Phoneme recognition in modulated maskers by normal-hearing and aided hearing-impaired listeners

Sandeep A. Phatak[a)] and Ken W. Grant
Scientific and Clinical Studies Section, Audiology and Speech Center, Walter Reed National Military Medical Center, Bethesda, Maryland 20889

(Received 27 December 2010; revised 10 July 2012; accepted 12 July 2012)

This study measured the influence of masker fluctuations on phoneme recognition. The first part of the study compared the benefit of masker modulations for consonant and vowel recognition in normal-hearing (NH) listeners. Recognition scores were measured in steady-state and sinusoidally amplitude-modulated noise maskers (100% modulation depth) at several modulation rates and signal-to-noise ratios. Masker modulation rates were 4, 8, 16, and 32 Hz for the consonant recognition task and 2, 4, 12, and 32 Hz for the vowel recognition task. Vowel recognition scores showed more modulation benefit and a more pronounced effect of masker modulation rate than consonant scores. The modulation benefit for word recognition from other studies was found to be more similar to the benefit for vowel recognition than that for consonant recognition. The second part of the study measured the effect of modulation rate on the benefit of masker modulations for vowel recognition in aided hearing-impaired (HI) listeners. HI listeners achieved as much modulation benefit as NH listeners for slower masker modulation rates (2, 4, and 12 Hz), but showed a reduced benefit for the fast masker modulation rate of 32 Hz.
[http://dx.doi.org/10.1121/1.4742718]

PACS number(s): 43.71.Es, 43.71.Ky, 43.71.An, 43.66.Dc [EB] Pages: 1646–1654

I. INTRODUCTION

Speech recognition by normal-hearing (NH) listeners is known to be better in a fluctuating masker than in a steady-state masker (Miller and Licklider, 1950). This benefit is thought to be due to audible glimpses of the target speech during the low intensity portions of the modulated masker (Cooke, 2006). The magnitude of this benefit depends on the type (i.e., sinusoidal vs square-wave) and rate of masker modulations. Although the modulation benefit for NH listeners varies with the task (Buss *et al.*, 2009), several studies (e.g., Miller and Licklider, 1950; Howard-Jones and Rosen, 1993) have reported that, in general, the benefit is greatest when gaps in the masker envelope occur at a rate of ~10 Hz. Lorenzi *et al.* (2006b) observed a similar rate-dependent modulation benefit for consonant recognition, but it was less pronounced than that observed for word recognition. Although the past studies examined the modulation benefit for words and consonants, the effect of masker modulation rate on vowel recognition is not known. Vowels are generally longer, more intense and relatively stationary, compared to consonants. These distinct acoustic characteristics may lead to significant differences in the way masker modulations might affect the recognition of vowels, as compared to consonants.

The current study consisted of two experiments that measured consonant and vowel recognition scores in steady-state and sinusoidally modulated maskers using isolated nonsense syllables. In the first experiment, modulation benefit was measured in NH listeners at different signal-to-noise ratios (SNRs) and masker modulation rates. Comparing the results for consonants and vowels with those for words would reveal the relative contributions of these phonemes to the overall modulation benefit. This may be particularly helpful for developing signal enhancement strategies for improving speech recognition in fluctuating noise. An advantage of using nonsense syllables instead of words as test stimuli is that recognition performance is less influenced by top-down cognitive processing of linguistic context. This is useful when evaluating the effects of sensorineural hearing loss on speech recognition.

In the second experiment, modulation benefit was measured in aided hearing-impaired (HI) listeners. A number of studies have demonstrated that HI listeners obtain significantly less benefit from masker fluctuations than NH listeners (Festen and Plomp, 1990; Bacon *et al.*, 1998; Lorenzi *et al.*, 2006b). This is most likely because unaided HI listeners are less able to take advantage of the low-intensity portions of noise because the target speech unmasked during this period is partially inaudible to them (Grose *et al.*, 2009). Bacon *et al.* (1998) demonstrated that elevating hearing thresholds in NH listeners by noise masking decreases their modulation benefit. On the other hand, when the loss of audibility is compensated for using a frequency-dependent gain, the modulation benefit for HI listeners increases, but is still less than that obtained by NH listeners (Peters *et al.*, 1998). Both Bacon *et al.* (1998) and Peters *et al.* (1998) used sentences from the Hearing In Noise Test (Nilsson *et al.*, 1994) to measure modulation benefit. Jin and Nelson (2006) found that aided HI listeners exhibit a deficit in modulation benefit not only for words, but also for consonants. However, whether such a deficit exists for vowel recognition in aided HI listeners is not known. The data collected in the second

[a)]Author to whom correspondence should be addressed. Electronic mail: s.a.phatak@gmail.com

 0001-4966/2012/132(3)/1646/9/$30.00

experiment will allow us to compare modulation benefits in NH (from the first experiment) and aided HI listeners for vowel recognition at several masker modulation rates and SNRs.

II. EXPERIMENT I

A. Methods

1. Listeners

Six NH listeners (three male and three female) in the age range of 30–60 years participated in this experiment. All listeners except one had audiometric thresholds less than ≤20 dBHL for frequencies between 250 Hz to 8 kHz. One listener had slightly elevated thresholds (i.e., 25–30 dBHL) at 6 and 8 kHz in both ears. The experiment consisted of two tasks measuring consonant recognition and vowel recognition, respectively. Four of the six listeners (one male, three female) completed both tasks. Of the remaining two listeners, one completed only the vowel recognition task, whereas the other completed only the consonant recognition task. Thus, each task had five NH listeners (two males, three females). The average age of listeners in consonant and vowel recognition tasks was 48.8 and 44.2 years, respectively. All listeners provided written informed consent before participating. The study was carried out in accordance with the regulations and ethical guidelines on experimentation with human subjects set forth by the Institutional Review Board and Human Use Committee, Walter Reed Army Medical Center, and Walter Reed National Military Medical Center.

