution of reduced hearing sensitivity might be possibly due to the strong association between hearing loss and re-

duced spectral resolution. Results have implications for the potential success of noise reduction strategies in hearing aids. Slow-acting algorithms that reduce gain in noisy channels may in fact reduce the information about speech in the dips of the noise. Further investigations of the causes of reduced MR, detailed analysis of MR at varying levels of audibility, models of the effects, and the implications for hearing aids are needed.

ACKNOWLEDGMENTS

We gratefully acknowledge the suggestions of the editor and two anonymous reviewers. This work was partially supported by NIH R01 DC0083086.

American National Standards Institute (ANSI) (1996). American National Standard Specification for Audiometers. ANSI S3.6-1996, New York.

Bacon, S. P., Opie, J. M., and Montoya, D. Y. (1998). "The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," *J. Speech Lang. Hear. Res.* **41**, 549–563.

Bashford, J. A., Meyers, M. D., and Brubaker, B. S. (1988). "Illusory continuity of interrupted speech: Speech rate determines durational limits," *J. Acoust. Soc. Am.* **84**, 1635–1638.

Dillon, H. (2001). "Prescribing hearing aid performance," *Hearing Aids* (Thieme, New York), pp. 234–238.

Dubno, J. R., and Dirks, D. D. (1989). "Auditory filter characteristics and consonant recognition for hearing-impaired listeners," *J. Acoust. Soc. Am.* **85**, 1666–1675.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (2002). "Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing," *J. Acoust. Soc. Am.* **111**, 2897–2907.

George, E. L., Festen, J. M., and Houtgast, T. (2006). "Factors affecting masking release for speeding modulated noise for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **120**, 2295–2311.

Glasberg, B. R. and Moore, B. C. J. (1989). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech," *Scand. Audiol.* **32**, 1–25.

Hedrick, M. S. and Carney, E. A. (1997). "Effects of relative amplitude and formant transitions on perception of place of articulation by adult listeners with cochlear implants," *J. Speech Lang. Hear. Res.* **40**, 1445–1457.

Hedrick, M. S. and Younger, M. S. (2007). "Perceptual weighting of stop consonant cues by normal and impaired listeners in reverberation versus noise," *J. Speech Lang. Hear. Res.* **50**, 254–269.

Jin, S.-H. and Nelson, P. B. (2006). "Speech perception in gated noise: The effects of temporal resolution," *J. Acoust. Soc. Am.* **119**, 3097–3108.

Kochkin, S. (2007). "Increasing hearing aid adoption through multiple environmental listening utility," *Hear. J.* **60**, 28–31.

Lorenzi, C., Micheyl, C., Berthommier, CF., and Portalier, S. (1997). "Modulation Masking in Listeners with Sensorineural Hearing Loss," *J. Speech Lang. Hear. Res.* **40**, 200–207.

Mackersie, C. L., Prida, T. L., and Stiles, D. (2001). "The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentence by listeners with sensorineural hearing loss," *J. Speech Lang. Hear. Res.* **44**, 19–28.

Moore, B. C. J. and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **75**, 536–544.

Moore, B. C., Peters, R. W., and Stone, M. A. (1999). "Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips," *J. Acoust. Soc. Am.* **105**, 400–411.

Nelson, P. B., and Jin, S.-H. (2004). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **115**, 2286–2294.

Nelson, P. B., Jin, S.-H., Carney, A. E., and Nelson, D. A. (2003). "Understanding speech in modulated interference: Cochlear implant users and normal hearing listeners," *J. Acoust. Soc. Am.* **113**, 961–968.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (2000). "Measuring the threshold for speech-reception by adaptive variation of the signal bandwidth. II. Hearing-impaired listeners," *J. Acoust. Soc. Am.* **107**, 1685–1696.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (1982). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," *J. Acoust. Soc. Am.* **72**, 1788–1803.

Peters, R. W., Moore, B. C., and Baer, T. (1998). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," *J. Acoust. Soc. Am.* **103**, 577–587.

Qin, M. K., and Oxenham, A. J. (2003). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," *J. Acoust. Soc. Am.* **114**, 446–454.

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**, 3988–3997.

Stone, M. A., Glasberg, B. R., and Moore, B. C. J. (1992). "Simplified measurement of auditory filter shapes using the notched-noise method," *Br. J. Audiol.* **26**, 329–334.

van Schijndel, N. H., Houtgast, T., and Festen, J. M. (2001). "Effects of degradation of intensity, time, or frequency content on speech intelligibility for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **110**, 529–542.

Warren, R. M. (1984). "Perceptual restoration of obliterated sounds," *Psychol. Bull.* **96**, 371–383.