

Masking release for words in amplitude-modulated noise as a function of modulation rate and task

Emily Buss,^{a)} Lisa N. Whittle, John H. Grose, and Joseph W. Hall III

Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

(Received 8 May 2008; revised 14 April 2009; accepted 15 April 2009)

For normal-hearing listeners, masked speech recognition can improve with the introduction of masker amplitude modulation. The present experiments tested the hypothesis that this masking release is due in part to an interaction between the temporal distribution of cues necessary to perform the task and the probability of those cues temporally coinciding with masker modulation minima. Stimuli were monosyllabic words masked by speech-shaped noise, and masker modulation was introduced via multiplication with a raised sinusoid of 2.5–40 Hz. Tasks included detection, three-alternative forced-choice identification, and open-set identification. Overall, there was more masking release associated with the closed than the open-set tasks. The best rate of modulation also differed as a function of task; whereas low modulation rates were associated with best performance for the detection and three-alternative identification tasks, performance improved with modulation rate in the open-set task. This task-by-rate interaction was also observed when amplitude-modulated speech was presented in a steady masker, and for low- and high-pass filtered speech presented in modulated noise. These results were interpreted as showing that the optimal rate of amplitude modulation depends on the temporal distribution of speech cues and the information required to perform a particular task. © 2009 Acoustical Society of America. [DOI: 10.1121/1.3129506]

PACS number(s): 43.66.Dc, 43.71.Es [KWG]

Pages: 269–280

I. INTRODUCTION

In normal-hearing listeners, masking of a speech signal presented in broadband noise can be reduced by the introduction of masker amplitude modulation (AM). It is widely believed that this masking release can be accounted for in terms of the reduced masker levels associated with modulation minima, providing the listener with brief “glimpses” of the signal at an improved signal-to-noise ratio (SNR) (Miller and Licklider, 1950; Dirks and Bower, 1970; Howard-Jones and Rosen, 1993). This explanation is analogous to looking through a picket fence—an observer can see through the gaps between slats in the fence, and that is often sufficient to build up an accurate impression of the scene behind that fence (Miller and Licklider, 1950). Similarly, with an amplitude-modulated masker, the listener hears brief portions of the signal in the masker minima, and under some conditions that provides enough information to decipher the speech signal.

The rate of AM has an effect on the magnitude of masking release; typically the biggest effects have been reported for relatively slow rates, in the vicinity of 10 Hz or lower (Miller and Licklider, 1950; Howard-Jones and Rosen, 1993; Bacon *et al.*, 1998). There is also some evidence that the optimal rate of AM may differ across speech materials. For example, the optimal rate for spondee words was found to be 1 Hz, lower than the optimal rate for other two-syllable words and monosyllabic words (Dirks *et al.*, 1969; Dirks and Bower, 1971). This result was interpreted as reflecting the

increased redundancy of spondee words. A slow rate of masker AM provides temporally sparse glimpses of the target speech, but each of those glimpses is of relatively high quality. Because forward masking decays at an approximately constant rate as a function of time (Plomp, 1964), the longer modulation minima associated with lower rates of AM result in less forward masking overall. Because of this reduction in non-simultaneous masking, speech cues coincident with long-duration modulation minima are encoded with greater fidelity. For redundant speech materials, these sparsely distributed high-quality glimpses may be sufficient to identify the target word, whereas less redundant material might require more temporally dispersed glimpses to support identification. The results of Dirks and his colleagues (Dirks *et al.*, 1969; Dirks and Bower, 1971) are consistent with an interaction between modulation rate and cue redundancy. However, the target speech materials used in those studies differed across the redundancy conditions, leaving open the possibility that other factors such as word frequency or acoustic differences across stimulus sets played a role in the pattern of masking release.

One way to influence the redundancy of cues sufficient to perform a speech task is to manipulate the context in which the target material is presented. In an open-set sentence recognition task, the semantic context in which a word is presented can reduce the acoustic cues necessary to identify that word. The speech in noise test employs this approach, with the phrase preceding a target word either strictly limiting the semantically plausible set of final words (high predictability) or allowing for a wide range of final words (low predictability). One drawback to this approach is that it is difficult to precisely quantify the effect of context, with

^{a)}Author to whom correspondence should be addressed. Electronic mail: ebuss@med.unc.edu

both linguistic and subjective factors playing a role in predictability of the target word. It is also possible to manipulate redundancy in a closed-set task by changing the size of the response set, a procedure which lends itself more easily to parametric manipulations. Miller *et al.* (1951) showed that identification of a target word changes in a comparable way whether cue redundancy is manipulated through semantic context or response set size. The current study used a set-size manipulation of cue redundancy to assess the role of masker AM rate, an approach which has the advantage that a single set of speech stimuli can be used across redundancy conditions.

The current experiments examined the role of cue redundancy in the masking release associated with masker AM in a population of normal-hearing adults. The first experiment estimated masking release as a function of masker AM rate for three tasks: detection, three-alternative forced-choice identification (3AFC-ID), and open-set identification (open-ID). The detection task can be performed based on a very coarse cue, such as an increment in stimulus energy at any frequency associated with speech. The 3AFC-ID task requires more detailed information, such as a phoneme or a stimulus feature that distinguishes the alternatives; in this case the information required to arrive at the correct response could be quite limited, perhaps even based on a single glimpse in time. In contrast, the open-ID requires a relatively complex set of cues, including multiple phonemes distributed over time. Because the masking release associated with masker AM is thought to be limited by temporal resolution, longer duration glimpses associated with lower modulation rates should provide higher quality acoustic information and support accurate performance to the extent that sparsely distributed cues support good performance of the task. This reasoning leads to the hypothesis that the optimal rate of masker AM depends on the task, with best performance for detection being associated with a single glimpse of high quality (i.e., low rate AM) and best performance for the open set requiring a larger number of more widely dispersed glimpses (i.e., high rate AM). Such a result would lend further support to the conclusions of Dirks and Bower (1971), who showed analogous effects using stimuli with inherently high cue redundancy (spondees) or low cue redundancy (non-spondee, two-syllable words).

II. EXPERIMENT 1

A. Methods

1. Observers

Five observers participated in this study (two females) ranging in age from 19.7 to 32.8 years (mean 23.8 years). All observers had pure tone thresholds of 20 dB hearing level (HL) or better at octave frequencies from 250 to 8000 Hz in the test ear (ANSI, 1996), and none reported a history of ear disease. Non-native English speakers were excluded from participation, and all observers spoke with an American accent.

2. Stimuli

Target speech material was a set of 500 consonant-nucleus-consonant (CNC) words (Peterson and Lehiste, 1962), spoken by an adult male with an American accent. These recordings were 444–992 ms, with a mean duration of 744 ms. The sampling rate was 24.4 kHz, and all signals were passed through an 8-kHz second order Butterworth low-pass (LP) filter. Recordings were digitally scaled to equal-rms level across tokens.

The masker was a Gaussian noise that was spectrally shaped to the long-term spectrum of the speech stimuli, referred to as speech-shaped noise. In some conditions the maskers were amplitude modulated via multiplication with a raised sinusoid, with modulation rates of 2.5, 5, 10, 20, or 40 Hz; in these conditions the peak masker level was 75-dB sound pressure level (SPL), for an overall level of 70.8 dB SPL. There were two steady masker conditions. In the *equal-peak* steady masker condition, the masker was presented at 75-dB SPL, matching the peak level of the AM masker. In the *equal-rms* conditions the masker was presented at 70.8 dB SPL, matching the overall level of the AM masker.