2. Signals

Speech stimuli for the consonant recognition task were isolated /ɑ/C/ɑ/ syllables spoken by a female talker, with 18 consonants [C = /p/, /t/, /k/, /f/, /θ/, /s/, /ʃ/, /b/, /d/, /g/, /v/, /ð/, /z/, /ʒ/, /m/, /n/, /ʧ/, /ʤ/] (Grant and Walden, 1996). Four utterances of each /ɑ/C/ɑ/ syllable were used. Speech stimuli for the vowel recognition task were /h/V/d/ syllables with nine vowels (V = /i/, /ɪ/, /ɛ/, /æ/, /o/, /ʊ/, /u/, /ʌ/, /ɝ/), spoken by one female and two male talkers (Fu and Galvin, 2006). Eight tokens of each /h/V/d/ (four from the female and two from each male talker) were used. Thus, there were a total of 72 tokens in each task. All stimuli were digitally stored and played back at a sampling rate of 24.414 kHz via a Tucker-Davis Technology system III controlled by a computer.

The left-hand panel in Fig. 1 shows average power spectra for the /ɑ/C/ɑ/ syllables calculated over the entire syllable (aCa), over the consonant region that includes formant transitions (C + FT), and over the consonant region without formant transitions (C − FT). The C − FT spectrum has less energy between 500 Hz and 2 kHz compared to the other two spectra due to the absence of formant energy. The right-hand panel in Fig. 1 shows average power spectra for h/V/d/ syllables calculated over the entire syllable (hVd) and over the vowel region (V) that includes formant transitions. Note that the two spectra are very similar.

The masker spectrum for the vowel recognition task was matched to the average /h/V/d/ spectrum so as to obtain

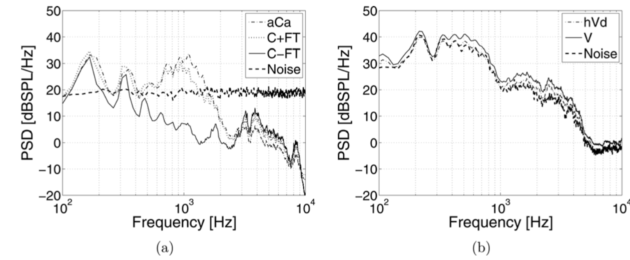

FIG. 1. Power spectral densities for (a) consonant stimuli at 60 dB SPL and (b) vowel stimuli at 65 dB SPL, calculated over entire tokens (aCa, hVd) and over specific regions (C − FT, C + FT, V; see the text for explanation). Spectral densities of the noise masker for consonant and vowel recognition tasks are represented by thick dashed lines for a 0 dB SNR.

uniform masking across frequency, which would presumably yield less variation in scores across individual vowels. In the case of consonant recognition, however, individual consonant scores have been reported to be more similar in a white-noise masker than in a speech-spectrum noise masker (Phatak and Allen, 2007). Therefore, the masker spectrum for the consonant recognition task was chosen to be white. Masker spectra for the two tasks are indicated by thick dashed lines in Fig. 1. Maskers were sinusoidally amplitude modulated (SAM) with 100% modulation depth at 4, 8, 16, and 32 Hz for the consonant recognition task, and at 2, 4, 12, and 32 Hz for the vowel recognition task. Additionally, a steady-state (SS) noise condition (i.e., no modulation) was included for both tasks. A new noise sample was generated before each presentation and the starting phase of the modulating sinusoid was randomized. The masker started 100 ms before the speech. Both speech and masker had a 30 ms raised-cosine ramp at onset and offset.

Consonant and vowel tokens were presented in quiet and in the presence of five types of noise maskers (4 SAM + 1 SS). The SNR was set using root mean square (rms) values of speech and noise samples. In the consonant recognition task, listeners were tested at −18, −12, −6, and 0 dB SNR. In the vowel recognition task, SAM maskers were presented at −24, −18, −12, and −6 dB SNR. Vowel recognition scores in the SS masker were significantly lower than those in any of the four SAM maskers. Therefore, SNRs for the SS masker were set 6 dB higher than the SAM masker SNRs (i.e., from −18 to 0 dB).

3. Procedure

Stimuli were presented diotically through Sennheiser HD-580 circumaural headphones. The speech signal was fixed at 60 and 65 dB SPL (Linear) for consonant and vowel recognition tasks, respectively. The lower presentation level in the consonant task was required in order to restrict the loudness of the white noise masker at low SNRs.

Listeners sat in front of a touch-screen monitor in a sound-treated room. After listening to a token in the consonant task, listeners entered their response by choosing from 18 buttons on the touch-screen, each labeled with a consonant. Similarly, in the vowel task, listeners chose from 9 vowel buttons on the touch-screen. Both tasks were self-paced. Listener responses were stored on a computer and were analyzed using MATLAB routines.

Before starting each task, listeners were trained in quiet to familiarize themselves with the talkers' voices and to associate the consonant and vowel sounds with the corresponding button labels. Each listener was then trained with 12 Hz SAM noise and the steady-state noise at one SNR (-12 dB for consonants and -18 dB for vowels). During the training session, visual feedback was provided following each response by highlighting the correct option on the touch-screen. No feedback was provided during testing sessions.

Each test block consisted of 72 tokens, presented in random order. The noise masker and SNR conditions were randomized across blocks. Each test run consisted of 21 such blocks (5 noise types $\times$ 4 SNRs + 1 quiet condition.). There were three test runs per listener in the consonant task and two tests runs in the vowel task. The total time per listener, including training, was $\sim$10 h. Listeners were tested in multiple 1–2 h sessions. Monetary compensation was provided on an hourly basis at the completion of each task.

