Simulations of cochlear-implant speech perception in modulated and unmodulated noise

Antje Ihlefeld and John M. Deeks
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

Patrick R. Axon
Addenbrooke's NHS Trust, Hills Road, Cambridge CB2 0QQ, United Kingdom

Robert P. Carlyon
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

(Received 3 September 2008; revised 21 May 2010; accepted 3 June 2010)

Experiment 1 replicated the finding that normal-hearing listeners identify speech better in modulated than in unmodulated noise. This modulated-unmodulated difference ("MUD") has been previously shown to be reduced or absent for cochlear-implant listeners and for normal-hearing listeners presented with noise-vocoded speech. Experiments 2–3 presented normal-hearing listeners with noise-vocoded speech in unmodulated or 16-Hz-square-wave modulated noise, and investigated whether the introduction of simple binaural differences between target and masker could restore the masking release. Stimuli were presented over headphones. When the target and masker were presented to one ear, adding a copy of the masker to the other ear ("diotic configuration") aided performance but did so to a similar degree for modulated and unmodulated maskers, thereby failing to improve the modulation masking release. Presenting an uncorrelated noise to the opposite ear ("dichotic configuration") had no effect, either for modulated or unmodulated maskers, consistent with the improved performance in the diotic configuration being due to interaural decorrelation processing. For noise-vocoded speech, the provision of simple spatial differences did not allow listeners to take greater advantage of the dips present in a modulated masker.

[DOI: 10.1121/1.3458817]

PACS number(s): 43.71.Ky, 43.66.Ts, 43.66.Pn, 43.66.Dc [RLF] Pages: 870–880

I. INTRODUCTION

Normal-hearing (NH) listeners identify target speech better in modulated noise than in unmodulated noise (Miller and Licklider, 1950; Festen and Plomp, 1990; Howard-Jones and Rosen, 1993a, 1993b; Summers and Molis, 2004; but see also Kwon and Turner, 2001; Apoux and Bacon, 2008). However, this benefit is reduced or absent in cochlear implant (CI) users and in NH listeners identifying "vocode"-simulations of speech processed by CIs (Qin and Oxenham, 2003; Nelson *et al.*, 2003; Nelson and Jin, 2004; Stickney *et al.*, 2004; Fu and Nogaki, 2005; Jin and Nelson, 2006). The release from masking for modulated versus unmodulated noise has been attributed to "dip-listening," a putative mechanism by which the auditory system identifies dips in the temporal envelope of the masking signal, allowing it to weight segments of the acoustic mixture with more favorable Target-to-Masker-ratios (TMRs) more strongly than those with lower TMRs (Buus, 1985). One possibility, already suggested by others (Kwon and Turner, 2001; Nelson *et al.*, 2003; Qin and Oxenham, 2003; Stickney *et al.*, 2004; Fu and Nogaki, 2005; Oxenham and Simonson, 2009), is that, even when the valleys in a modulated masker are preserved, CI processors and vocode simulations do not allow the listener effectively to segregate the target from the acoustic mixture.[1] This may impair a listener's ability to judge which parts of the acoustic mixture have predominantly target energy. The present study considers whether simple binaural differences between target and masker can improve a listener's ability to listen to speech in the dips of a fluctuating masker.

Previous studies using unprocessed speech presented in free field to normal-hearing (NH) listeners have shown a large advantage when the target and masker are spatially separated, compared to when they are presented from the same location (for reviews see, e.g., Bronkhorst, 2000; Darwin 2008). This advantage arises for a number of reasons, including the head shadow effect, a reduction in similarity and uncertainty produced by differences in perceived location between the target and masker, and, particularly at negative target-to-masker-ratios (TMRs), interaural decorrelation produced by the target (Levitt and Rabiner, 1967; Zurek 1993; Freyman *et al.*, 1999; Arbogast *et al.*, 2002; Brungart and Simpson, 2002; Shinn-Cunningham *et al.*, 2005; Edmonds and Culling, 2006). A smaller, but still significant, spatial release from masking has been observed in bilateral CI users (Tyler *et al.*, 2002; Gantz *et al.*, 2002; Müller *et al.*, 2002; van Hoesel and Tyler, 2003; Schleich *et al.*, 2004; Long *et al.*, 2006; Buss *et al.*, 2008; Loizou *et al.*, 2009). Two recent studies have also shown a masking release when NH listeners are presented with vocoded speech. Freyman *et al.* (2008) measured performance for noise-vocoded speech presented over loudspeakers in the presence of babble. They found that spatially separating the target and masker improved performance when listeners detected isolated words, but not when required to report whole sentences, and attributed this difference to the lower overall TMR in the former case. It is possible that spatial masking effects are reduced at high TMRs because the louder target might allow subjects to segregate it from the masker even in the absence of spatial cues; in addition, interaural decorrelation effects introduced by adding a spatially separated target to a masker are likely to be greatest at low TMRs (Brungart, 2001; Brungart *et al.*, 2001; Freyman *et al.*, 2004; Ihlefeld and Shinn-Cunningham, 2008a). Garadat *et al.* (2009) required NH listeners to identify spondees presented over headphones in a background of concatenated sentences, with all stimuli sine-wave-vocoded either before or after convolution with head related transfer functions (HRTFs) that simulated various spatial locations of the target and the masker. They also observed a spatial release from masking. Interestingly, spatial release did not depend markedly on whether the vocoding took place before the HRTFs—in which case interaural temporal fine structure (TFS) cues were preserved—or after the HRTFs, in which case interaural TFS cues were absent (Shinn-Cunningham *et al.*, 2005; Drennan *et al.*, 2007; Garadat *et al.*, 2009).

Although the above studies shed light on how listeners, under conditions broadly similar to CI processing, may be able to exploit binaural cues, they each used only a single type of masker. As such, it is hard to differentiate between an improved ability to listen in the dips of the masker, and other advantages—such as the processing of interaural decorrelation and differences in spectral shape that can be introduced by spatial differences—that do not require the listener selectively to weight different portions of the input waveform. In the present study, we specifically investigate this "dip listening" mechanism, under different spatial configurations, using two maskers—noise that is either square-wave modulated at 16 Hz or unmodulated—designed to differ maximally in the extent to which a dip-listening strategy is useful.

In Experiment 1, establishing a control condition, we measured the difference in performance between modulated and unmodulated masker conditions (modulated-unmodulated difference, "MUD;" Carlyon *et al.*, 1989) for identification of unprocessed speech in noise with monaural presentation. Experiments 2 and 3 measured MUD for identification of noise-vocoded speech in noise, and tested whether it could be increased by the introduction of simple spatial differences between target and masker. This included a comparison of configurations under which interaural correlation cues based on the masker fine structure were present or absent.

II. EXPERIMENT 1—FULL SPEECH IN MODULATED NOISE

A. Stimuli

All stimuli consisted of a speech target and noise masker in the right ear, and were processed using MATLAB 7.1 (The Mathworks Inc., Natick, MA) prior to the experiment.