3. Procedures

There were three tasks. In the *detection* task there were three listening intervals, visually indicated by lights on a hand-held response box. Listening intervals were 1 s in duration and separated by 500 ms. The observer was asked to select the interval in which a CNC word was presented; a randomly selected speech token was equally likely to be presented in each listening interval, and token onset coincided with the onset of the listening interval. In the 3AFC-ID task the observer was presented with a randomly selected word and then asked to identify the word from three alternatives presented visually after the listening interval, the foils being selected randomly from the remaining 499 tokens. In the open-ID condition the observer was presented with a randomly selected word and asked to repeat that word aloud; at that point the observer was visually presented with the correct response and prompted to score his response as correct or incorrect using buttons displayed on the computer screen.¹ An experimenter monitored each experimental session, including spot checks for correct self-scoring in the open-ID task; in no case did the experimenter have to re-instruct an observer in any of these procedures.

In all conditions the masker was presented continuously. The signal level was adjusted following a one-down, one-up tracking rule estimating 50% correct (Levitt, 1971). At the beginning of a track, signal level was adjusted in 4-dB steps; that stepsize was reduced to 2 dB after the second reversal. Each track continued until a total of 12 reversals had been obtained, and the threshold estimate associated with a track was the average signal level at the last 10 track reversals.

During the first testing session data were collected in one of the two speech identification tasks, selected at random (either 3AFC-ID or open-ID). The second session consisted of data collection for the alternate speech identification task. The detection conditions were completed in the third and final session for all observers. In each of the three sessions,

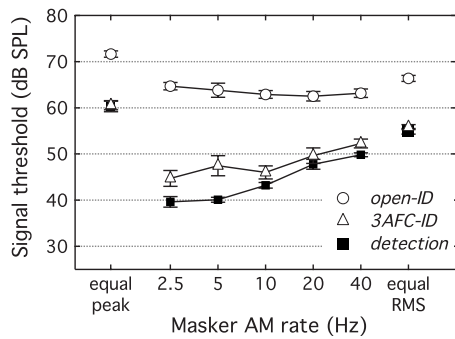


FIG. 1. The mean threshold across observers is plotted as a function of condition, with error bars indicating ± 1 standard error of the mean. Symbols indicate the task: detection (■), 3AFC-ID (△), or open-ID (○). Thresholds for the equal-peak steady masker condition are plotted at the far left and those for the equal-rms condition at the far right of the panel.

observers completed one threshold estimate in each of the seven masker conditions in random order, including five modulation rates and two levels of steady masker. A second estimate was then obtained in all seven conditions in a new random order. As time allowed, a third estimate was obtained in conditions for which the previous two estimates varied widely, at the discretion of the experimenter on a case-by-case basis. All estimates obtained in a condition were averaged to produce a final threshold estimate for each observer. Each listening session lasted for 1 h.

B. Results

Figure 1 shows mean thresholds, averaged across the five observers, plotted as a function of masker condition. Symbols reflect the task, as indicated in the figure legend. These thresholds indicate the signal level required to achieve 50% correct in a fixed-level masker, so low thresholds represent good performance. Results in the baseline, steady masker conditions will be considered first. For the equal-peak masker conditions, thresholds are very similar for the detection and the 3AFC-ID tasks, with means of 60.3 and 60.6 dB SPL, respectively. Threshold for the open-ID task is higher, with a mean of 71.6 dB SPL. Thresholds in the equal-rms condition are on average 5.1 dB lower than those in the associated equal-peak conditions. This difference is consistent with the 4.2-dB reduction in level. A set of three t-tests was performed, comparing the difference in steady masker thresholds within observer for each task; in no case was the mean difference significantly different from 4.2 dB ($p > 0.36$, two-tailed).

Masker AM tended to reduce thresholds relative to both the equal-peak and equal-rms conditions, but this effect was dependent on both AM rate and task. The effect of rate was broadly similar for the detection and 3AFC-ID tasks, with best thresholds for lower rates of masker AM and elevation with increasing modulation rate. There is some indication that the trend for poorer performance with increasing AM rates begins at a lower rate for the detection than the 3AFC-ID task, with knee points at 5 and 10 Hz, respectively. The effect of masker AM rate on these two conditions was large, with thresholds for the 3AFC-ID task spanning a range of 7.5 dB, and those for the detection task varying by 10.2

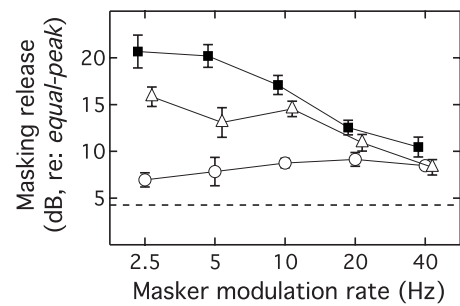


FIG. 2. Masking release is plotted as a function of masker AM rate relative to the threshold obtained in the equal-peak steady masker condition associated with each task. Symbols indicate the three response conditions as in Fig. 1: detection (■), 3AFC-ID (△), or open-ID (○). Error bars indicate ± 1 standard error around the mean, and the horizontal dashed line shows the 4.2 dB improvement expected due to the reduction in masker level associated with AM. Data points are slightly offset on the abscissa to aid visual inspection.

dB. In contrast, best performance for the open-ID task was obtained at the highest AM rates of 10–40 Hz, with poorer thresholds for lower AM rates. The effect of AM rate on open-ID thresholds was more modest than in the other conditions, with mean thresholds spanning a range of just 2.2 dB.

To assess this pattern of results thresholds were submitted to a repeated-measures analysis of variance (ANOVA), with five levels of AM RATE (2.5, 5, 10, 20, and 40 Hz) and three levels of TASK (detection, 3AFC-ID, open-ID). There were significant main effects of RATE ($F_{4,16}=14.0$, $p < 0.0001$) and TASK ($F_{2,8}=218.5$, $p < 0.0001$), and there was a significant interaction ($F_{8,32}=13.5$, $p < 0.0001$). A pre-planned linear contrast indicated a significant interaction with AM rate for the open-ID and 3AFC-ID tasks ($F_{1,4}=69.3$, $p < 0.001$), consistent with the task-by-rate interaction described above. The interaction with AM rate for the 3AFC-ID and detection tasks just failed to reach significance ($F_{1,6}=7.1$, $p=0.056$).

Because the change in open-ID thresholds with AM rate was modest, it was of interest to determine whether the significant interaction between 3AFC-ID and open-ID tasks was due solely to variability in the degree to which increasing rates of AM increased 3AFC-ID thresholds, or whether the open-ID thresholds actually improved as a function of AM rate. The significance of the decrease in open-ID thresholds as a function of AM rate was assessed in two stages. First, thresholds were fitted with a linear regression, with four dummy variables coding for observer. This analysis resulted in a significant effect of observer ($F_{4,20}=3.82$, $p < 0.05$). A correlation between the residuals from this analysis and the logarithm of modulation rate was computed. This second analysis resulted in a significant linear association between masker AM rate and open-ID thresholds ($r=-0.36$, $p < 0.05$, one-tailed). While this fit accounted for only 13% of the variance in the data, it is consistent with an improvement in thresholds with increasing AM rate for this task.