B. Results

1. Recognition scores

Figure 2 shows consonant and vowel recognition scores as a function of SNR in each type of masker. The percent correct scores were converted to rau scores (Sherbecoe and Studebaker, 2004) in order to minimize ceiling effects, especially in the case of vowel recognition. As talker-sex differences did not affect vowel scores, vowel data were combined across talkers. The effect of the masker modulation rate on recognition scores was noticeably different for consonants and vowels. Consonant scores [Fig. 2(a)] in modulated maskers, although greater than those in the SS masker, were not different from each other, whereas vowel scores [Fig. 2(b)] showed a clear dependence on the modulation rate. Vowel scores in 2, 4, and 12 Hz maskers were different from each other, and from those in the SS masker condition, for SNRs < -6 dB. However, vowel scores in the 32 Hz masker showed higher slope (in rau/dB) than those for other modulated maskers. A four (modulation rates) $\times$ four (SNRs) repeated-measures analysis of variance (ANOVA) was conducted on consonant and vowel scores in the modulated masker conditions. Consonant scores had a significant main effect of SNR [$F(3,12) = 385.0$, $p < 0.001$], but the main effect of modulation rate was not significant [$F(3,12) = 2.2$, $p = 0.136$]. Vowel scores had significant main effects of both SNR [$F(3,12) = 386.2$, $p < 0.001$] and modulation rate [$F(3,12) = 132.5$, $p < 0.001$], and a significant interaction between the two [$F(9,36) = 19.6$, $p < 0.001$]. This analysis illustrates that the rate of masker modulations affected vowel recognition, but not consonant recognition. The variations in recognition scores of individual vowels with modulation rate were similar to that for the average vowel scores shown in Fig. 2(b). No consistent trend in rate dependence was observed across individual consonant scores.

2. Benefit of masker modulations

The benefit of masker fluctuation is commonly reported in terms of a score improvement, i.e., the modulation masking release (MMR). The MMR is defined as the recognition score in a modulated masker minus that in a steady-state masker, at a given SNR. Figure 3 shows the MMR for consonants and vowels. In general, the MMR increased with a decrease in the SNR. Above 0 dB SNR (not displayed in Fig. 3), no significant MMR was obtained. Figure 2(b) shows that at -18 dB SNR, the SS scores for vowels were close to the chance performance level, while for SNR ≥ -12 dB, scores in 4, 12, and 32 Hz SAM maskers were close to the ceiling. Although this may result in underestimation of the vowel MMR, vowel MMR values were clearly greater in magnitude than the consonant MMR at SNR ≤ -12 dB. The variation in vowel MMR with masker modulation rate was also greater than that in consonant MMR. The maximum vowel MMR was observed in the 12 Hz SAM masker condition.

3. Comparison with other studies

Figure 4(a) compares the consonant MMR values for NH listeners obtained in the present study with those reported by Füllgrabe *et al.* (2006) and Lorenzi *et al.* (2006b) using SAM maskers. Füllgrabe *et al.* (2006) measured consonant recognition at -15 dB SNR, whereas Lorenzi *et al.* (2006b) used listener-specific SNRs, ranging between -9 and -15 dB. Both studies obtained NH consonant MMRs that were close to those observed in the current study in a comparable SNR range (i.e., -12 and -18 dB). It is important to note that the masker spectrum in the current study was white, whereas Füllgrabe *et al.* (2006) and Lorenzi *et al.* (2006b) used speech-spectrum maskers. Thus, at least for

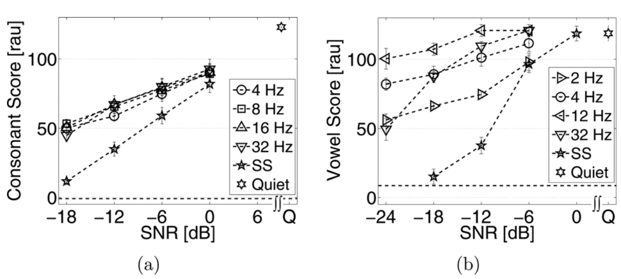

FIG. 2. (a) Consonant recognition scores and (b) vowel recognition scores, in rau, as a function of SNR and in the quiet condition (Q). Symbols and error bars represent the mean and $\pm$1 standard deviation across listeners in each condition. Horizontal dashed lines indicate chance performance (i.e., 5.56% = -0.88 rau for consonants and 11.11% = 8.59 rau for vowels).

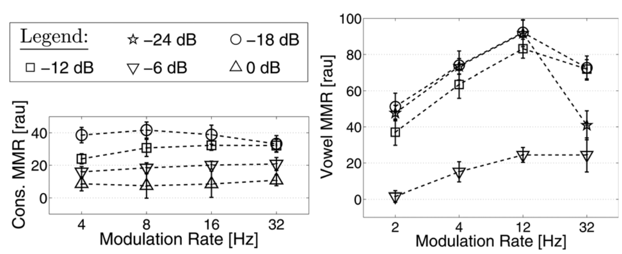

FIG. 3. MMR, i.e., recognition scores in SAM noise minus the score in SS noise, obtained by NH listeners for consonants (left) and vowels (right). Symbols and error bars represent the mean and $\pm$1 standard deviation, respectively, in each condition. Vowel MMR values at -24 dB SNR were estimated by assuming the SS scores to be at the chance level (i.e., 8.59 rau).

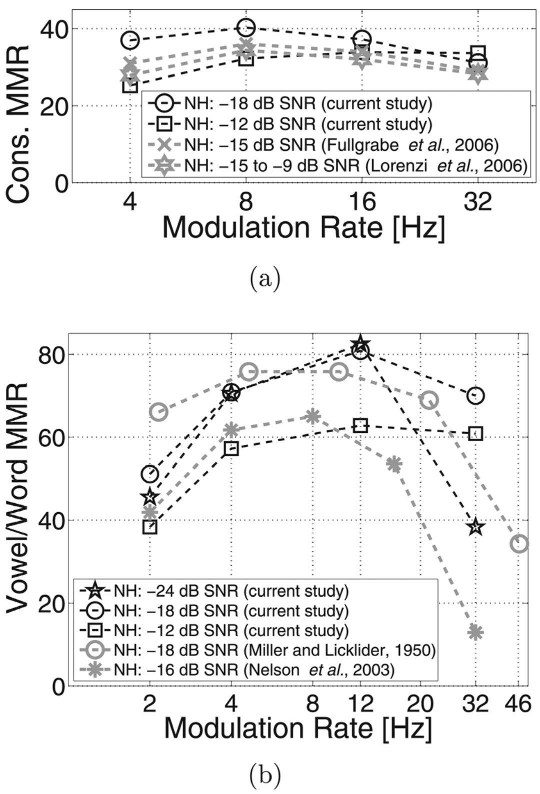

FIG. 4. The MMR in percentage points, observed in the present study (black) and in other published studies (gray). The vowel MMR at −24 dB SNR in the current study was estimated by assuming the SS score to be at the chance performance level. Miller and Licklider (1950) data points were estimated from the published figures.

these two maskers, consonant MMR patterns do not seem to depend on the masker spectrum.