1. Targets

Speech stimuli were derived from a recording of the Coordinate Response Measure ("CRM") corpus with British talkers (Bolia *et al.*, 2000; Kitterick and Summerfield, 2007). Only utterances from the four male talkers of the corpus were used throughout the study. Sentences were of the form "Ready ⟨*call sign*⟩, go to ⟨*color*⟩ ⟨*number*⟩ now." ⟨*Color*⟩ was one of the set [white, red, blue, and green]. ⟨*Number*⟩ was one of the digits between one and eight. ⟨*Call sign*⟩ was one of [Arrow, Baron, Charlie, Eagle, Hopper, Laker, Ringo, and Tiger]. Listeners were instructed to ignore the call sign. This was done so that the keywords ⟨*color*⟩⟨*number*⟩ formed a limited well-known set with relatively little decision-making and working-memory demands.

Utterances were time-windowed at the beginning and end of each recording (2-ms squared cosine windows). Each utterance was processed by four band pass filters (4th-order Butterworth; 24 dB per octave attenuation) with 3-dB-down points at 100–300, 300–500, 500–1700, and 1700–6000 Hz; these cut-off frequencies were the same as in Nelson *et al.* (2003). To produce the speech target stimuli used in Experiment 1, the four narrow bands of speech were summed. For simplicity, we refer to these sounds as "unprocessed" speech, in order to distinguish them from the vocoded stimuli used in Experiments 2 and 3. Finally, the broadband root-mean square (RMS) was equalized across all utterances.

2. Maskers

On each trial, the presentation of the masker began 250 ms before the onset of the target utterance, and stopped together with the target. There were two types of masker: unmodulated and modulated noise. For each target speech utterance in the corpus, matched unmodulated noise maskers were generated by processing tokens of uniformly distributed white noise with the same four band-pass filters that were used for the target speech stimuli. In each frequency band, the RMS of the noise was equalized to the RMS in the corresponding band of the target speech utterance. Afterwards, the four bands were summed, creating unmodulated, spectrally-matched noise maskers.

To generate modulated noise maskers, for each token, an unmodulated noise masker token was modulated with 2-msec cosine-squared windows at a rate of 16 Hz (50% duty cycle, 100% modulation depth).[2] The modulated masker was then scaled such that its RMS equaled that of the corresponding token of the unmodulated masker. Therefore, the peak level of the modulated noise was 3 dB higher than that of the unmodulated noise. An alternative method would have been to equate the peak level of the modulated and unmodulated noises, thereby causing the RMS to be lower in the unmodulated case. We chose not to do this because, if listeners were not able to listen selectively in the dips (e.g., by smoothing the input over a temporal window longer than the modulation period), then performance would still be higher in the modulated than in the unmodulated noise.

For both unmodulated and modulated maskers, 1024 tokens of noise were generated off-line, and randomly drawn with replacement on each trial.

B. Listeners

Eight normal-hearing, fluent speakers of British English (ages 18 to 42, average age 27, median age 22) were paid to participate. Throughout all experiments in this study, all lis-

TABLE I. Listeners and the experiments that they had participated in.

	Experiment		
Listener Number	1 Full speech	2 Noise vocoded	3 Noise vocoded
1	x	x	x
2	x	x	x
3	x	x	x
4	x		x
5	x		x
6	x		x
7	x		
8	x		
9		x	
10		x	
11		x	
12		x	
13		x	
14		x	
15			x
16			x
17			x
Total Number of listeners	8	9	9

teners had pure-tone thresholds in quiet of less than 20 dB HL between 250 Hz and 4 kHz in both ears, as determined by 2I-2AFC adaptive threshold measurements. All listeners gave written informed consent prior to the experiment. All but two listeners who participated in Experiment 1 had previously participated in Experiments 2 and/or 3. Table I lists all listeners and the experiments that they had participated in.

C. Procedures

Stimuli were D/A converted with a sound card (Turtle Beach Sonic Fury; 16 bit resolution, 44.1 kHz sampling rate) and amplified using separate programmable attenuators (TDT PA4) and different channels of a headphone buffer (TDT HB6) for each ear. Stimuli were then presented over Sennheiser HD 250 II headphones to the listener seated in a double-walled sound-treated booth. The target and masker were presented monotically to the right ear. The masker level was fixed at 46 dB SPL; the target level was set to 30, 38 or 46 dB SPL, resulting in TMRs of −16, −8 and 0 dB. The order in which TMRs were presented was randomized across trials within a block such that each TMR was played once before all of them were played again in a new random order. The left ear contained no signal.

The TMR and the talker voice of the original utterance were randomly selected on each trial. The task was 32-alternative forced-choice closed-set speech identification. Throughout all experiments in this study, listeners were instructed: "Report the color and number you heard on the right side." Listeners were instructed to ignore the masker in the right ear. Following each trial, listeners indicated perceived target keywords using a graphical user interface (GUI), after which the GUI indicated the correct response. Correct-answer feedback was provided after each trial. A trial was scored as correct and listeners were given feedback that they were correct if and only if they reported both target keywords (color and number).

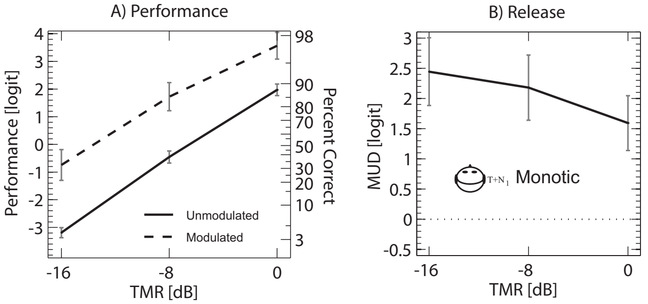

FIG. 1. Experiment 1 (noise masker in the right ear, and unprocessed speech in the right ear). (A) Mean correct performance in unmodulated and modulated noise (solid and dashed lines, respectively). (B) Mean masking release (logit-transformed performance in modulated minus unmodulated noise). Errorbars show 95% confidence intervals of across-listener means. Chance performance was 3%.

Listeners were each tested in a single session lasting two hours. The session consisted of 17 blocks of about five minutes duration each. One block contained 36 trials. Listeners performed 102 trials each for both masker conditions, at −16 dB, −8 dB and 0 dB TMR.

D. Data analysis

Although some previous studies have analyzed data in terms of percent correct scores, we chose to analyze our data in terms of logit-transformed scores. These scores conformed better with the homoscedasticity assumption of the analyses of variance (ANOVAs), and can partially counteract the floor and ceiling effects inherent in percent-correct scores (e.g., Morrison and Kondaurova, 2009). The logit-transform for probability correct p is $\mathrm{logit}(p)=\ln[p/(1-p)]$, transforming the percent correct scores into a variable of theoretically infinite range. Chance performance was 1/32. To avoid undefined values of the logit transform, $p<1/32$ was set to 1/32 (3%), and $p>1\text{-}1/N$ was set to 1–1/N (98%), where N was 51, a conservative value for the number of trials. For each masker configuration in each experiment we calculated the MUD from the (vertical) difference between logit-transformed psychometric functions for modulated and unmodulated noise. Throughout this study, statistical analyses were also calculated with raw percent correct scores, and led to similar conclusions.