The masking release associated with masker AM is shown in Fig. 2. The reduction in threshold relative to the equal-peak masker is plotted as a function of rate, with symbols following the same convention as in Fig. 1. The dashed

horizontal line indicates the reduction in masker level associated with AM (4.2 dB), and hence the improvement in threshold that would be expected if reductions in overall masker level were responsible for masking release. Values above this line reflect masking release associated with the transient improvements in SNR as a result of masker AM. Consistent with observations of absolute thresholds, above, these values of masking release show an interaction between masker AM rate and task; there is more masking release for low AM rates under conditions of high cue redundancy (detection and 3AFC-ID) and better performance at high rates under conditions of reduced redundancy (open-ID). This depiction of the data also highlights the fact that whereas there was a masking release in the mean data in all conditions, there was a nearly 15-dB difference in the masking release obtained across tasks with 2.5 Hz AM, and a relatively consistent masking release at 40-Hz. This result is consistent with the interpretation that performances in all three conditions benefit approximately equally for temporally dispersed brief modulation minima, but that the temporally sparse high-quality cues associated with low rates of AM are much more beneficial in high-redundancy (detection and 3AFC-ID) as compared to low-redundancy (open-ID) tasks.

C. Discussion

The results of experiment 1 support the hypothesis that the peak masking release associated with masker AM occurs at different rates of AM in tasks requiring differing levels of signal detail. Thresholds in tasks that can be performed based on sparse or coarse cues, in this case word detection or 3AFC-ID, are lowest for relatively low rates of AM, whereas thresholds in the open-set task, which requires more detailed information, are lowest at higher rates of AM.

1. Effects of task

Three speech tasks (detection, 3AFC-ID, and open-ID) were employed in order to manipulate the speech cues necessary to perform the task; it was argued that detection requires the sparsest cues, closed-set identification requires relatively minimal encoding of the speech signal, and open-ID requires relatively detailed encoding of multiple phonemes. Performance in these three tasks provides support for this ordering of task difficulty. Averaging across all masker conditions, performance is rank ordered following this presumed hierarchy of difficulty, with mean thresholds of 65.0 dB (open-ID), 50.9 dB (3AFC-ID), and 48.0 dB (detection). One caveat to this ranking of task difficulty is that the results of the detection and 3AFC-ID conditions were more similar to each other than to the open-ID condition, as reflected in the mean across all masker conditions and in the nearly identical steady masker thresholds in the detection and 3AFC-ID conditions.

Performance in the open-ID task in steady noise is consistent with published results. For example, [Studebaker et al. \(1994\)](#) measured percent correct for CID W-22 words as a function of SNR and found that a SNR of -5.5 dB was associated with 50% correct in speech-shaped noise, comparable to the average threshold of -3.9 dB SNR found for the

open-ID steady masker conditions of the present experiment. It is commonly observed that speech recognition requires a higher SNR than detection of speech. In one demonstration of this effect [Hawkins and Stevens \(1950\)](#) measured thresholds for running discourse and found a 10-dB difference in detection as compared to recognition. This effect size is comparable to the approximately 10-dB threshold difference observed in the present open-ID and detection tasks. While these results are consistent, interpretation of this parallel is complicated by ambiguity in quantifying the relative context effects in the two paradigms, and, in particular, whether the open-ID of the present experiment is comparable the running discourse used in the [Hawkins and Stevens \(1950\)](#) study.

Differences between the open-ID and 3AFC-ID identification conditions are consistent with previous observations that words are easier to understand in the context of a sentence and that high-predictability sentences are recognized more accurately than low-predictability sentences ([Miller et al., 1951](#); [Kalikow et al., 1977](#)). For example, [Miller et al. \(1951\)](#) reported percent correct for monosyllabic word recognition at a range of SNRs for open set and closed set, with a wide range of response options; these results indicate an ~ 18 dB difference in the 50% correct point between open-set and four-alternative forced-choice conditions (with SNRs of +4 and -14 dB, respectively), very close to the equal-peak masker results in the present data set for the open-ID and 3AFC-ID tasks (with SNRs of -3.4 and -14.4 dB, respectively).

The effect of context on recognition is sometimes described in terms of the “linguistic entropy” of the speech sample ([van Rooij and Plomp, 1991](#); [Bronkhorst et al., 1993](#)). When entropy is high, the listener has very little information with which to narrow the range of possible targets, but when entropy is low the listener can use the information present in the surrounding segment of speech to help interpret the sensory signal; entropy is the inverse of redundancy. The effects of linguistic entropy have been modeled in terms of variable levels of cognitive noise ([Müsch and Buus, 2001](#)). One way to think about the performance advantage associated with low linguistic entropy is in terms of a template-matching algorithm. When the pool of templates is small the odds of identifying the correct template are relatively good, whereas a very large pool of templates introduces more opportunities for error.

2. Effects related to modulation rate

The most striking finding related to masker modulation rate was that whereas there was a trend for better performance at lower modulation rates in the 3AFC-ID and detection tasks, the trend in the open-ID task was for better performance at the higher rates of modulation. This is also broadly consistent with the findings of [Dirks and his colleagues. Dirks et al. \(1969\)](#) reported that masking release for spondee words and sentences was greater for a masker AM rate of 1 Hz than 10 Hz. The opposite was observed for monosyllabic words, where a masker AM rate of 10 Hz was associated with greater masking release than a rate of 1 Hz. This rate effect was interpreted in terms of the minimal cues necessary to correctly identify a speech token given semantic

constraints inherent to the speech materials. In the present paradigm, the level of sensory detail necessary to correctly identify the speech token was manipulated by task, with 3AFC-ID requiring minimal cues (similar to high-redundancy/restrictive context materials) and open-ID requiring more complex cues (similar to low-redundancy/minimal context materials). In the present data set, the pattern of masking release as a function of AM rate is very similar for the 3AFC-ID and detection tasks. There is a non-significant trend for better performance at lower rates for the detection task, an effect that would be consistent with the interpretation of the effect of AM rate across the two identification tasks.

3. Effects of forward masking

Whereas the results are consistent with an interpretation of the AM rate effect in terms of the temporal distribution of glimpses of the speech, this account may be complicated by the fact that the SNR at threshold differs substantially across tasks. For example, at a 20-Hz rate of masker AM, threshold in the open-ID condition is 62.5 dB, whereas those in the 3AFC-ID and detection conditions are 49.6 and 47.8 dB, respectively. In poor SNR conditions, the speech signal is likely to be audible only during the temporal center of a masker modulation minimum, whereas in the higher SNR conditions, the speech signal may be audible for a larger proportion of the modulation period. As a consequence, the glimpses of speech associated with high rate AM could be effectively briefer at low as compared to high SNRs. If that is the case, then forward masking could play a larger role in performance of detection and 3AFC-ID tasks as compared to open-ID at high masker AM rates.

Supplemental data were collected in order to test the possible role of forward masking in the effect of AM rate. Stimuli were identical to those in the main experiment except that AM was applied to the speech signal instead of to the masker. In these conditions the masker was a 75-dB SPL speech-shaped noise, and the signal was a CNC word that was sinusoidally amplitude modulated at either 5 or 20 Hz. Signal level was adjusted adaptively to estimate threshold for 50% correct response, as described above, in both the 3AFC-ID or open-ID task. Thresholds were collected in random order, with a total of three to four estimates in each of four conditions (2 rates \times 2 tasks). Seven normal-hearing observers were recruited to complete the supplemental conditions, all meeting the inclusion criteria noted above for the primary study. Observers ranged in age from 21.0 to 40.5 years (mean 29.5 years), and all had previously participated in a speech study using CNC words, including experiments 1 and 2 described in the present report, as well as very similar pilot experiments.