To the best of our knowledge, there are no published results regarding the effect of masker modulation rate on vowel perception. However, past studies of masking release for word recognition (Miller and Licklider, 1950; Nelson *et al.*, 2003) reported results similar to the vowel data obtained in the current study. Figure 4(b) shows the word recognition MMR for NH listeners at −18 dB SNR from Miller and Licklider (1950) and at −16 dB SNR from Nelson *et al.* (2003), both measured with square-wave gated maskers (50% duty cycle). Miller and Licklider (1950) used phonetically balanced words, whereas Nelson *et al.* (2003) used IEEE sentences. These word MMR values were more similar to the vowel MMR than the consonant MMR in the current study, in both magnitude and rate dependence. For example, word MMRs showed high overall magnitudes and a sharp decrease for modulation rates greater than 20 Hz, similar to that observed in the vowel MMR at −24 dB in the current study.

III. EXPERIMENT II

The main purpose of this experiment was to determine whether HI listeners can derive the same degree of modulation benefit and the same dependence on modulation rate as do NH listeners, once the target speech is made audible. Jin and Nelson (2006) observed lower-than-normal modulation benefit in HI listeners even after providing gain. However, that study did not test vowel recognition, and the number of modulation rates and SNR conditions were too few to draw any conclusion regarding the dependence of modulation benefit on the masker modulation rate. To expand on the results of previous studies, experiment II measured vowel recognition by aided HI listeners in sinusoidally modulated maskers at several SNRs and modulation rates.

A. Methods

Five HI listeners (three male, two female) with bilateral hearing loss completed the vowel recognition task described in experiment I. The average age for HI listeners was 71.8 years (range: 58–89 years). Because of the large age difference between NH listeners (from experiment I) and HI listeners, another group of four elderly listeners with normal-to-near-normal hearing (ENNH) was also recruited. These ENNH listeners were 65–68 years old, with an average of 66.8 years. To achieve a better age-match between HI and ENNH groups, data from an 89 year old HI listener was omitted. As this listener's recognition scores were close to the average scores for the HI group, excluding his data had no noticeable effect on recognition scores. A comparison of ANOVA results with and without this listener's data revealed that excluding his data marginally altered significance values, but not enough to cross the threshold of statistical significance (i.e., $p = 0.05$) for any test. The average age for the remaining four HI listeners (two male, two female) was 67.5 years (range: 58–79 years). Figure 5 compares means and standard errors of pure-tone thresholds averaged across listeners within each listener group.

Individualized linear gain was provided to each ear of HI and ENNH listeners[1] according to the Cambridge formula for moderate speech levels (Moore and Glasberg, 1998).[2] The gain was applied after combining speech [at 65 dB sound pressure level (SPL)] and noise (at the required SNR). At low SNRs, noise levels after the required amplification for HI listeners were not only high (>100 dB SPL), but also resulted occasionally in signal clipping. Therefore, SNRs for HI listeners were set 6 dB higher than those for NH listeners from experiment I (i.e., −18 to 0 dB for SAM maskers and −12 to +6 dB for the SS masker). The SNRs for ENNH listeners

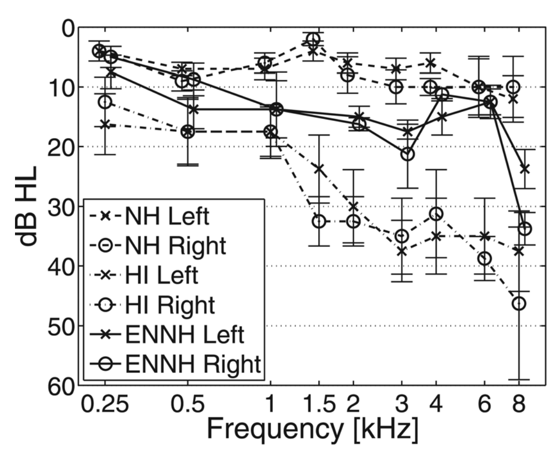

FIG. 5. Mean and ±1 standard errors for audiometric thresholds in left and right ears of NH, HI, and ENNH listeners.

were the same as those presented to NH listeners in experiment I.

B. Results

1. Recognition scores

Figure 6 shows vowel recognition scores (in rau) of NH (dashed lines) and HI (solid lines) listeners for each masker type. Data for NH subjects are re-plotted from Fig. 2(b). The average scores in SS, 2, 4, and 12 Hz modulation conditions were practically identical for NH and aided HI listeners. A repeated-measures ANOVA was conducted on scores at three common SNRs (viz., −18, −12, and −6 dB), with modulation rate and SNR as within-subject factors and listener group as the between-subject factor. The analysis revealed significant main effects of modulation rate [$F(3,21) = 260.8$, $p < 0.001$] and SNR [$F(2,14) = 372.9$, $p < 0.001$], but no significant main effect of listener group [$F(1,7) = 0.9$, $p = 0.363$]. Interactions between rate and SNR [$F(6,42) = 24.6$, $p < 0.001$], between rate and listener groups [$F(3,21) = 5.5$, $p = 0.006$], and the three-way interaction among rate, SNR and groups [$F(6,42) = 3$, $p = 0.017$] were significant. *Post hoc* unpaired *t*-tests indicate that the difference between the two groups was statistically significant only for the 32 Hz, −18 dB SNR condition [$t(7) = 8.2$, $p < 0.001$]. These data indicate that for modulation rates of 12 Hz and below, elderly HI listeners provided with sufficient gain perform identically to younger NH listeners.

2. Modulation benefit

Figure 4(b) plots the modulation benefit for vowel recognition in NH listeners in terms of the MMR, which is the difference between rau scores in modulated and SS maskers. The vowel recognition data in Fig. 6 show that listeners were at chance performance for the SS condition at ∼−18 dB. Conversely, performance was near ceiling at an SNR of −12 (12 Hz masker) to 0 dB (SS masker). Because of these ceiling and floor effects, it is not possible to accurately evaluate the modulation benefit in terms of a score-difference metric like the MMR. To circumvent this problem one can also express the modulation benefit in terms of the SNR difference (in dB) between the SS and modulated masker conditions for a given level of performance. To differentiate this measure of modulation benefit (in decibel difference) from a measure based on the percent correct (or rau) difference, it is termed hereafter as MMR_{dB}.

