The effect of spatial separation on informational and energetic masking of speech[a)]

Tanya L. Arbogast,[b)] Christine R. Mason, and Gerald Kidd, Jr.
Department of Communication Disorders and The Hearing Research Center, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

(Received 28 February 2002; revised 13 July 2002; accepted 6 August 2002)

The effect of spatial separation of sources on the masking of a speech signal was investigated for three types of maskers, ranging from energetic to informational. Normal-hearing listeners performed a closed-set speech identification task in the presence of a masker at various signal-to-noise ratios. Stimuli were presented in a quiet sound field. The signal was played from 0° azimuth and a masker was played either from the same location or from 90° to the right. Signals and maskers were derived from sentences that were preprocessed by a modified cochlear-implant simulation program that filtered each sentence into 15 frequency bands, extracted the envelopes from each band, and used these envelopes to modulate pure tones at the center frequencies of the bands. In each trial, the signal was generated by summing together eight randomly selected frequency bands from the preprocessed signal sentence. Three maskers were derived from the preprocessed masker sentences: (1) different-band sentence, which was generated by summing together six randomly selected frequency bands out of the seven bands not present in the signal (resulting in primarily informational masking); (2) different-band noise, which was generated by convolving the different-band sentence with Gaussian noise; and (3) same-band noise, which was generated by summing the same eight bands from the preprocessed masker sentence that were used in the signal sentence and convolving the result with Gaussian noise (resulting in primarily energetic masking). Results revealed that in the different-band sentence masker, the effect of spatial separation averaged 18 dB (at 51% correct), while in the different-band and same-band noise maskers the effect was less than 10 dB. These results suggest that, in these conditions, the advantage due to spatial separation of sources is greater for informational masking than for energetic masking.
[DOI: 10.1121/1.1510141]

PACS numbers: 43.66.Dc, 43.66.Lj, 43.66.Pn [MRL]

I. INTRODUCTION

The ''cocktail party'' (Cherry, 1953) environment contains multiple speech sources and often involves the problem of understanding one talker while ignoring one or more speaking at the same time. This environment is acoustically complex. There are multiple paths of direct sound, one from each source, and in a nonanechoic room, there are also several additional paths of reverberant sound associated with each sound source. The ability to separate simultaneous spectral components into individual sound sources and attend only to the source of interest is both an impressive and essential task for successful communication in the cocktail party environment.

When speech is embedded in background speech, as in the cocktail party, two different types of masking occur—*energetic* and *informational* (Freyman *et al.*, 1999, 2001; Brungart, 2001a; Brungart *et al.*, 2001). Traditional masking theory posits the existence of a bank of overlapping auditory filters through which sound is filtered. Energetic or peripheral masking occurs because the masker energy, or portion thereof, falls within the same auditory filter as the signal energy, resulting in a degraded representation of the signal in the auditory system from the cochlea and beyond. Informational masking occurs in the absence of, or in addition to, overlapping spectra between the signal and masker; there is good representation of the signal in the neural firing patterns beyond the cochlea. Traditionally, informational masking is believed to stem from uncertainty in the stimulus (e.g., Pollack, 1975; Watson *et al.*, 1976; Watson, 1987; Neff and Green, 1987; Lutfi, 1989), although qualitative similarity between the signal and masker may also play a role (e.g., Leek *et al.*, 1991; Neff, 1995; Brungart, 2001a; Kidd *et al.*, 2002). The interpretation is that the uncertainty in the stimulus makes it difficult to detect the signal, or the signal fuses with the masker due to similarity, rendering it difficult to extract information about any one frequency component. Informational masking is thought to operate at a higher level in the auditory system than the periphery. Properties of informational masking have been well-studied using nonspeech stimuli, which affords good control over the type of masking produced. Similar research using speech stimuli is difficult to find, although several studies have attributed certain outcomes to the presence of informational masking in concurrent speech tasks (e.g., Brungart, 2001a; Brungart *et al.*, 2001; Freyman *et al.*, 1999, 2001).

Cherry (1953) identified spatial separation between mul-

[a)]Portions of this research were presented at the Midwinter Meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, FL, January, 2002. This work has also been submitted by the first author in partial fulfillment of the requirements for the ScD degree.
[b)]Electronic mail: rohtla@bu.edu

tiple sound sources as one of several potential means by which listeners might solve the cocktail party problem. A large body of research exists on the benefits of spatial separation between speech and noise, and to a lesser extent between two speech sources. For a normal-hearing listener, a 90° spatial separation between signal speech and the masker (speech or noise) improves speech recognition performance, often by 5 to 10 dB depending on room characteristics, speech materials, etc. (Duquesnoy, 1983; Gelfand *et al.*, 1988; Bronkhorst and Plomp, 1990; Nilsson *et al.*, 1992; Bronkhorst and Plomp, 1992; Peissig and Kollmeier, 1997; Hawley, 2000). Spatial separation between a signal and masker gives rise to several cues that can account for the improved speech recognition: better ear advantage due to headshadow, binaural interaction, and auditory object formation. The latter cue is a function of auditory scene analysis (cf. Bregman, 1990), the organization of simultaneous sound components into auditory objects, which once formed, can be segregated (perceived by the listener as separate objects). A large proportion of the improvement in speech recognition performance reported due to spatial separation is often attributed to headshadow and binaural interaction cues only (Zurek, 1993).