Results of these supplemental conditions are shown in Fig. 3, with mean threshold plotted as a function of modulation rate and error bars indicating one standard deviation. Symbol shape reflects task, consistent with Fig. 1. As in the AM-masker conditions, thresholds are higher in the open-ID than in the 3AFC-ID conditions, with a mean difference of 15.9 dB. Thresholds also differ for the two signal modulation rates, with better performance for 5-Hz AM in the 3AFC-ID

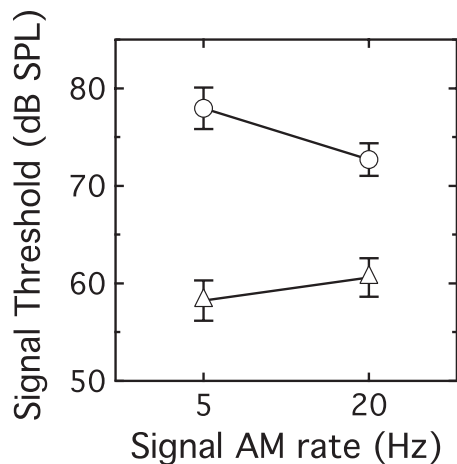


FIG. 3. Word identification thresholds are plotted as a function of signal AM rate. Symbols reflect the response condition, following the conventions of Fig. 1: 3AFC-ID (Δ) or open-ID (\circ). Error bars indicate ± 1 standard deviation.

condition and 20-Hz AM in the open-ID condition. A repeated-measures ANOVA was performed to assess the significance of this interaction. There were two levels of TASK (3AFC-ID and open-ID) and two levels of RATE (5 and 20 Hz). There was a main effect of TASK ($F_{1,6}=292.49$, $p < 0.0001$), a main effect of RATE ($F_{1,6}=8.19$, $p < 0.05$), and a significant interaction ($F_{1,6}=40.07$, $p < 0.001$). Paired t-tests were performed to assess whether this interaction is due to rate effects on one or both of the tasks. These analyses confirmed that thresholds at the two signal AM rates were significantly different (one-tailed) for both the open-ID ($t_6=8.00$, $p < 0.001$) and 3AFC-ID ($t_6=-2.68$, $p < 0.05$) tasks.

Because the speech rather than the masker was modulated in the supplemental conditions, forward masking would not be expected to play a large role in the results. The finding of an interaction between modulation rate and task lends support to the idea that the distribution of information required to perform each of the speech tasks is responsible for the AM-masker rate effects obtained in the main experiment. Previous work on amplitude-modulated speech, sometimes described as interrupted speech, has uncovered a non-monotonic relationship between modulation rate and performance that is dependent on both modulation duty cycle and speaking rate (Huggins, 1964; Powers and Wilcox, 1977). These results have largely been explained in terms of the temporal distribution of cues required to support recognition. For example, if the “off” portion of the modulation period is long relative to the duration of a word, then some words will be missed in an open-set task. A similar explanation appears to be valid for the present data. A slow modulation period is associated with infrequent but relatively long-duration glimpses, sufficient to perform the 3AFC-ID task. Those cues are too temporally sparse to perform the open-ID task, however, where better performance is obtained when glimpses are more widely distributed over time.

The finding of better performance in the 3AFC-ID task at 5 than 20 Hz implies that the short glimpses associated with 20-Hz AM are less effective than those at 5 Hz even in the absence of forward masking. One factor that could un-

derlie better performance at the lower AM rates in the 3AFC-ID condition is the detrimental effect of sidebands associated with signal modulation. This suggestion is analogous to the switching artifact that has been proposed to limit performance with interrupted speech at high rates of interruption (Huggins, 1964).

4. The role of frequency region

It has been argued that whereas vowels are more perceptually salient, with higher mean energy content, consonants are more important in word identification (Bonatti *et al.*, 2005; Toro *et al.*, 2008). If that is the case, then it is also possible that detection and 3AFC-ID could be performed based on acoustically salient vowel information, whereas open-ID relies more on consonants. This possibility gets some support from the finding that across tokens, identification of vowels in speech-shaped noise is better than that for consonants, though there is wide variability across the different types of consonants (Phatak and Allen, 2007). There is also evidence that low- and high-frequency regions contribute differently to the perception of vowels and consonants. For example, recognition of vowels in speech-shaped noise is dominated by low-frequency cues related to F1 and to a lesser extent F2 (Parikh and Loizou, 2005), whereas the spectral cues, which differentiate consonants, are more robust at high than low frequencies (Phatak and Allen, 2007). These observations support the possibility that the available frequency region of speech could play an important role in the pattern of masker AM effects observed in the present experiment.

In normal-hearing listeners it has been suggested that temporal resolution may limit access to low-frequency cues more than those at high frequencies, a hypothesis motivated by several considerations: the possibility of inherently better temporal resolution in high- than low-frequency channels (Festen, 1987; Stuart and Phillips, 1996), better representation of the masker envelope at the output of wider high-frequency channels (Haggard *et al.*, 1990; Bacon *et al.*, 1997), or increased suppression at high frequencies in normal-hearing listeners (Bacon *et al.*, 1997; Lee and Bacon, 1998). In the context of the present paradigm, this line of reasoning would suggest that relatively high-frequency speech information could play a dominant role in the masking release observed with a full-spectrum speech target.

It is well known that masker AM provides greatly reduced benefits for listeners with moderate sensorineural hearing impairment as compared to normal-hearing listeners (e.g., Festen and Plomp, 1990; Gustafsson and Arlinger, 1994). Whereas some previous studies have implicated temporal resolution as a limitation in AM-related masking release, other work suggests that this factor is insufficient to explain the performance of hearing impaired listeners (Jin and Nelson, 2006), implicating reduced spectral resolution or the interaction between reduced spectral resolution and reduced speech redundancy as the dominant factor (Baer and Moore, 1994; Hall *et al.*, 2008). Several studies have shown a diminished benefit of masker AM minima for speech that has been degraded using a vocoder simulation (Kwon and Turner, 2001; Nelson *et al.*, 2003; Qin and Oxenham, 2003;

Nelson and Jin, 2004). In some conditions performance worsens with the introduction of masker AM (Kwon and Turner, 2001), an effect attributed to masker modulation interfering with the processing of speech envelope fluctuations. One explanation for the poor intelligibility of vocoded speech in AM noise is that the spectral coarseness of the target fails to provide sufficient cues to segregate it from the masker. Qin and Oxenham (2005) suggested that cues related to voice fundamental frequency are very important to this segregation process, and that poor performance in these conditions is due to the fact that vocoded speech discards fine-structure cues to F0; this hypothesis is bolstered by the finding that restoring those cues significantly improves performance (Chang *et al.*, 2006; Qin and Oxenham, 2006). Other data suggest that similar effects may limit masking release obtained with introduction of masker AM for unprocessed speech (Lorenzi *et al.*, 2006). These results are consistent with the hypothesis that low-frequency fine-structure cues are critical for stream segregation, without which envelope patterns associated with masker AM could exert modulation masking that interferes with the use of envelope-based speech cues. By this reasoning, it is possible that the masking release trends observed in experiment 1 might have arisen largely from processes related to relatively low-frequency speech information.

A third and final possibility regarding the importance of different frequency regions of speech to masking release is that the release observed in experiment 1 is inextricably related to relatively *wideband* processes, and therefore depends critically on the synthesis of speech information across low and high spectral regions. This might be the case if spectral redundancy of speech cues were a precondition for AM-related masking release. This possibility is in accord with previous reports of greater masking release under conditions of high speech cue redundancy (Dirks *et al.*, 1969; Kwon and Turner, 2001), as well as the finding that masking release is reduced to a comparable degree for LP and high-pass (HP) speech stimuli when baseline performance in steady noise is comparable across filter conditions (Oxenham and Simonson, 2008). Fullgrabe *et al.* (2006) recently showed that different speech cues are perceived best at different rates of AM. The availability of a wide range of cues could increase the chances of correct identification in a masker characterized by a single rate of AM. Conversely, a spectrally impoverished cue set could severely restrict the speech information available at any single masker AM rate.