Referring to Fig. 6, the MMR_{dB} for NH listeners between SS and the 32 Hz condition for a rau of ∼80 is ∼11 dB. Similarly, the MMR_{dB} for HI listeners for this same level of performance is ∼8 dB. The result of this procedure for all NH, ENNH, and HI listeners at many levels of performance (55–95 rau) is plotted in Fig. 7. MMR_{dB} at specific rau values were obtained by piece-wise cubic interpolation of recognition score (in rau) vs SNR functions for each listener.

Figure 7 shows that, on average, the NH group obtained more modulation benefit and the HI group obtained less modulation benefit than the ENNH group. A repeated-measures ANOVA of these MMR_{dB} values revealed significant main effects of rau score [$F(4,40) = 36.4$, $p < 0.001$] and listener group [$F(2,10) = 187.9$, $p = 0.031$], and a significant interaction between the two [$F(8,40) = 3.2$, $p = 0.007$]. After Bonferroni correction, *post hoc* tests revealed that the difference between NH and HI group was significant ($p = 0.031$), but the ENNH group was not significantly different from the other two groups ($p > 0.05$). The fact that ENNH listeners obtained a relatively lower average MMR_{dB} than NH listeners does not rule out aging as a contributing factor for the reduced benefit. A partial correlation analysis was conducted on the MMR_{dB} values at 55–95 rau scores for 13 listeners (5 NH, 4 ENNH, and 4 HI) with two input variables—age and the Speech Intelligibility Index (SII), which is a measure of speech audibility (ANSI, 1997). When SII was controlled, correlations between MMR_{dB} and age were small ($R^2 < 0.13$) and not significant ($p > 0.69$). Similarly, when age was controlled, partial correlations between MMR_{dB} and SII were low ($R^2 < 0.36$) and barely significant for rau scores of 65–95 ($0.041 \le p \le 0.048$), but not significant at 55 rau ($p = 0.07$). These results suggest that the reduced MMR_{dB} observed in HI listeners is more likely due to auditory processing deficits other than audibility loss and age-related deficits.

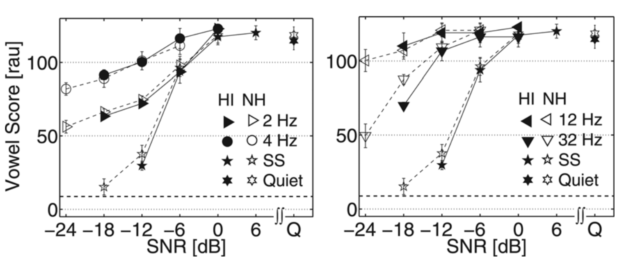

FIG. 6. Vowel recognition scores as a function of SNR and in quiet condition (*Q*), for NH (dashed lines, open symbols) and aided HI (solid lines, filled symbols) listeners. Four modulation rates are divided into two panels to avoid visual cluttering. Scores in quiet and in the SS masker are plotted in both panels for reference. Symbols and error bars represent mean and ±1 standard deviation across listeners in each condition. Horizontal dashed lines indicate chance performance.

IV. SUMMARY AND DISCUSSION

There were two main experimental questions in this study—*does the rate of masker fluctuations affect consonant and vowel recognition differently?* (experiment I) and, *do aided HI subjects demonstrate the same degree and rate dependence of modulation benefit as NH subjects?* (experiment

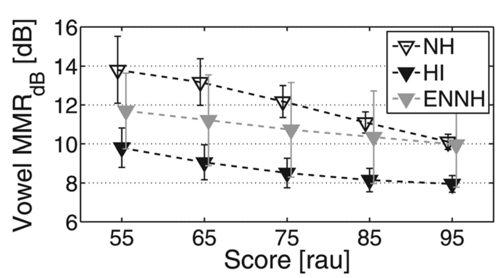

FIG. 7. MMR_{dB} as a function of the recognition score, for NH, ENNH, and HI listeners. Symbols and error bars represent mean and ±1 standard deviation within each listener group.

II). Experiment I compared consonant and vowel recognition in NH listeners, whereas experiment II compared vowel recognition in NH and aided HI listeners. A third group of listeners (ENNH) was tested in experiment II in an attempt to isolate the effects of aging and hearing loss on modulation benefit.

A. Consonants vs vowels

Results from experiment I illustrate that the masker modulation rate affected vowel recognition more than it affected consonant recognition. Vowel recognition not only showed greater benefit from masker fluctuations than consonant recognition, but also showed a greater dependence on the rate of modulation. One important difference between the two tasks was in the masker spectrum, i.e., white noise for consonant recognition and speech-spectrum noise for vowel recognition. As noted in Sec. II B 3, this difference in the masker spectrum did not affect consonant MMR. In addition, pilot tests indicated that there were no statistically significant differences between vowel scores in white noise and in speech-spectrum noise (see the Appendix). Thus, the differences in consonant and vowel MMR patterns were not due to different masker spectra.

The higher rate dependence of vowel MMR is most likely due to differences in temporal durations of consonant and vowel cues. For example, the stop release burst, which is the primary cue for identifying stop plosives (Blumstein and Stevens, 1980), is a temporally short feature that lasts ∼10–20 ms. The left panels in Fig. 8 show auditory spectrograms of an /ɑtɑ/ token in quiet (top) and in the presence of a 16 Hz modulated masker at −12 dB SNR (bottom). Unmasking of this short release burst, which is critically important for recognizing the consonant, does not depend on the masker modulation rate as long as the gap between masker peaks is long enough to unmask the burst. In other words, the temporal location of such short consonant cues, relative to the low intensity portions of the masker envelope, determines whether the consonant can be recognized successfully, rather than the masker modulation rate.