E. Results

For each of the eight listeners, percent correct was calculated separately for each noise condition as a function of TMR. Figure 1(A) shows the group mean proportion of correct responses. The left ordinate shows the proportion in logit-units; the right ordinate is in units of percent correct. Error bars show the 95% confidence interval around the mean (1.96 times the standard error of the mean across listeners). A repeated measures two-way ANOVA on the logit transformed percent correct scores showed a significant release from masking for modulated versus unmodulated noise $[F(1,7)=86.890,\ p<0.001]$.

TABLE II. Across-listener average slopes of line fits of logit-transformed percent correct scores for each experiment and stimulus condition. Values in round brackets are 95% confidence intervals of the across-listener mean.

Stimulus condition		Slope (1/dB)
Experiment 1. TARGET: UNPROCESSED SPEECH *anechoic target and masker*		
Masker	Monotic unmodulated noise	0.3226 (0.1185)
	Monotic modulated noise	0.2693 (0.1522)
Experiment 2. TARGET: NOISE-VOCODED SPEECH *anechoic target and masker*		
Masker	Monotic unmodulated noise	0.1957 (0.1154)
	Monotic modulated noise	0.1515 (0.1038)
	Diotic unmodulated noise	0.1867 (0.1174)
	Diotic modulated noise	0.1309 (0.1112)
Experiment 2. TARGET: NOISE-VOCODED SPEECH *reverberant target and masker*		
Masker	Monotic unmodulated noise	0.1869 (0.0994)
	Monotic modulated noise	0.1318 (0.1021)
	Diotic unmodulated noise	0.1627 (0.0944)
	Diotic modulated noise	0.1127 (0.1129)
Experiment 3. TARGET: NOISE-VOCODED SPEECH *anechoic target and masker*		
Masker	Diotic unmodulated noise	0.2130 (0.1107)
	Diotic modulated noise	0.1626 (0.1289)
	Dichotic unmodulated noise	0.16255 (0.12892)
	Dichotic modulated noise	0.17713 (0.12666)

There was also a significant interaction between TMR and noise condition $[F(2,14)=5.579,\ p=0.017]$, likely to be caused by ceiling effects at 0 dB in the modulated masker conditions. Secondary two-tailed paired t-tests found significant differences between modulated and unmodulated maskers at all TMRs ($df=7$, $t=8.559$, 7.947 and 6.647 at -16, -8 and 0 dB TMR, respectively; $p<0.001$ at all TMRs).

Logit transformed percentage correct scores were fitted with lines using a minimum least-squares method (command polyfit in Matlab 7.4.0, The Mathworks, Natick, MA). Table II lists across-subject averages of the slopes of these fits. A two-tailed paired t-test revealed that slopes were significantly steeper in unmodulated than in modulated noise ($df=7$, $t=2.966$, $p=0.021$).

Figure 1(B) shows the modulated-unmodulated difference (MUD), defined as the difference between the logit transformed percent correct scores for performance with modulated and unmodulated maskers, establishing a baseline condition for Experiments 2 and 3. As noted above, ceiling effects may have influenced scores in the modulated condition at the 0 dB TMR, and this may explain the lower MUD at that TMR.

F. Discussion

Previous studies show that when target speech is masked by concurrent noise, performance generally improves when the noise is modulated compared to when the noise is unmodulated (for a review see, e.g., Assmann and Summerfield, 2004). Experiment 1 confirmed these findings. Using British recordings of the CRM corpus, mixed with either unmodulated or 16-Hz-modulated noise, presented to the right ear, here we found that performance was better with 16-Hz-modulated noise than with unmodulated noise, similar to the results of previous experiments (e.g., Nelson *et al.*, 2003, Nelson and Jin, 2004; Fu and Nogaki, 2005). Moreover, the psychometric functions were shallower with modulated than with unmodulated noise maskers. For half of the time, during the dips of the masker, the target is energetically masked only through non-simultaneous masking from the peaks of the masker, making overall performance less steeply dependent on masker energy. Moreover, the flattening of the modulated performance function is consistent with the idea that modulated noise is perceptually more similar to the speech stimuli than unmodulated noise, perhaps causing listeners to confuse the noise fluctuations with those of the speech (Kwon and Turner, 2001; Qin and Oxenham, 2003; Stickney *et al.*, 2004; Apoux and Bacon, 2008). This modulation interference may have increased variability in listeners' detection and identification of target speech, flattening the psychometric functions in modulated compared to unmodulated noise conditions (cf., Lutfi *et al.*, 2003).

In addition, the duration of the noise bursts in the modulated noise condition was 31.25 ms; the duration of each of the keywords (colors or numbers) was roughly between 160 and 350 msec. By chance from trial to trial, phonetic features that listeners used to identify the speech tokens were or were not energetically masked by the modulated masker, increasing randomness in the responses and further decreasing the slope of the psychometric function (cf., Howard-Jones and Rosen, 1993a).

Overall, Experiment 1 confirmed that, in the current paradigm, NH listeners can take advantage of the dips in a modulated masker. In contrast, CI listeners often struggle in this type of task. The following experiments were motivated by the idea that part of the difficulty that CI listeners experience when listening to target speech in modulated background interferers results from their reduced ability to segregate target and interferers. Perhaps this problem is aggravated by the fact that most CI users are implanted only on one side. For both NH and HI listeners, spatial differences can improve segregation of perceptually similar sources (Arbogast, *et al.* 2002). Implanting CI users bilaterally should give them access to spatial cues, adding perceptual evidence for segregating competing sources. In fact, if spatial cues provide a particular "dip listening" benefit when listening in temporally fluctuating maskers, then the benefits of bilateral implantation might be underestimated when studied using unmodulated noise maskers.

For Experiment 2, noise-vocoded speech was presented to the right ear, together with either modulated or unmodulated monotic or diotic noise. Based on the results in the literature and because noise-vocoded speech is perceptually more similar to noise maskers than unprocessed speech, we expected to see less MUD for noise-vocoded speech than for the unprocessed speech stimuli in Experiment 1. Our hypothesis was that introducing spatial differences between target and masker would improve segregation and therefore increase MUD for noise-vocoded speech in noise, causing greater MUD when the spatial features of target and masker differed (monotic target and diotic noise), than when their spatial features were similar (monotic target and monotic noise).

III. EXPERIMENT 2—NOISE VOCODED SPEECH, SIMPLE SPATIAL CUES

A. Stimuli

The targets were noise-vocoded speech, generated from the four bands of speech used in Experiment 1. The envelope of each narrow band of speech was extracted with half-wave rectification, followed by low-pass filtering with 50-Hz cutoff frequency and 24 dB/octave frequency roll-off [processing after Nelson *et al.* (2003), except that here, a lower cutoff frequency was chosen for the envelope extraction to keep voice pitch cues in the envelope minimal]. Each envelope was then multiplied with a noise carrier. To generate the noise carriers, for each utterance, a token of uniformly distributed white noise was generated and processed with the same four band-pass filters that were used for the speech stimuli in Experiment 1. The four amplitude-modulated bands were then summed to produce the vocoded target speech signal. Finally, the broadband RMS was equalized across all utterances and set to the same value as in the processing regimes of Experiment 1. These vocoded speech signals were presented to each listener's right ear. Modulated or unmodulated maskers, similar to those in Experiment 1, were presented either to the right ear only or diotically.