In summary, the discussion above highlights the importance of considering how the masking release effects such as those found in the first experiment may be related to absolute frequency region. The relative contribution of low- and high-frequency speech information to masking release was explored further in the second experiment.

III. EXPERIMENT 2

In experiment 1 it was shown that the masking release associated with masker AM varied as a function of AM rate, with different patterns of masking release for different speech tasks. These results were discussed in terms of audi-

bility, by virtue of unmasking of the cues required to perform each task as a function of time via masker AM. One potentially important factor in that paradigm was the relative contribution of low- versus high-frequency cues. Experiment 2 further assessed masking release for a LP and a HP filtered stimulus. If the effects observed in experiment 1 were driven solely by effects related to instantaneous signal-to-masker ratio, then the masking release due to masker AM should be very similar for LP and HP filtered speech in a speech-shaped noise masker under matched conditions, to the extent that the temporal distribution of cues is similar in these two frequency regions. If, on the other hand, the encoding of low- and high-frequency speech information is qualitatively different, then masking release could differ substantially across spectral regions.

One caveat that should be considered with respect to the present experiment is that the interaction of task and masker AM rate may rely on spectral as well as temporal redundancies. Up to this point the good performance at low rates of masker AM in the 3AFC-ID task has been discussed in terms of the temporal redundancy of cues, such that the cues available during a single, relatively long-duration glimpse could support correct identification, regardless of when that glimpse occurs in the word. Such good performance could also rely on the spectral redundancy of speech, however, as suggested by Oxenham and Simonson (2008). Reducing spectral redundancy by LP or HP filtering the stimulus, as in the present paradigm, could reduce the quality of each glimpse, such that low rates of masker AM no longer support good performance. Based on this reasoning, changes in the pattern of task-by-AM rate effects might be predicted for the present experiment.

A. Methods

1. Observers

A total of 17 observers participated in this experiment (15 females) ranging in age from 18.5 to 47.8 years (mean 27.1 years). All observers had pure tone thresholds of 20 dB HL or better at octave frequencies from 250 to 8000 Hz in the test ear (ANSI, 1996), and none reported a history of ear disease. Non-native English speakers were excluded from participation, and all listeners spoke with an American accent. Some observers had previously participated in psychoacoustic studies, but none using CNC speech materials.

2. Stimuli

As in experiment 1, testing involved either detection, 3AFC-ID, or open-ID for CNC words. All testing was performed in the presence of a continuous speech-shaped noise masker. Masker AM, when present, was achieved via multiplication with a raised sinusoid at a modulation rate of 2.5, 5, 10, 20, or 40 Hz. Steady masker conditions included equal-peak and equal-rms comparison conditions, associated with the same peak or the same rms level as the comparable AM-masker conditions. In contrast to experiment 1, both the signal and masker were passed through a LP or a HP filter. Filtering was achieved by passing the stimuli through a fourth order Butterworth filter twice, once forward and once

backward, with a 1700-Hz cut-off frequency. This cut-off frequency was selected based on pilot listening in which this cutoff resulted in comparable open-set thresholds for both LP and HP filtered speech; a steady masker was used for these pilot conditions. This value is also consistent with the observation that the centroid of speech information is typically cited as falling between 1000 and 2000 Hz (Studebaker *et al.*, 1987; DePaolis *et al.*, 1996; Henry *et al.*, 1998). In the LP and HP conditions the masker was 75 or 70.8-dB SPL prior to filtering; HP filtering in the HP condition reduced the overall masker level by 16 dB. In order to assess the significance of this level reduction in the pattern of results, an additional HP filter condition was included, wherein the masker level was increased by 16 dB, for a level of 75 or 70.8-dB SPL after filtering. This condition is referred to here as HP+16.

3. Procedures

Observers were randomly assigned to the LP, HP, or HP+16 conditions. Following the procedures of experiment 1, each observer began with one of the two identification conditions (either open-ID or 3AFC-ID), completing thresholds in random order. The second session was spent on the alternate identification condition, and all observers completed detection conditions in the third and final listening session. In all cases signal level associated with 50% correct was estimated using a one-down, one-up track. Initial level adjustments were made in steps of 4 dB; this stepsize was reduced to 2 dB after the second track reversal. Tracks continued for 12 reversals, and the average signal level at the last 10 reversals was used as an estimate of threshold. Between two and three estimates were obtained in each condition, and the means of all thresholds obtained are reported below.

B. Results

The mean SNRs at threshold for the equal-peak steady masker conditions are reported in Table I; the standard error of the mean appears to the right of each value, and comparable results from the full-spectrum conditions of experiment 1 appear in the top row of the table for comparison. Thresholds were relatively consistent across the three filter conditions (LP, HP, and HP+16). As in experiment 1, thresholds tended to fall in rank order, from open-ID, to 3AFC-ID, to detection, and thresholds in the final two tasks were relatively similar. The SNR at threshold for the open-ID task differed by about 5 dB between the results of experiments 1 and 2, indicating a detrimental effect of LP and HP filtering on masked open-set recognition. The significance of this difference in the open-ID performance between full-spectrum and filter conditions was evaluated using a set of three t-tests, one for each of the filter conditions; all three were significant ($\alpha=0.05$, two-tailed, and Bonferroni correction). In contrast, filtering had little or no effect on performance in the 3AFC-ID and detection conditions, for which SNR at threshold was relatively consistent across the two experiments and across filter conditions within experiment 2. Thresholds in the equal-rms condition (not shown) were on average 3.8 dB

TABLE I. Mean thresholds (SNR, in dB) for the equal-peak steady masker condition as a function of target filter conditions, masker level, and task. The standard error of the mean across thresholds ($n=5$ or $n=6$) appears in parentheses to the right of each estimate.

Filter condition	Task		
	Open-ID	3AFC-ID	Detection
Experiment 1, full-spectrum	-3.37 (0.70)	-14.44 (0.98)	-14.69 (1.13)
Experiment 2, low-pass (LP)	2.13 (1.03)	-14.56 (0.80)	-14.80 (0.57)
Experiment 2, high-pass (HP)	2.95 (1.07)	-13.73 (0.80)	-15.63 (0.93)
Experiment 2, high-pass (HP+16)	3.44 (1.01)	-14.54 (0.98)	-16.83 (0.63)

lower than those in the associated equal-peak conditions, with a standard error of 0.43 dB. This is consistent with the 4.2-dB reduced masker level.

The consistency of performance in the LP and HP filter conditions for the steady masker of the present experiment was assessed using a mixed model ANOVA. The mean threshold in the two steady masker conditions (equal-peak and equal-rms) was used in this analysis since the difference in masker level across these steady masker conditions did not significantly affect the SNR at threshold. There was one across-subject factor of COND (LP, HP, and HP+16) and one within-subject factor of TASK (open-ID, 3AFC-ID, and detection). The results of this analysis indicated a significant main effect of TASK ($F_{2,18}=518.88$, $p<0.0001$), but no main effect of COND ($F_{1,9}=1.14$, $p=0.31$) and no interaction ($F_{1,18}=1.28$, $p=0.30$). These results confirm that baseline performance is comparable across the three filter conditions, with comparable performance above and below the 1700-Hz filter cutoff.

Masking release was computed as the difference between thresholds measured with an AM masker and those from the associated equal-peak steady masker conditions. Results appear in Fig. 4, with masking release plotted as a function of modulation rate and error bars indicating one standard error of the mean. Masking release in all three filter conditions exhibited some common features: masking release associated with open-ID conditions was modest and that for detection was largest, with 3AFC-ID usually falling intermediate between these two. In all cases masking release varied across task more for low rates than for high rates of masker AM, where values tend to converge. These shared features were also noted in the results obtained with full-spectrum speech in experiment 1 (Fig. 2).