The primary cues for vowel identification are formants, which are relatively stationary over time. The right panels in Fig. 8 show auditory spectrograms of an utterance of /hæd/ presented in quiet and in a 12 Hz modulated masker at −18 dB SNR. Even at −18 dB SNR, many pitch periods of the vowel formants are always unmasked, irrespective of the exact temporal location of the formats relative to the low intensity portions of the masker envelope. As the masker modulation rate increases from 2 to 12 Hz, the probability of masking the entire vowel formant region decreases, resulting in an increase in the vowel MMR. The reduction in vowel MMR above 12 Hz could potentially be due to two factors. First, as the masker modulation rate increases, the temporal gaps become shorter, which may make the estimation of formant frequency more difficult. Second, forward masking from the preceding masker peak may persist into the gap region and effectively "fill in" the gap making it difficult to glimpse vowel formant cues (Wilson and Carhart, 1971).

The consonant MMR observed in the current study was consistent with previously published data for NH listeners [Fig. 4(a)]. The MMR for words from past studies were much closer to the MMR for vowel than to the MMR for consonant, in both magnitude and rate dependence, which suggests that the masking release obtained for words could be dominated by unmasking of vowels. These findings are consistent with the observation of Diehl *et al.* (1987) that under adverse listening conditions, vowel sounds form "islands of reliability" due to their intense and relatively stationary formants. This may be one reason why some studies found that there is relatively more speech information in vowel sounds than in consonant sounds (Kewley-Port *et al.*, 2007).

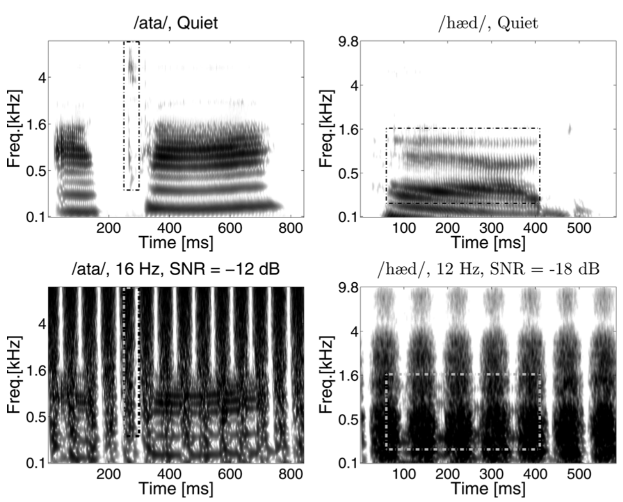

FIG. 8. Auditory spectrograms of a consonant token (/ɑtɑ/, left) and a vowel token (/hæd/, right) in quiet (top) and in the presence of a modulated masker (bottom). Spectrograms were calculated using a bank of 200 gammatone filters (Slaney, 1998). The grayscale intensity is proportional to the logarithm of filter output. Dashed-dotted rectangles indicate the time-frequency location of the release burst in consonant /t/ and formants in vowel /æ/ in left and right panels, respectively.

B. Amplification, audibility, and SNR loss

Vowel recognition data in experiment II showed that aided HI listeners obtained as much modulation benefit (MMR_{dB}) as NH listeners for the modulation rates of up to 12 Hz. In contrast, Jin and Nelson (2006) found aided HI listeners to have less modulation benefit than NH listeners for both consonant and word recognition in 8 and 16 Hz square-wave modulated maskers. Differences in the masker modulation type (i.e., square vs sinusoidal) may be partially responsible for differences in results across the two studies. In a modulated masker, energy from the masker peak spreads into subsequent masker valley due to forward masking (Wilson and Carhart, 1971). This spread of energy would be more prominent in the case of a square-wave modulated masker, compared to a sine-wave modulated masker, due to more abrupt peak-to-valley transitions. Thus, HI listeners, who have greater susceptibility to forward masking than NH listeners (Kidd *et al.*, 1984), are likely to obtain relatively lower modulation benefit in square-wave modulated maskers.

A second and more prominent difference between the two studies is in speech stimuli (i.e., consonants/words vs vowels). Jin and Nelson contended that the reduced modulation benefit for aided HI listeners in their study was because their linear gain amplification strategy (half-gain rule) was not sufficient to make the entire dynamic range of speech audible to HI listeners during noise gaps. However, linear gain was able to provide sufficient audibility to vowel stimuli in the current study, as demonstrated in the following analysis.

Figure 9 plots the amplified vowel spectrum presented to the HI listener with the lowest speech audibility (i.e., lowest SII) in the current study, along with the listener's audiometric thresholds (in dB SPL). To obtain this spectrum, the average vowel spectrum (V) from Fig. 1(b) was adjusted according to the gain prescription for this lowest-audibility listener. The amplified vowel spectrum was above audiometric thresholds for frequencies up to 4 kHz. Thus, the most important vowel cues, i.e., first and second formants (Hillenbrand *et al.*, 1995), were audible to all aided HI listeners in the current study. A similar spectral analysis was done for the consonant stimuli in the current study, but with the presentation level (70 dB SPL) and the amplification strategy (half-gain rule) of Jin and Nelson (2006) study. The spectral analysis revealed that even in for the best-audibility (i.e., highest SII) HI listener in the current study, the amplified consonant spectrum would be below or barely above the audiometric thresholds for most of the signal bandwidth. This corroborates Jin and Nelson's claim that the entire dynamic range of speech, especially that of consonants, was perhaps not audible to their aided HI listeners.

It is important to verify that the audibility of relevant speech cues is restored in HI listeners via amplification before measuring their suprathreshold hearing performance. The SNR loss is considered as a measure of the effect of suprathreshold hearing deficits on speech-in-noise performance (Plomp, 1978; Killion, 1997). Clinically, it is defined as the difference between speech-in-noise thresholds of NH and HI listeners, when the speech is presented at audible levels (Killion *et al.*, 2004). The SNR loss is ideally an audibility-independent phenomenon, but Phatak *et al.* (2009) cautioned that the audibility loss may be sometimes misinterpreted as the SNR loss, due to insufficient gain. Current study verified that HI listeners indeed do not exhibit SNR loss for vowel recognition in the SS masker and in low-rate (≤12 Hz) modulated maskers, when audibility is restored.