B. Listeners

Nine normal-hearing, fluent speakers of British English (ages 18 to 42, average age 27, median age 22) were paid to participate (see Table I). Listeners completed two sessions on two different days; each session consisting of 17 blocks of 48 trials each. Overall, they performed 51 trials in each noise condition and binaural configuration at −16, −8, 0 and 8 dB TMR. In addition, at the beginning of the first session, listeners completed one block of 48 trials of listening to vocoded speech stimuli in the right ear in quiet. Testing in quiet served two purposes: to establish a baseline performance and to familiarize the listener with the experimental task. All of the listeners were extremely quick in picking up the task. Therefore, we did not include any additional practice trials.

C. Procedures

The measurements described here were interleaved with another condition, in which the stimuli were processed so as to simulate mild reverberation. The results obtained in that condition showed a similar pattern to that observed in the main conditions described here, and are presented in the Appendix.

A run of four or five consecutive blocks always had the same masker configuration consisting of either monotic or diotic noise. The order of the conditions was randomized across listeners, and sessions were structured such that listeners heard five blocks of one configuration, then five blocks of the other configuration, followed by another four blocks for each configuration. TMR and the presence/absence of simulated reverberation were randomly selected from trial to trial such that all TMRs in each of the two reverberation cases were presented once before they were repeated.

In the monotic configuration, listeners were instructed to listen for the target in the right ear, and ignore the masker in the right ear. In the diotic configuration, listeners were instructed to listen for the target in the right ear and ignore the masker from the center of the head.

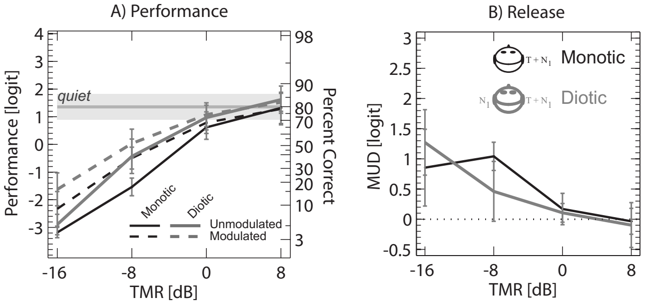

FIG. 2. Experiment 2, Anechoic conditions only. (A) Mean percent correct performance in unmodulated versus modulated noise (solid versus dashed lines, respectively) and for monotic and diotic noise maskers (black versus gray lines, respectively). (B) MUD. Errorbars show 95% confidence intervals of the across-listener mean.

D. Results

Figure 2(A) shows performance as a function of TMR for each noise condition (unmodulated vs. modulated; solid and dashed lines, respectively), and masker configuration (monotic vs. diotic noise; black and gray lines, respectively). As in Fig. 1, the ordinate is labeled both in logit and in percent correct units. The horizontal light gray lines show quiet performance, with gray shaded horizontal bars indicating the 95% confidence interval around the mean; the across-listener average of correct responses was 77%. Figure 2(B) shows MUDs (with error bars showing the 95% confidence interval around the mean).

A repeated measures ANOVA of logit-transformed percent correct scores, with within-listener factors of modulation condition, masker configuration, and TMR, showed main effects of all factors. Performance was significantly better in the modulated than in the unmodulated noise conditions [$F(1,8)=31.434$, $p=0.001$; dashed lines fall above solid lines in Fig. 2(A)]. Furthermore, performance was significantly better with the diotic than with the monotic masker [main effect of masker configuration: $F(1,8)=33.719$, $p<0.001$; dark lines Fig. 2(A) generally fall below gray lines].

The improvement in performance for the diotic compared to the monotic masker was greatest at −8 dB and −16 dB TMR, as indicated by a significant interaction between masker configuration and TMR on the logit-transformed percent correct scores [$F(3,24)=3.427$, $p=0.033$]. This is consistent with listeners exploiting interaural decorrelation cues in the diotic masker condition, and with the finding that these cues tend to be most valuable at low TMRs (Levitt and Rabiner, 1967; Zurek, 1993). Note that while performance is consistently better for diotic than for monotic noise configurations, at −16 dB TMR this difference is smaller than at −8 dB TMR, a result likely caused by a floor effect in the unmodulated noise condition at −16 dB TMR.

If our hypothesis that spatial cues help to listen in the dips was true, MUD should have been greater with diotic than with monotic masking. However, a repeated measures ANOVA on MUDs with masker configuration and TMR as factors found no main effect of masker configuration $[F(1,8)=0.293;\ p=0.603]$, but confirmed the interaction between masker configuration and TMR $[F(3,24)=3.593,\ p=0.028]$. Secondary paired two-tailed t-tests of MUDs at each TMR suggested that MUD might be *smaller* at −8 dB for the diotic than the monotic masker, but this finding did not survive Bonferroni correction (t-test, $df=8$, $t=2.411$, $p=0.042$ uncorrected, 0.17 corrected). Overall, the results provide no evidence that simple spatial differences between target and masker improve the listener's ability to utilize the advantageous TMR in the dips of the fluctuating noise used here.

When lines were fitted to the logit transformed data,[3] the slopes were shallower in modulated than in unmodulated noise. A repeated measures ANOVA on the fitted slopes of the logit-transformed percent correct scores confirmed this $[F(1,8)=47.775,\ 11.306;\ p<0.001,\ 0.01$ for main effect modulation condition and masker configuration, respectively].

E. Discussion

The MUDs obtained in Experiment 2 using noise-vocoded speech were smaller, both with monotic and diotic maskers, than those obtained with unprocessed speech in Experiment 1. This could be due to listeners having difficulty, during the gaps in the modulated noise, in distinguishing between the target speech and the masker. If so, then our finding that binaural cues failed to increase MUD suggests that they are not effective at improving a "dip listening" strategy. In this regard, it is worth pointing out that, both with monotic and diotic modulated maskers, listeners were aware that two sources were present; this sounded like a regular pulsing sound heard in the middle of the head or in the right ear, depending on the masker configuration, plus "something else." However, we should stress that being aware of two sources does not necessarily mean that listeners were able effectively to weight those portions of the input corresponding to the dips in the masker. Indeed, other aspects of the auditory scene analysis literature strongly suggest that grouping and segregation is not "all or none." For example, when a component of a harmonic complex is mistuned by about 3%, listeners clearly hear it as a separate source but it nevertheless contributes to the pitch of that complex (Moore *et al.*, 1985; Moore *et al.*, 1986; Ciocca and Darwin, 1999; Carlyon and Gockel, 2008). Another example comes from the fact that when two same-sex voices are combined, the listener can clearly hear that there are two people talking, but performance still improves when the speakers are of different sexes (Brungart, 2001).