Despite these general commonalities, some aspects of the LP and HP filter data differed from the full-spectrum data of experiment 1, most notably a trend for smaller magnitudes of masking release in some conditions in the LP and HP conditions: averaging across all conditions, the mean masking release in experiment 1 was 12.3 dB, a value that can be compared to 11.2 dB in the LP and 10.6 dB in the HP conditions. In the LP condition the reduction in masking release is subtle and evident primarily in the open-ID conditions: whereas masking release in the open-ID condition of experiment 1 ranged from 6.9 to 9.1 dB, masking release in the LP conditions spanned from 4.8 to 7.9 dB.

The significance of the LP and HP filter on the pattern of masking release was assessed with a pair of mixed model ANOVAs. In each case the masking release obtained with full-spectrum stimuli from experiment 1 was compared with that from either the LP or the HP filter condition of the present experiment. These analyses included an across-subjects factor of COND (full-spectrum and filtered), a within-subjects factor of RATE (2.5, 5, 10, 20, and 40 Hz), and a within-subjects factor of TASK (open-ID, 3AFC-ID, and detection). As in the previous analyses, both of these analyses resulted in a highly significant main effect of RATE and TASK, as well as an interaction ($p<0.0001$). The result of interest here was in terms of the effect of COND. For the LP analysis, there was no main effect of COND ($F_{1,8}=0.58$, $p=0.47$), and none of the interactions with COND approached significance ($p>=0.36$). For the HP analysis, there was a significant main effect of COND ($F_{1,9}=18.38$, $p<0.005$) and a significant interaction between RATE and COND ($F_{4,36}=4.74$, $p<0.005$). These results indicate that the reduction in mean masking release with stimulus filtering

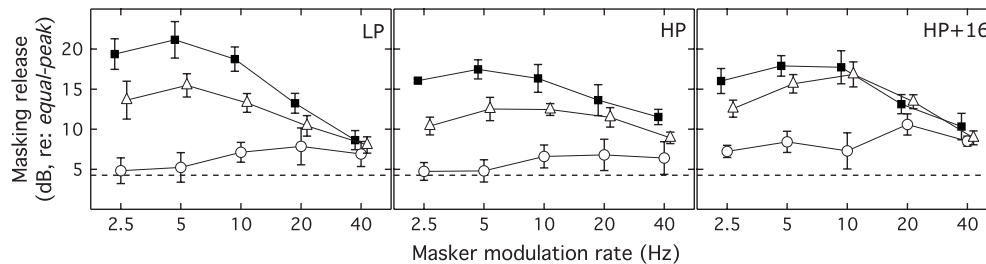


FIG. 4. Masking release is plotted as a function of masker AM rate relative to the threshold obtained in a steady masker with equal-peak level. Each panel shows results of a different filter condition, as indicated in the upper right corner. Symbols indicate the task: detection (■), 3AFC-ID (△), or open-ID (○). Error bars indicate ± 1 standard error around the mean, and the horizontal dashed line shows the 4.2 dB improvement expected due to the reduction in masker level associated with AM. Data points are slightly offset on the abscissa to aid visual inspection.

was significant for the HP but not the LP conditions. The significant interaction between RATE and COND for the HP analysis reflects the fact that, compared to masking release in the full-spectrum conditions, masking release for the HP conditions did not vary as widely across tasks for the low rates of AM, and these functions did not converge as closely for the highest rate of masking release (compare Fig. 2 with middle panel of Fig. 4).

It is possible that the reduced masking release and reduced effect of masker AM rate on the HP condition are related to the 16-dB reduction in stimulus level associated with the 1700-Hz HP filter. A third ANOVA was therefore performed to compare the HP+16 and full-spectrum data. Again, there were highly significant main effects of RATE and TASK, as well as a significant interaction ($p < 0.0001$). Of particular interest here, there was no main effect of COND ($F_{1,8} = 0.05$, $p = 0.82$), and none of the interactions with COND reached significance ($p > 0.12$). This result is consistent with the conclusion that the effect of the stimulus filter on the HP condition was mediated to some extent by a level effect rather than the elimination of speech cues below 1700 Hz.

The effect of masker AM rate on masking release in the open-ID condition was modest in the results of experiment 1, with AM rate accounting for only 13% of the variance in those data. Because of the small size of that result and its importance for interpretation of the task-by-AM rate interaction, an analysis was undertaken to assess the effect of masker AM rate on the open-ID task across all conditions of experiments 1 and 2. For this analysis, the differences between equal-peak and AM-masker conditions were computed for individual listeners. The resulting estimates of masking release were submitted to a mixed model ANOVA, with one across-subjects factor of COND (experiment 2: LP, HP, HP+16 and experiment 1) and one within-subjects factor of RATE (2.5, 5, 10, 20, and 40 Hz). This analysis resulted in a significant effect of RATE ($F_{4,68} = 2.75$, $p < 0.05$), but no significant effect of COND ($F_{3,17} = 1.45$, $p = 0.26$) and no interaction ($F_{12,68} = 0.21$, $p = 1.00$). A linear contrast performed on the effect of RATE was significant ($F_{1,17} = 12.77$, $p < 0.01$), consistent with the visual impression that masking release rose approximately linearly as a function of masker AM rate in the open-ID condition, with no reliable difference across filter conditions.

C. Discussion

Thresholds in the steady masker for both LP and HP filter conditions are statistically indistinguishable, consistent with the idea that the 1700-Hz cutoff corresponds to the centroid of speech cues for these stimuli. The SNR at threshold for the steady masker is nearly constant across filter conditions for a given task. Averaging across filter conditions, these thresholds are -15.3 dB for detection, -14.0 dB for 3AFC-ID, and 2.7 dB for the open-ID task. This pattern of results closely resembles that obtained for the full-spectrum conditions of experiment 1, with the exception of the open-ID task, for which thresholds were reliably about 5 dB higher in the filtered than full-spectrum conditions.

1. The role of frequency region in masking release

Masking release in the HP conditions of experiment 2 tended to be smaller than full-spectrum results from experiment 1, whereas those in the LP conditions were not reduced. In principle, this result could be due to the disproportionate contribution of low-frequency cues to masking release, such as cues based on the use of temporal fine structure. If this were the case, then increasing stimulus level should have little if any effect on this pattern of results. Contrary to that interpretation, masking release in the HP+16 condition was comparable to that obtained in the full-spectrum (experiment 1) and LP (experiment 2) conditions. This result supports the conclusion that masking release with the HP filtered target is smaller than that in comparable LP conditions not because of a failure of high-frequency information to convey the necessary speech cues, but rather due to the lower overall level of the stimulus in that frequency region. This result is consistent with published data indicating significant level effect on the intelligibility speech in AM noise (e.g., Dirks *et al.*, 1969; de Laat and Plomp, 1983; Festen, 1993; Stuart and Phillips, 1997).

The present paradigm bears some resemblance to the paradigm of Scott *et al.* (2001) and Elangovan and Stuart (2005). In those studies, masking release associated with an aperiodic masker was measured for NU6 words. LP filtering was shown to reduce the benefit of masker AM (Scott *et al.*, 2001), whereas HP filtering had a more modest effect (Elangovan and Stuart, 2005). These findings were interpreted as showing the importance of high-frequency channels for masking release, a consequence of better temporal resolution at high than low frequencies. Whereas the conclusions of these two studies by Stuart and his colleagues (Scott *et al.*, 2001; Elangovan and Stuart, 2005) are contradictory to the present results, the finding of comparable masking release above and below 1700 Hz in the present study is consistent with the results of Oxenham and Simonson (2009). In one set of conditions in that study, sentence recognition was assessed in speech-shaped noise and noise that was modulated by the envelope of a one-talker masker. Masking release was comparable for LP and HP filter conditions, a result supported and extended in the present experiments.