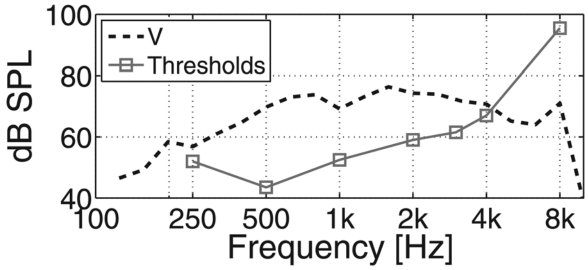

FIG. 9. Spectrum of amplified vowel stimuli for the HI listener with the lowest speech audibility. The audibility was calculated in terms of the SII. The average vowel spectrum was calculated over the vowel region that includes formants (*V*).

C. Suprathreshold deficits

A three-way comparison among NH, HI, and ENNH data (Fig. 7) showed that the reduced modulation benefit observed for HI listeners in 32 Hz masker is not entirely due to aging. The reduced benefit could result from peripheral or central auditory processing deficits that remain after correcting for the elevated audiometric thresholds. Some examples of peripheral processing deficits are reduced time and frequency resolutions (Humes and Christopherson, 1991; Gordon-Salant and Fitzgibbons, 1993; George *et al.*, 2006) and reduced ability to use temporal fine structure cues (Lorenzi *et al.*, 2006a; Lorenzi *et al.*, 2008). These processing deficits seem to be co-morbid with elevated thresholds, but persist after compensating for elevated thresholds. Dubno *et al.* (2003) found that speech recognition scores for isolated CV and VC syllables in a 25 Hz interrupted masker were correlated with forward masking thresholds at 2 and 4 kHz. Jin and Nelson (2006) also found a correlation between forward masking at 2 and 4 kHz and suprathreshold consonant MMR in aided HI listeners.

In addition to these peripheral deficits, age-related central and cognitive deficits could further reduce the benefit of masker fluctuations. Central processing deficits in auditory attention (Shinn-Cunningham and Best, 2008), working memory (Rudner *et al.*, 2003), and speed of processing (Jerger *et al.*, 1991) are known to affect speech-in-noise performance. It is possible that central deficits, which are often associated with aging (Pichora-Fuller, 2003), could also limit the benefit of short speech glimpses. Further investigation is warranted to identify the specific deficits, both age-related and age-independent, responsible for reducing modulation benefit. Nevertheless, experiment II demonstrated that neither hearing loss nor aging affects the benefit of slow masker fluctuations (≤12 Hz) for vowel recognition.

V. CONCLUSIONS

The main conclusions from this study can be summarized as follows.

(1) The MMR for vowel recognition is greater in magnitude and is more affected by the rate of masker modulations, than that for consonant recognition.
(2) Word MMRs from published literature are closer to vowel MMR than to consonant MMR observed in the current study.
(3) When speech is presented at suprathreshold levels, neither aging nor hearing loss affects the benefit of slow masker fluctuations (≤12 Hz) for vowel recognition.

ACKNOWLEDGMENTS

We thank Mary Cord and Jennie Canan for audiometric screenings. John Galvin from the House Ear Institute provided original recordings of vowel stimuli. We are grateful to Christian Lorenzi and Peggy Nelson for sharing their data. Sincere thanks to Emily Buss and two anonymous reviewers for their constructive comments. This research was supported by Cooperative Research and Development

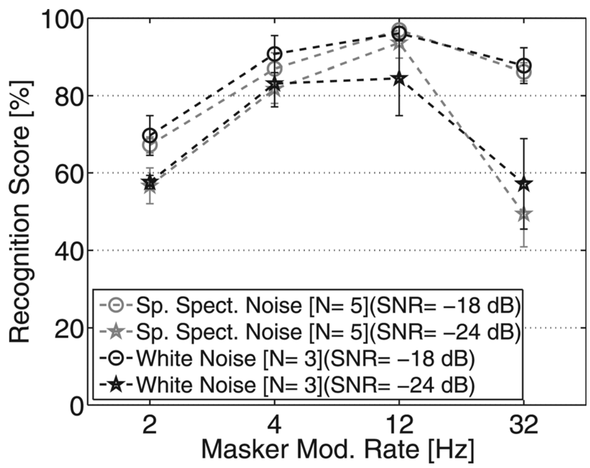

FIG. 10. A comparison of vowel recognition scores in speech-spectrum noise and white noise at −18 and −24 dB SNR.

Agreements between the Clinical Investigation Regulatory Office, U.S. Army Medical Department and School, and the Oticon Foundation, Copenhagen, Denmark. The views presented in this paper are those of the authors and do not necessarily reflect the official policy or position of the Army, Navy, or Air Force, the Department of Defense, or the U.S. Government.

APPENDIX

The differences in magnitudes and rate-dependence patterns between consonant and vowel MMRs could be partially due to different masker spectra in the two tasks. To test this, vowel recognition was measured with three NH listeners in white noise at −18 and −24 dB SNR. Figure 10 compares vowel recognition scores in speech-spectrum noise (from experiment I) and white noise. A repeated-measures ANOVA was conducted with SNR and modulation rate as within subject factors and masker spectrum as the between-subject factor. There was no significant main effect of the masker spectrum [$F(1,1) = 0.03$, $p = 0.869$], and the three-way interaction among rate, SNR and masker type was also not significant [$F(1,3) = 1.4$, $p = 0.272$]. *Post hoc* t-tests showed that differences between vowel recognition scores in the two types of maskers were not significant at any rate or SNR tested. Thus, differences between consonant and vowel scores were not due to different masker spectra.

[1]Due to their slightly elevated thresholds, ENNH listeners were provided gain in order to remove the potentially confounding factor of audibility loss. ENNH listeners were also tested unaided in quiet and in 32 Hz masker at −18 dB SNR. Amplification did not affect scores in these two conditions. Gains provided to ENNH listeners were much lower than those provided to HI listeners and did not cause any signal clipping even at the lowest SNR.

[2]The gain prescribed by the Cambridge formula at a frequency f is $0.48\text{HL}(f) + \text{INT}(f)$, where $\text{HL}(f)$ and $\text{INT}(f)$ are hearing threshold and intercept, respectively, at that frequency. The standard prescription for this formula provides intercepts only up to 5 kHz. In the current study, gain targets at audiometric frequencies of 4, 6, and 8 kHz were calculated using the intercept at 4 kHz. Thus, although the intercept was fixed at high frequencies, the gain varied with the degree of hearing loss.