A second interesting finding is that although diotic presentation did not increase MUD, it did improve performance overall. Hence, although performance in the presence of a modulated masker was improved by adding a copy of the masker to the other ear, Experiment 2 provides no evidence to suggest that this resulted from an improved ability to listen in the dips. The improvement could, instead, be due either to differences in perceived spatial location between the target and diotic masker (cf. Arbogast *et al.*, 2002; Kidd *et al.*, 2005b; Ihlefeld and Shinn-Cunningham, 2008b), with these differences being useful even for unmodulated maskers, or to the monotic target reducing the interaural correlation between the noise samples in each ear in the diotic configuration (e.g., Levitt and Rabiner, 1967; Zurek 1993; Edmonds and Culling 2005; Edmonds and Culling 2006). Experiment 3 aimed to distinguish between these explanations.

IV. EXPERIMENT 3—NOISE VOCODED SPEECH: THE EFFECTS OF LOCATION AND INTERAURAL DECORRELATION CUES

A. Rationale

In order to tease apart the effects of spatial separation and interaural decorrelation, Experiment 3 studied performance both with diotic and dichotic maskers. In this latter configuration, in which independent samples of noise were presented to the two ears, the perceived location of the noise was also different from that of the target speech, but no benefit from interaural correlation processing should occur. A particularly useful comparison is between performance in dichotic noise in Experiment 3 with that with monotic noise in Experiment 2; this allows us to study the effects of a large spatial location cue (albeit with a more diffuse spatial image than in the diotic configuration) in two conditions in which the target speech did not produce any interaural decorrelation.

B. Stimuli

Speech stimuli were similar to those in Experiment 2. Four types of masker, differing in whether they were diotic or dichotic and in whether they were modulated or unmodulated, were randomly interleaved from trial to trial.[4] In the diotic masker configuration, which was similar to the diotic condition in Experiment 2, the masker was presented identically to both ears, presumably producing a spatial image in the center of the listener's head. In the dichotic configuration the left ear masker was an independent token of noise that was statistically identical to the noise in the right ear, so that the spatial image of the noise presumably was centered at the listener's head, but less compact compared to diotic noise configurations. Both the diotic and the dichotic maskers were either unmodulated or 16-Hz modulated. When modulated, the localization of both maskers may have been further improved by the interaurally coherent onsets and offsets pro-

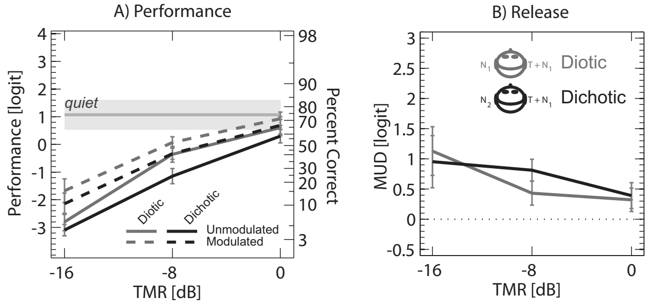

FIG. 3. Experiment 3 (noise maskers in both ears, and noise-vocoded speech in the right ear). (A) Mean percent correct performance for diotic and dichotic noise (gray and black lines, respectively) in unmodulated and modulated noise conditions (solid and dashed lines, respectively). (B) Mean masking release. Error bars show 95% confidence intervals of the across-listener mean.

duced by the cosine-tapered square-wave modulation. In addition, the presence of the signal would have interaurally decorrelated the masker envelopes.

C. Listeners

Nine normal-hearing, fluent speakers of British English (ages 18 to 42, average age 27, median age 22) were paid to participate (see Table I). Listeners completed two sessions on two different days, of 17 blocks with 45 trials each. Overall, they each performed 102 trials in each condition at −16 dB, −8 dB and 0 dB TMR. As in Experiment 2, we measured speech identification in quiet at the beginning of each session.

D. Results

The across-listener average performance in quiet was 72% correct. Percent correct for target speech in noise was calculated separately for each listeners and noise condition as a function of TMR. Figures 3(A) and 3(B) show the logit-transformed proportion of correct responses and MUD, respectively (with error bars showing the 95% confidence interval around the mean; the horizontal line and gray shaded bar show quiet performance and 95% confidence interval around mean quiet performance, respectively). Performance with the diotic and dichotic masker is shown by thick gray and black lines, respectively.

Performance was better in the diotic than in the dichotic noise configurations. A repeated-measures ANOVA on the logit-transformed percent correct scores with diotic/dichotic masker configuration, modulation and TMR as factors showed a main effect of masker configuration [$F(1,8)=51.02$, $p<0.001$]. The main effect of modulation was also significant [$F(1,8)=79.177$, $p<0.001$]. There were significant interactions between TMR and masker configuration (diotic versus dichotic), and between TMR and modulation condition [$F(2,16)=3.606;11.318$ and $p=0.051;0.001$, for TMR x masker and TMR x modulation, respectively]. This is consistent with floor effects in unmodulated dichotic noise at −16 dB, and with ceiling effects in diotic noise at 0 dB TMR.

To compare overall performance in the dichotic configuration with that observed with monaural maskers in Experiment 2, we performed two ANOVAs, one for the three participants common to both experiments, with the different configurations as a within-listener factor, and one for the remaining participants with the configurations treated as between-listener factors. In neither case was there a main effect of configuration [$F(1,2)=0.46$, $p=0.567$; $F(1,10)=0.014$, $p=0.910$], and combining these two probabilities using Stouffer's method (Stouffer *et al.*, 1949) gave a nonsignificant overall p value of 0.857. This suggests that a substantial change in the spatial location of the masker does not improve performance. This was also true when we considered performance only in the modulated conditions [Stouffer's test: $p=0.8515$; for $F(1,2)=0.203$, $p=0.696$, and $F(1,10)=0.048$; $p=0.832$]. In this latter case, the difference in perceived location between the modulated dichotic masker and the target may have been enhanced by the abrupt and regular onsets and offsets within the square-wave-modulated noise. Furthermore, the target speech would have produced an interaural decorrelation in the envelope of the masker. Despite this, performance for modulated noise was very similar in the dichotic and monotic conditions.

For each listener and condition, the logit-data were fitted with lines using the least-squares method (see Table II for slopes of the fits). Buttressing and extending results from Experiment 2, slopes were steeper in the unmodulated than in the modulated conditions and steeper in the dichotic than in the diotic configurations; however, statistical analysis only confirmed a significant effect of modulation [repeated measures ANOVA, with modulation and diotic/dichotic masker configuration as factors; $F(1,8)=18.501,0.712$; $p=0.003,0.423$].

E. Discussion

Here, targets and maskers always differed in their perceived spatialization, with the target sounding from the listener's right ear, and with the masker heard more centrally in the listener's head. Nevertheless, MUDs were substantially smaller than with the unprocessed speech of Experiment 1, and were no larger than with the purely monaural presentation of Experiment 2. Dichotic masking led to worse performance than diotic masking. Moreover, performance for dichotic noise was similar to that obtained in the monotic noise configuration of Experiment 2. This suggests that the superior performance in the diotic configuration, compared to that with monaural or dichotic maskers, did not result directly from perceived spatial differences between target and masker, and was indeed due to binaural decorrelation processing (e.g., Edmonds and Culling 2005).