2. Effect of modulation rate on filtered stimuli

The effect of modulation rate was very similar across filter conditions, with largest effects at low rates for detection and 3AFC-ID. Masking release in the open-ID conditions was smaller, but increased as a function of masker AM rate. This general trend in masking release was similar across filter conditions and was also seen in the results of experiment 1 with full-spectrum stimuli. There was some indication of a reduced effect of masker AM rate on the HP as compared to previous full-spectrum data, but this difference was not apparent in the HP+16 condition. These results are consistent with the conclusion that the task-by-rate effect was consistent across low- and high-frequency regions of the speech signal, aside from a modest level effect.

One expectation touched on above, and proposed by Oxenham and Simonson (2009), is that masking release as-

sociated with masker AM requires some degree of cue redundancy relative to the cues necessary to perform the speech task; severely filtering the speech signal could reduce that redundancy and therefore reduce masking release. To the extent that LP and HP filtering the speech eliminates some of that redundancy, periodically unmasking “glimpses” of the signal could be less beneficial to performance. There was little evidence of a substantially reduced masking released for filtered stimuli in the current data set, however. Whereas the mean masking release associated with the open-ID, LP condition was approximately 2 dB less than that in the comparable full-spectrum conditions of experiment 1, this difference was not significant. It is possible that increasing the number of subjects might have revealed a small but significant effect. Whereas a reduction in redundancy might affect the optimal masker AM rate, it is possible that LP or HP filtering at 1700 Hz did not sufficiently reduce redundancy to allow such an effect to be observed. Future work will pursue more severe filtering conditions to assess this possibility.

IV. GENERAL DISCUSSION

In the present studies using CNC words, the pattern of masking release observed with the introduction of sinusoidal masker AM depends on the observer’s task. Consistent with previous results obtained using low- and high-redundancy speech materials, lower rates of AM support better performance when the degree of detail required to correctly perform the task is relatively coarse, and high rates support better performance when fine detail is required. A similar task-by-AM rate interaction is seen when the signal (instead of the masker) is amplitude modulated. Therefore, the interaction between masker AM rate and task is unlikely to reflect the differential effects of forward masking at different SNRs. The pattern of masking release as a function of AM rate is relatively unaffected by LP or HP filtering the speech at 1700 Hz, provided that the data are compared for similar masker levels. This result was interpreted as indicating that absolute frequency effects, such as ability to encode temporal fine structure at low frequencies or hypothetically superior temporal resolution at high frequencies, are not required to account for the pattern of masking release as a function of AM rate.

There is continuing interest in the finding that masker AM provides greatly reduced benefits for listeners with moderate sensorineural hearing impairment as compared to normal-hearing listeners (Lorenzi *et al.*, 2006; Hopkins and Moore, 2009). While audibility plays some role in this result, deficits in temporal or spectral resolution might also limit performance in this task (Eisenberg *et al.*, 1995; Peters *et al.*, 1998), including effects related to the coding of temporal fine-structure cues (e.g., Hopkins and Moore, 2009). One interpretation of these findings is that reduced spectral and temporal resolution could reduce cue redundancy, such that sparse glimpses at the speech signal are not sufficient to support recognition, an idea recently proposed by Oxenham and Simonson (2009). The paradigm of the experiments described here could provide a tool for exploring the role of redundancy in the hearing impaired population.

The present results demonstrating an interaction between masker AM rate and task are difficult to reconcile with efforts to predict speech performance using the speech intelligibility index (SII) (ANSI, 1997). This model uses estimates of audibility of various spectral regions of the speech signal, in combination with speech importance functions, to compute a SII; the relationship between SII and percent correct is then obtained empirically. This basic model has been adapted for use with non-stationary maskers, by either averaging the SII associated with masker modulation maxima and minima (e.g., Horwitz *et al.*, 2007) or by computing the instantaneous SII as a function of time (Rhebergen and Versfeld, 2005; Rhebergen *et al.*, 2006). While not inherent to the SII model, context effects can be incorporated by taking into account the fact that the function relating percent correct to SII is steeper in high than low-predictability speech materials (Hargus and Gordon-Salant, 1995).

These adaptations of the SII model for use with a non-stationary signal are not consistent with the interaction between task and masker modulation rate observed here. Because the SII is based on audibility, it is not sensitive to the distribution of cues over time. An effect of modulation rate would be predicted by a SII model incorporating the limitations to audibility associated with forward masking and temporal resolution (such as Rhebergen *et al.*, 2006), but the results of the supplemental conditions of experiment 1 indicate that forward masking is not likely to be responsible for the task-by-rate interaction. Findings of the present study suggest that models of speech perception in modulated noise might benefit from inclusion of information regarding the temporal distribution of speech cues, including the degree of cue redundancy relative to the other sources of information available to the observer.

ACKNOWLEDGMENTS

This work was supported by a grant from NIH NIDCD (Grant No. R01 DC000418). We thank associate editor Ken Grant and two anonymous reviewers for their helpful comments and suggestions.

¹Pilot testing of the self-scoring method indicated that verbalizing the response helped the subject commit to an unambiguous response prior to scoring. This method was judged to be superior to the more standard procedure, where an experimenter outside the booth monitors subject responses and performs scoring, because of difficulties differentiating odd pronunciation from errors in speech recognition and the possibility of experimenter bias. Whereas self-scoring could conceivably introduce errors, there is no reason to believe that such errors would vary systematically across stimulus conditions of the present study, as all observers were blind to the predictions of the study. Because the primary data of interest involved the pattern of results across masker conditions, any inaccuracy introduced by self-scoring is very unlikely to affect the results presented here.

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

ANSI (1997). *ANSI S3.5-1997, American National Standards Methods for Calculation of Speech Intelligibility Index* (American National Standards Institute, New York).

Bacon, S. P., Lee, J., Peterson, D. N., and Rainey, D. (1997). “Masking by modulated and unmodulated noise: Effects of bandwidth, modulation rate, signal frequency, and masker level.” *J. Acoust. Soc. Am.* **101**, 1600–1610.