ANSI (**1997**). S3.5, *American National Standard Methods for Calculation of the Speech Intelligibility Index* (Acoustical Society of America, 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.

Blumstein, S. E., and Stevens, K. N. (**1980**). "Perceptual invariance and onset spectra for stop consonants in different vowel environments," J. Acoust. Soc. Am. **67**(2), 648–662.

Buss, E., Whittle, L. N., Grose, J. H., and Hall, J. W., III. (**2009**). "Masking release for words in amplitude-modulated as a function of modulation rate and task," J. Acoust. Soc. Am. **126**(1), 269–280.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**(3), 1562–1573.

Diehl, R. L., Kluender, K. R., Foss, D. J., Parker, E. M., and Gernsbacher, M. A. (**1987**). "Vowels as islands of reliability," J. Mem. Lang. **26**(5), 564–573.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2003**). "Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with normal hearing," J. Acoust. Soc. Am. **113**(4), 2084–2094.

Festen, J. M., and Plomp, R. (**1990**). "Effect of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Fu, Q.-J., and Galvin, I., J. J. (**2006**). "Recognition of simulated telephonic speech by cochlear implant users," Am. J. Audiol. **15**(2), 127–132.

Füllgrabe, C., Berthommier, F., and Lorenzi, C. (**2006**). "Masking release for consonant features in temporally fluctuating background noise," Hear. Res. **211**, 74–84.

George, E. L. J., Festen, J. M., and Houtgast, T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**(4), 2295–2311.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1993**). "Temporal factors and speech recognition performance in young and elderly listeners," J. Speech Hear. Res. **36**, 1276–1285.

Grant, K. W., and Walden, B. E. (**1996**). "Evaluating the articulation index for auditory-visual consonant recognition," J. Acoust. Soc. Am. **100**(4), 2415–2424.

Grose, J. H., Mamo, S. K., and Hall, J. W., III (**2009**). "Age effects in temporal envelope processing: Speech unmasking and auditory steady state responses," Ear Hear. **30**(5), 568–575.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**(5), 3099–3111.

Howard-Jones, P. A., and Rosen, S. (**1993**). "The perception of speech in fluctuating noise," Acustica **78**, 258–272.

Humes, L. E., and Christopherson, L. (**1991**). "Speech identification difficulties of hearing-impaired elderly persons: The contributions of auditory processing deficits," J. Speech Hear. Res. **34**, 686–693.

Jerger, J., Jerger, S., and Pirozzolo, F. (**1991**). "Correlation analysis of speech audio-metric scores, hearing loss, age, and cognitive abilities in the elderly," Ear Hear. **12**(2), 103–109.

Jin, S., and Nelson, P. B. (**2006**). "Speech perception in gated noise: The effects of temporal resolution," J. Acoust. Soc. Am. **119**(5), 3097–3108.

Kewley-Port, D., Burkle, T. Z., and Lee, J. H. (**2007**). "Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners," J. Acoust. Soc. Am. **122**(4), 2365–2375.

Kidd, G. J., Mason, C. R., and Feth, L. L. (**1984**). "Temporal integration of forward masking in listeners having sensorineural hearing loss," J. Acoust. Soc. Am. **75**(3), 937–944.

Killion, M. C. (**1997**). "SNR loss: 'I can hear what people say, I can't understand them.'," Hear. Rev. **4**(12), 8–14.

Killion, M. C., Niquette, P., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (**2004**). "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **116**(4), 2395–2405.

Lorenzi, C., Debruille, L., Garnier, S., Fleuriot, P., and Moore, B. C. J. (**2008**). "Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal," J. Acoust. Soc. Am. **125**(1), 27–30.

Lorenzi, C., Gilbert, G., Carn, C., Garnier, S., and Moore, B. C. J. (**2006a**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.

Lorenzi, C., Husson, M., Ardoint, M., and Debruille, X. (**2006b**). "Speech masking release in listeners with flat hearing loss: Effects of masker fluctuation rate on identification scores and phonetic feature reception," Int. J. Audiol. **45**, 487–495.

Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**(2), 167–173.

Moore, B. C. J., and Glasberg, B. R. (**1998**). "Use of a loudness model for hearing aid fitting. I. Linear hearing aids," Brit. J. Audiol. **32**, 301–319.

Nelson, P. B., Jin, S., 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**(2), 961–968.

Nilsson, M. J., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of a hearing in noise test for the measurement of speech reception threshold," J. Acoust. Soc. Am. **95**, 1085–1099.

Peters, R. W., Moore, B. C. J., 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**(1), 577–587.

Phatak, S. A., and Allen, J. B. (**2007**). "Consonant and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am. **121**(4), 2312–2326.

Phatak, S. A., Yoon, Y., Gooler, D. M., and Allen, J. B. (**2009**). "Consonant recognition loss in hearing-impaired listeners," J. Acoust. Soc. Am. **126**(5), 2683–2694.

Pichora-Fuller, M. K. (**2003**). "Cognitive aging and auditory information processing," Int. J. Audiol. **42**, S26–S32.

Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**(2), 533–549.

Rudner, M., Rönnberg, J., and Lunner, T. (**2003**). "Working memory supports listening in noise for persons with hearing impairment," J. Am. Acad. Audiol. **22**(3), 156–167.

Sherbecoe, R. L., and Studebaker, G. A. (**2004**). "Supplementary formulas and tables for calculating and interconverting speech recognition scores in transformed arcsine units," Int. J. Audiol. **43**(8), 442–448.

Shinn-Cunningham, B. G., and Best, V. (**2008**). "Selective attention in normal and impaired hearing," Trends Amplif. **12**(4), 283–299.

Slaney, M. (**1998**). "Auditory Toolbox, Version 2," Technical Report No. 1998-010, Interval Research Corporation, Palo Alto, CA.

Wilson, R. H., and Carhart, R. (**1971**). "Forward and backward masking: Interactions and additivity," J. Acoust. Soc. Am. **49**(4), 1254–1263.