V. GENERAL DISCUSSION

A. Dip listening

The results described here replicate previous reports that, for NH listeners, the MUD obtained for noise-vocoded speech is considerably smaller than that for unprocessed speech. Moreover, our results are consistent with recent work showing that MUD decreases both with overall performance level and with TMR, with a sweet spot occurring at lower TMRs where speech is in part audible and not perfectly intelligible (Gnansia *et al.*, 2008; Oxenham and Simonson, 2009; Bernstein and Grant, 2009). Therefore it is perhaps not too surprising that overall, MUDs were larger than those previously obtained at more positive TMRs and with babble noise (cf., Qin and Oxenham, 2003; Nelson and Jin, 2004; Whitmal *et al.*, 2007).

Our data also extend previous findings showing reduced modulation masking release with degraded speech (Kwon and Turner, 2001; Nelson *et al.*, 2003; Qin and Oxenham, 2003; Stickney *et al.*, 2004; Fu and Nogaki, 2005; Oxenham and Simonson, 2009). Here, results show that the introduction of simple spatial cues fails to restore the modulation advantage. Hence, if the reduced MUD observed for vocoded stimuli is indeed due to listeners "not knowing when to listen," then a spatial separation between the masker and target fails to alleviate this problem, at least for the stimuli used here. In the case of the diotic masker, a potential complication is that, for the modulated masker, the improvement in performance afforded by interaural decorrelation may have been reduced or absent during the dips in the masker. In contrast, the decorrelation advantage would have been present throughout the unmodulated masker, and so this could have reduced the MUD, counteracting any possible advantage to be gained from more effective dip-listening. However, this argument does not apply to the case with the dichotic masker, for which adding a contralateral copy of the masker also failed to increase the MUD. Indeed, the MUDs in the dichotic and diotic conditions were statistically indistinguishable $[F(1,8)=0.484;\ p=0.506]$.

B. Interaural decorrelation

As noted above, adding a copy of the masker to the contralateral ear produced a substantial improvement in speech identification in unmodulated noise. The fact that this benefit disappeared when the contralateral noise was uncorrelated with that in the target ear suggests that the advantage resulted from listeners exploiting interaural decorrelation cues in the diotic condition. Furthermore, no such advantage occurred for the dichotic modulated noise, compared with monaural modulated noise. In the first condition, masker fine structure was decorrelated across the ears. However, the 16-Hz modulation created a binaurally correlated envelope that was partially decorrelated by the introduction of the monotic target. This means that the present results provide no evidence that listeners can exploit interaural decorrelation in the envelope to extract speech from a fluctuating masker.

It is interesting to compare this finding with recent evidence that both NH and CI listeners can use envelope decorrelation in some circumstances. For instance, when NH listeners are required to detect a 125-Hz tone in a diotic 25-Hz wide narrowband noise centered on 125 Hz, with both the noise and tone "transposed" by halfwave rectification, low-pass filtering and multiplication with a 4-kHz sinusoid, thresholds are lower when the tone is out of phase at the two ears than when it is in phase (van de Par and Kohlrausch, 1997). A similar finding has been obtained for bilateral CI users: when stimulated with a CI processed mixture of diotic noise and a tone that was either binaurally out of phase or binaurally in phase, tone detection thresholds were better for the out-of-phase conditions (Long *et al.*, 2006).

These previously reported data show that both NH and CI listeners can use interaural decorrelation when detecting a signal. Here, listeners could not use interaural decorrelation cues from the envelopes of 16-Hz modulated noise to better extract speech information from the dips of the masker. The difference in the ability of our listeners and of those in previous studies to exploit interaural envelope decorrelation could be due to a number of factors, including modulation waveform shape, masker bandwidth, signal bandwidth, and task (speech recognition versus tone detection). One indication that it was not solely due to the higher modulation frequencies present in the maskers used in the previous studies (difference in modulation frequency: 16 Hz vs 125 Hz) comes from Long *et al.*'s finding that the binaural sensitivity of their CI users was most pronounced for envelopes fluctuations slower than about 50 Hz. An important issue for future research is to examine the degree to which the ability of listeners to exploit interaural envelope cues depends crucially on the nature of the task (tone detection *vs* speech understanding), and/or on the particular physical characteristics of the stimuli.

VI. CONCLUSIONS

(i) Introducing dips in a masker via tapered square-wave amplitude modulation improves performance markedly for unprocessed speech, but this modulation masking release is reduced when the speech is noise-vocoded so as to simulate aspects of the information available to cochlear implant listeners.

(ii) Binaural differences introduced between noise-vocoded speech and a continuous noise masker improved speech recognition performance in a manner expected from the literature on both tonal signals and unprocessed speech. However, these same binaural differences did not help listeners exploit the dips created in the masker when it was modulated. Results are consistent with the idea that, at least for the masker parameters used here, simple spatial differences do not help listeners exploit the dips in a modulated masker in order to identify vocoded speech. Instead, the benefits arise from the processing of interaural decorrelation in the waveform fine structure.

ACKNOWLEDGMENTS

Grants from the Otology Research Fund supported this work. The authors would like to thank three anonymous reviewers, the associate editor Richard Freyman, Jonathan Peelle, Barbara Shinn-Cunningham, and Eric Thompson for helpful feedback on earlier versions of this manuscript, and Christine Mason and Gerald Kidd Jr. for sharing the head-related transfer functions used to simulate reverberation in Experiment 2.

TABLE III. Reverberation times measured in the IAC chamber as a function of center frequency, in one-octave wide bands.

Center frequency (Hz)	Reverberation time (s)
125	0.15
250	0.12
500	0.10
1000	0.09
2000	0.09
4000	0.08
8000	0.08

APPENDIX: EFFECTS OF REVERBERATION IN EXPERIMENT 2

An additional objective of Experiment 2 was to simulate the effects of mild reverberation on modulation release. In some previous experiments, CI users were tested in sound-treated chambers that presumably were not anechoic (Nelson *et al.*, 2003; Fu and Nogaki, 2005). In general, even mild reverberation smears acoustic energy in time and frequency, reducing the depth of modulation in fluctuating maskers, increasing the amount of spectro-temporal overlap between target and masker during the dips of the noise masker and perhaps rendering the modulation cues less useful (Poissant *et al.*, 2006; Lavandier and Culling, 2008). Therefore, a second hypothesis driving Experiment 2 was that reverberation should reduce the amount of release from masking compared to anechoic conditions.