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.
- Baer, T., and Moore, B. C. J. (1994). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," *J. Acoust. Soc. Am.* **95**, 2277–2280.
- Bonatti, L. L., Pena, M., Nespore, M., and Mehler, J. (2005). "Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing," *Psychol. Sci.* **16**, 451–459.
- Bronkhorst, A. W., Bosman, A. J., and Smoorenburg, G. F. (1993). "A model for context effects in speech recognition," *J. Acoust. Soc. Am.* **93**, 499–509.
- Chang, J. E., Bai, J. Y., and Zeng, F. G. (2006). "Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise," *IEEE Trans. Biomed. Eng.* **53**, 2598–2601.
- de Laat, J. A. P. M., and Plomp, R. (1983). "The reception threshold of interrupted speech for hearing-impaired listeners," *Hearing, Physiological Bases and Psychophysics, Proceedings of the Sixth International Symposium on Hearing* (Springer-Verlag, Berlin), pp. 357–363.
- DePaolis, R. A., Janota, C. P., and Frank, T. (1996). "Frequency importance functions for words, sentences, and continuous discourse," *J. Speech Hear. Res.* **39**, 714–723.
- Dirks, D. D., and Bower, D. (1970). "Effect of forward and backward masking on speech intelligibility," *J. Acoust. Soc. Am.* **47**, 1003–1008.
- Dirks, D. D., and Bower, D. R. (1971). "Influence of pulsed masking on spondee words," *J. Acoust. Soc. Am.* **50**, 1204–1207.
- Dirks, D. D., Wilson, R. H., and Bower, D. R. (1969). "Effect of pulsed masking on selected speech materials," *J. Acoust. Soc. Am.* **46**, 898–906.
- Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (1995). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing," *J. Speech Hear. Res.* **38**, 222–233.
- Elangovan, S., and Stuart, A. (2005). "Interactive effects of high-pass filtering and masking noise on word recognition," *Ann. Otol. Rhinol. Laryngol.* **114**, 867–878.
- Festen, J. M. (1987). "Speech-reception threshold in a fluctuating background sound and its possible relation to temporal auditory resolution," *The Psychophysics of Speech Perception* (Nijhoff, Dordrecht, The Netherlands), pp. 461–466.
- Festen, J. M. (1993). "Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice," *J. Acoust. Soc. Am.* **94**, 1295–1300.
- Festen, J. M., and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *J. Acoust. Soc. Am.* **88**, 1725–1736.
- Fullgrabe, C., Berthommier, F., and Lorenzi, C. (2006). "Masking release for consonant features in temporally fluctuating background noise," *Hear. Res.* **211**, 74–84.
- Gustafsson, H. A., and Arlinger, S. D. (1994). "Masking of speech by amplitude-modulated noise," *J. Acoust. Soc. Am.* **95**, 518–529.
- Haggard, M. P., Hall, J. W., and Grose, J. H. (1990). "Comodulation masking release as a function of bandwidth and test frequency," *J. Acoust. Soc. Am.* **88**, 113–118.
- Hall, J. W., III, Buss, E., and Grose, J. H. (2008). "The effect of hearing impairment on the identification of speech that is modulated synchronously or asynchronously across frequency," *J. Acoust. Soc. Am.* **123**, 955–962.
- Hargus, S. E., and Gordon-Salant, S. (1995). "Accuracy of speech intelligibility index predictions for noise-masked young listeners with normal hearing and for elderly listeners with hearing impairment," *J. Speech Hear. Res.* **38**, 234–243.
- Hawkins, J. E., and Stevens, S. S. (1950). "The masking of pure tones and of speech by white noise," *J. Acoust. Soc. Am.* **22**, 6–13.
- Henry, B. A., McDermott, H. J., McKay, C., James, C. J., and Clark, G. M. (1998). "A frequency importance function for a new monosyllabic word test," *Aust. J. Audiol.* **20**, 79–86.
- Hopkins, K., and Moore, B. C. (2009). "The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise," *J. Acoust. Soc. Am.* **125**, 442–446.
- Horwitz, A. R., Ahlstrom, J. B., and Dubno, J. R. (2007). "Speech recognition in noise: Estimating effects of compressive nonlinearities in the basilar-membrane response," *Ear Hear.* **28**, 682–693.
- Howard-Jones, P. A., and Rosen, S. (1993). "The perception of speech in fluctuating noise," *Acustica* **78**, 258–272.
- Huggins, A. W. (1964). "Distortion of the temporal pattern of speech: Interruption and alternation," *J. Acoust. Soc. Am.* **36**, 1055–1064.
- 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.
- Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (1977). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *J. Acoust. Soc. Am.* **61**, 1337–1351.
- Kwon, B. J., and Turner, C. W. (2001). "Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?," *J. Acoust. Soc. Am.* **110**, 1130–1140.
- Lee, J., and Bacon, S. P. (1998). "Psychophysical suppression as a function of signal frequency: Noise and tonal maskers," *J. Acoust. Soc. Am.* **104**, 1013–1022.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (2006). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18866–18869.
- Miller, G. A., Heise, G. A., and Lichten, W. (1951). "The intelligibility of speech as a function of the context of the test materials," *J. Exp. Psychol.* **41**, 329–335.
- Miller, G. A., and Licklider, J. C. R. (1950). "The intelligibility of interrupted speech," *J. Acoust. Soc. Am.* **22**, 167–173.
- Müsch, H., and Buus, S. (2001). "Using statistical decision theory to predict speech intelligibility. I. Model structure," *J. Acoust. Soc. Am.* **109**, 2896–2909.
- 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.
- Oxenham, A. J., and Simonson, A. M. (2009). "Masking release for low- and high-pass filtered speech in the presence of noise and single-talker interference," *J. Acoust. Soc. Am.* **125**, 457–468.
- Oxenham, A. J., and Simonson, A. M. (2009). "Masking release for low- and high-pass filtered speech in the presence of noise and single-talker interference," *J. Acoust. Soc. Am.* **125**, 457–468.
- Parikh, G., and Loizou, P. C. (2005). "The influence of noise on vowel and consonant cues," *J. Acoust. Soc. Am.* **118**, 3874–3888.
- 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**, 577–587.
- Peterson, G. E., and Lehiste, I. (1962). "Revised CNC lists for auditory tests," *J. Speech Hear. Disord.* **27**, 62–70.
- Phatak, S. A., and Allen, J. B. (2007). "Consonant and vowel confusions in speech-weighted noise," *J. Acoust. Soc. Am.* **121**, 2312–2326.
- Plomp, R. (1964). "Rate of decay of auditory sensation," *J. Acoust. Soc. Am.* **36**, 277–282.
- Powers, G. L., and Wilcox, J. C. (1977). "Intelligibility of temporally interrupted speech with and without intervening noise," *J. Acoust. Soc. Am.* **61**, 195–199.
- 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.
- Qin, M. K., and Oxenham, A. J. (2005). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," *Ear Hear.* **26**, 451–460.
- Qin, M. K., and Oxenham, A. J. (2006). "Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech," *J. Acoust. Soc. Am.* **119**, 2417–2426.
- Rhebergen, K. S., and Versfeld, N. J. (2005). "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," *J. Acoust. Soc. Am.* **117**, 2181–2192.
- Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (2006). "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," *J. Acoust. Soc. Am.* **120**, 3988–3997.
- Scott, T., Green, W. B., and Stuart, A. (2001). "Interactive effects of low-pass filtering and masking noise on word recognition," *J. Am. Acad. Audiol.* **12**, 437–444.
- Stuart, A., and Phillips, D. P. (1996). "Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-

- hearing, and presbycusis listeners," *Ear Hear.* **17**, 478–489.
- Stuart, A., and Phillips, D. P. (1997). "Word recognition in continuous noise, interrupted noise, and in quiet by normal-hearing listeners at two sensation levels," *Scand. Audiol.* **26**, 112–116.
- Studebaker, G. A., Pavlovic, C. V., and Sherbecoe, R. L. (1987). "A frequency importance function for continuous discourse," *J. Acoust. Soc. Am.* **81**, 1130–1138.
- Studebaker, G. A., Taylor, R., and Sherbecoe, R. L. (1994). "The effect of noise spectrum on speech recognition performance-intensity functions," *J. Speech Hear. Res.* **37**, 439–448.
- Toro, J. M., Shukla, M., Nespor, M., and Endress, A. D. (2008). "The quest for generalizations over consonants: Asymmetries between consonants and vowels are not the by-product of acoustic differences," *Percept. Psychophys.* **70**, 1515–1525.
- van Rooij, J. C., and Plomp, R. (1991). "The effect of linguistic entropy on speech perception in noise in young and elderly listeners," *J. Acoust. Soc. Am.* **90**, 2985–2991.