To test the above hypothesis, stimuli were processed so as to simulate reverberation in a medium-sized IAC chamber. For brevity, we refer to these stimuli as "reverberant," even though headphone presentation was always used. Targets in the reverberant condition were similar to the anechoic stimuli in Experiment 2, except that the noise vocoded speech tokens were further processed with head-related transfer functions (HRTFs). The HRTFs were identical to those in the BARE condition used by Kidd *et al.* (2005a). They were measured on an acoustic manikin in a mildly reverberant room (single-walled IAC booth 12′4″ × 13′ × 7′6″; length, width, height; the direct-to-reverberant ratio, averaged across speaker locations, was 6.3 dB; reverberation times listed in Table III), for a sound source at 5 foot distance, in front of the manikin, in the horizontal plane containing the ears. The first 10 ms of these two recordings are dominated by direct energy; the remainder contains mostly reverberant energy. After processing with HRTFs, each resulting reverberant stimulus was scaled such that the RMS of the direct portion of the sound at the right ear (i.e., stimulus convolved with only the first 10 ms of the head-related impulse response) equaled the RMS of the corresponding anechoic stimulus. The maskers were the same as in the anechoic condition.

Results for the reverberation cases are listed in Table IV. Performance for noise-vocoded speech in unmodulated noise was virtually identical for monotic anechoic versus monotic reverberant, and for diotic anechoic versus diotic reverberant conditions. Speech identification in modulated noise was slightly worse in reverberation relative to the anechoic conditions, consistent with the idea that reverberant energy spread into the dips of the modulated noise masker. However, MUDs were statistically identical for anechoic and reverberant conditions, as confirmed by repeated measures ANOVA with main factors of reverberation, monotic/diotic masker configuration, modulation condition and TMR [$(F \times (1,8)=0.066,\ p=0.804$].

Ignoring results from anechoic conditions, repeated measures ANOVA of logit-transformed percent correct scores with within-listener factors of modulation condition, masker configuration, and TMR, showed main effects of all factors. Performance was significantly better in diotic than in monotic noise, and better in the modulated than in the unmodulated noise conditions [$F(1,8)=74.155, 21.785$; $p<0.001$, $p=0.002$, for diotic/monotic and modulation condition, respectively]. In other words, here, reverberant energy from the IAC chamber, did not dramatically affect performance and MUDs did not decrease.

TABLE IV. Performance in reverberation. Mean performance in bold fonts; 95% confidence intervals in brackets.

	Monotic noise				Diotic noise			
TMR	−16 dB	−8 dB	0 dB	8 dB	−16 dB	−8 dB	0 dB	8 dB
				Unmodulated noise				
% correct	3.2 (1.6)	19.3 (4.7)	60.8 (10.1)	73.4 (7.8)	6.54 (2.2)	43.8 (8.4)	70.4 (6.1)	76.9 (6.7)
				Modulated Noise				
% correct	12.5 (2.6)	38.6 (8.4)	56.7 (11.0)	75.2 (8.5)	22.7 (7.6)	47.9 10.8)	68.6 (8.3)	76.2 (8.2)
				MUD				
MUD [logit-units]	1.2 (0.4)	1.0 (0.5)	−0.2 (0.5)	0.2 (0.5)	1.2 (0.7)	0.2 (0.4)	−0.1 (0.5)	0.0 (0.4)

[1] Other explanations for the reduced masking release, observed with degraded speech have also been proposed by these authors. Alternative explanations include a lack of redundancy in the degraded speech signal, impoverished spectral information in that signal, and interference from modulations in the masker envelope with the processing of amplitude modulations in the envelope of the speech signal (Kwon and Turner, 2001; Nelson *et al.*, 2003; Qin and Oxenham, 2003; Stickney *et al.*, 2004; Fu and Nogaki, 2005; Oxenham and Simonson, 2009). In addition, several studies suggest that modulation masking release can only occur at relatively low target to masker energy ratios (Bernstein and Grant, 2009; Oxenham and Simonson, 2009).

[2] Previous studies observed masking release for modulation rates between

10 Hz and 32 Hz for whole sentence speech material. In the current study, we chose a masker modulation of 16 Hz, corresponding to a masker cycle length of 62.5 ms that is shorter than half of the typical duration of the target keywords in this study.

[3] The method differed slightly from that used in Experiment 1, because it is difficult to estimate psychometric functions when asymptotic performance does not reach 100% correct. Theoretically, logit transforming a psychometric function asymptoting to less than 100% correct results in curved rather than linear shape. To quantify the resulting error, we calculated the RMS difference between hypothetical, sigmoidally shaped psychometric functions and their inverse logit transformed minimum least square line fits. Based on this analysis, fitted psychometric functions were slightly steeper than veridical functions around 50% correct performance level and slightly shallower near their asymptote. Importantly, for 77.7% correct asymptotic performance the resulting root mean square error was 3.2%, a precision deemed reasonable for the purpose of this study.

[4] These conditions were interleaved with a fifth condition, which, because of a programming error, is not reported here.

Apoux, F., and Bacon, S. P. (**2008**). “Selectivity of modulation interference for consonant identification in normal-hearing listeners,” J. Acoust. Soc. Am. **123**, 1665–1672.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). “The effect of spatial separation on informational and energetic masking of speech,” J. Acoust. Soc. Am. **112**, 2086–2098.

Assmann, P. F., and Summerfield, Q. (**2004**). “The perception of speech under adverse acoustic conditions,” in *Springer Handbook of Auditory Research: Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, Berlin), Vol. **18**. pp. 231–308.

Bernstein, J. G. W., and Grant, K. W. (**2009**). “Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners,” J. Acoust. Soc. Am. **125**, 3358–3372.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). “A speech corpus for multitalker communications research,” J. Acoust. Soc. Am. **107**, 1065–1066.

Bronkhorst, A. (**2000**). “The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,” Acustica **86**, 117–128.

Brungart, D. S. (**2001**). “Informational and energetic masking effects in the perception of two simultaneous talkers,” J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., and Simpson, B. D. (**2002**). “The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal,” J. Acoust. Soc. Am. **112**, 664–676.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). “Informational and energetic masking effects in the perception of multiple simultaneous talkers,” J. Acoust. Soc. Am. **110**, 2527–2538.

Buss, E., Pillsbury, H. C., Buchman, C. A., Pillsbury, C. H., Clark, M. S., Haynes, D. S., Labadie, R. F., Amberg, S., Roland, P. S., Kruger, P., Novak, M. A., Wirth, J. A., Black, J. M., Peters, R., Lake, J., Wackym, P. A., Firszt, J. B., Wilson, B. S., Lawson, D. T., Schatzer, R., D’Haese, P. S. C., and Barco, A. L. (**2008**). “Multicenter U.S. bilateral MED-EL cochlear implantation study: Speech perception over the first year of use,” Ear Hear. **29**, 20–32.

Buus, S. (**1985**). “Release from masking caused by envelope fluctuations,” J. Acoust. Soc. Am. **78**, 1958–1965.

Carlyon, R. P., Buus, S., and Florentine, M. (**1989**). “Comodulation masking release for three types of modulator as a function of modulation rate,” Hear. Res. **42**, 37–45.

Carlyon, R. P., and Gockel, H. (**2008**). “Effects of harmonicity and regularity on the perception of sound sources,” in *Springer Handbook of Auditory Research: Auditory Perception of Sound Sources*, edited by W. A. Yost (Springer, New York), Vol. **29**. pp. 191–213.

Ciocca, V., and Darwin, C. J. (**1999**). “The integration of nonsimultaneous frequency components into a single virtual pitch,” J. Acoust. Soc. Am. **105**, 2421–2430.

Darwin, C. (**2008**). “Spatial hearing and perceiving sources,” in *Springer Handbook of Auditory Research: Auditory Perception of Sound Sources*, edited by W. A. Yost (Springer, New York), Vol. **29**. pp. 215–232.

Drennan, W. R., Won, J. H., Dasika, V. K., and Rubinstein, J. T. (**2007**). “Effects of temporal fine structure on the lateralization of speech and on speech understanding in noise,” J. Assoc. Res. Otolaryngol. **8**, 373–383.

Edmonds, B. A., and Culling, J. F. (**2005**). “The spatial unmasking of speech: Evidence for within-channel processing of interaural time delay,” J. Acoust. Soc. Am. **117**, 3069–3078.

Edmonds, B. A., and Culling, J. F. (**2006**). “The spatial unmasking of speech: Evidence for better-ear listening,” J. Acoust. Soc. Am. **120**, 1539–1545.

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.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). “Effect of number of masking talkers and auditory priming on informational masking in speech recognition,” J. Acoust. Soc. Am. **115**, 2246–2256.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2008**). “Spatial release from masking with noise-vocoded speech,” J. Acoust. Soc. Am. **124**, 1627–1637.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). “The role of perceived spatial separation in the unmasking of speech,” J. Acoust. Soc. Am. **106**, 3578–3588.

Fu, Q. J., and Nogaki, G. (**2005**). “Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing,” J. Assoc. Res. Otolaryngol. **6**, 19–27.

Gantz, B. J., Tyler, R. S., Rubinstein, J. T., Wolaver, A., Lowder, M., Abbas, P., Brown, C., Hughes, M., and Preece, J. P. (**2002**). “Binaural cochlear implants placed during the same operation,” Otol. Neurotol. **23**, 169–180.

Garadat, S. N., Litovsky, R. Y., Yu, G., and Zeng, F. (**2009**). “Role of binaural hearing in speech intelligibility and spatial release from masking using vocoded speech,” J. Acoust. Soc. Am. **126**, 2522–2535.

Gnansia, D., Jourdes, V., and Lorenzi, C. (**2008**). “Effect of masker modulation depth on speech masking release,” Hear. Res. **239**, 60–68.

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

Howard-Jones, P. A., and Rosen, S. (**1993b**). “Uncomodulated glimpsing in ‘checkerboard’ noise,” J. Acoust. Soc. Am. **93**, 2915–2922.

Ihlefeld, A., and Shinn-Cunningham, B. G. (**2008a**). “Spatial release from masking in a selective speech identification task,” J. Acoust. Soc. Am. **123**, 4369–4379.

Ihlefeld, A., and Shinn-Cunningham, B. G. (**2008b**). “Disentangling the effects of spatial cues on selection and formation of auditory objects,” J. Acoust. Soc. Am. **124**, 2224–2235.

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.

Kidd, G., Jr., Arbogast, T., Mason, C., and Gallun, F. (**2005b**). “The advantage of knowing where to listen,” J. Acoust. Soc. Am. **118**, 3804–3815.

Kidd, G., Jr., Mason, C. R., Brughera, A., and Hartmann, W. M. (**2005a**). “The role of reverberation in release from masking due to spatial separation of sources for speech identification,” Acustica **91**, 526–536.

Kitterick, P. T., and Summerfield, A. Q. (**2007**). “The role of attention in the spatial perception of speech,” Assoc. Res. Otolaryngol. Abstr. **30**, 423.

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.

Lavandier, M., and Culling, J. F. (**2008**). “Speech segregation in rooms: Monaural, binaural, and interacting effects of reverberation on target and interferer,” J. Acoust. Soc. Am. **123**, 2237–2248.

Levitt, H., and Rabiner, L. R. (**1967**). “Binaural release of masking for speech and gain in intelligibility,” J. Acoust. Soc. Am. **42**, 601–608.

Loizou, P. C., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., and Roland, P. (**2009**). “Speech recognition by bilateral cochlear implant users in a cocktail-party setting,” J. Acoust. Soc. Am. **125**, 372–83.

Long, C. J., Carlyon, R. P., Litovsky, R. Y., and Downs, D. H. (**2006**). “Binaural unmasking with bilateral cochlear implants,” J. Assoc. Res. Otolaryngol. **7**, 352–360.

Lutfi, R. A., Kistler, D. J., Callahan, M. R., and Wightman, F. L. (**2003**). “Psychometric functions for informational masking,” J. Acoust. Soc. Am. **114**, 3273–3282.

Miller, G., and Licklider, J. (**1950**). “Sensitivity to changes in the intensity of white Gaussian noise and its relation to masking and loudness,” J. Acoust. Soc. Am. **22**, 167–193.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). “Relative dominance of individual partials in determining the pitch of complex tones,” J. Acoust. Soc. Am. **77**, 1853–1860.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). “Thresholds for hearing mistuned partials as separate tones in harmonic complexes,” J. Acoust. Soc. Am. **80**, 479–483.

Morrison, G. S., and Kondaurova, M. V. (**2009**). “Analysis of categorical response data: Use logistic regression rather than endpoint-difference scores or discriminant analysis," J. Acoust. Soc. Am. **126**, 2159–2162.

Müller, J., Schön, F., and Helms, J. (**2002**). "Speech understanding in quiet and noise in bilateral users of the MEDEL COMBI 40/401 cochlear implant system," Ear Hear. **23**, 198–206.

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.

Poissant, S. F., Whitmal, N. A., III, and Freyman, R. L. (**2006**). "Effects of reverberation and masking on speech intelligibility in cochlear implant simulations," J. Acoust. Soc. Am. **119**, 1606–1615.

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.

Schleich, P., Nopp, P., and D'Haese, P. (**2004**). "Head shadow, squelch and summation effects in bilateral users of the Med-El Combi 40/40 + cochlear implant," Ear Hear. **25**, 197–204.

Shinn-Cunningham, B. G., Ihlefeld, A., Satyavarta, and Larson, E. (**2005**). "Bottom-up and top-down influences on spatial unmasking," Acustica **91**, 967–979.

Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., and Williams, R. M., Jr. (**1949**). "Adjustment during army life," *Studies in Social Psychology in World War II: The American Soldier* (Princeton University Press, Princeton), Vol. **1**. p. 45.

Summers, V., and Molis, M. R. (**2004**). "Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level," J. Speech Lang. Hear. Res. **47**, 245–256.

Tyler, R. S., Gantz, B. J., Rubinstein, J. T., Wilson, B. S., Parkinson, A. J., Wolaver, A., Preece, J. P., Witt, S., and Lowder, M. W. (**2002**). "Three month results with bilateral cochlear implants," Ear Hear. **23**, 80S–89S.

van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," J. Acoust. Soc. Am. **101**, 1671–1680.

van Hoesel, R. J. M., and Tyler, R. S. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.

Whitmal, N. A., III, Poissant, S. F., Freyman, R. L., and Helfer, K. S. (**2007**). "Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience," J. Acoust. Soc. Am. **122**, 2376–2388.

Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance II*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), pp. 255–276.