However, a handful of studies have found spatial advantages that cannot be completely accounted for by a combination of headshadow and binaural interaction effects. In a nonanechoic room, Kidd *et al.* (1998) found spatial advantages of 5 to 40 dB for the identification of nonspeech pure-tone frequency patterns in the presence of a multitone, informational masker. By comparison, the spatial advantage in Gaussian noise was 10 dB or less, the majority of which was accounted for by headshadow. In a virtual sound field, Ericson and McKinley (1997) studied the spatial advantage of separating approximately equal-level signal and masker talkers by 90° in the presence of a diotic pink noise of varying levels. The size of the separation advantage was not directly reported, but was estimated from their graphs of diotic and 90° separation psychometric functions at a performance level of 50% correct. The advantage was approximately 14 dB for opposite-sex signal and masker talkers, and about 21 dB for two male talkers. When the two talkers were female, the advantage appeared to be at least as large as for the two male talkers, but because the slopes of the functions were vastly different and the diotic functions did not exceed 45% correct, it was difficult to determine. These advantages, specified in terms of the level of the pink noise, and not the signal-to-masker ratio of the two talkers, are difficult to compare with other research which used a noise and/or a speech masker separately.

When both the signal and masker are speech stimuli, the *perception* of spatial separation between the signal and masker can be sufficient for a significant speech recognition advantage to occur, consistent with auditory object formation. Freyman *et al.* (1999) used localization dominance in the precedence effect to demonstrate this. When listeners were asked to identify nonsense sentences spoken by a female talker in the presence of speech-shaped noise, they showed an average 8-dB spatial release from masking with a 60° separation. In the presence of another female talker, the release was nearly 14 dB. This larger release was attributed to the presence of informational masking in the speech masker, which allowed the listener to use the perceived separation of sources as a cue to source segregation, leading to a larger spatial release than possible with headshadow and binaural interaction cues alone. When the precedence effect was used to create a perceived spatial separation of 60° between the signal and speech-shaped noise, there was no release from masking. However, when the masker was another talker, a significant release of 4 to 9 dB was obtained, even though both stimuli were presented at both locations. Spatial release from the speech masker was 5 to 8 dB greater than that predicted by detection thresholds in the speech-shaped noise for the same location/perceived location configurations. In a subsequent study of perceived spatial separation, Freyman *et al.* (2001) reported a similar spatial release (in the precedence condition) from a two-talker, same-sex masker in English, a foreign language, and a smaller release in reversed English. But, similar to a flat-envelope speech-shaped noise, essentially no spatial advantage was found for speech masked by wideband speech envelope modulated noise, or by 8-band speech envelope modulated noise, both derived from the two-talker masker.

The above studies indicate that spatial release from a speech masker is not necessarily subject to the same rules and limitations as release from a noise masker. Informational masking in the speech masker may be the key to this difference between the maskers. Many previous studies on the spatial release from a speech masker have generally found moderate effects not much larger than that for a noise masker. This may be due to the presence of a significant amount of energetic masking in the task. There may be a larger spatial release from an informational masker as a result of perceptual segregation of signal and masker into separate auditory objects. It is difficult, however, to isolate the informational portion of the masking using speech stimuli since speech is broadband. Two speech sources will have considerable spectral overlap and therefore portions of each speech source will fall within the same auditory filters, resulting in peripheral masking. However, the advantages of using speech stimuli include a less abstract task for the listener and a more direct application to the problem of the cocktail party.

In order to study informational masking using speech stimuli, peripheral masking ideally should be eliminated to remove its influence from the results. Therefore, spectral overlap between the signal and competing speech should be minimal. One approach to this problem is use of cochlear implant simulation processing on the speech stimuli. Cochlear implantees receive a spectrally degraded version of the speech stimulus through their implants. The processing strategy utilized in many implant processors filters speech into several adjacent frequency channels or bandpass filters, extracts the temporal amplitude envelope within each channel, and uses this envelope to modulate the series of electrical pulses sent to each channel. Therefore, the spectral information in speech is binned into a number of channels and is not fully represented. Simulations of cochlear implant processing in normal-hearing listeners have revealed that speech recognition performance is very good (90% correct or better) with as few as 4–6 channels (Shannon *et al.*, 1995; Dorman *et al.*, 1997). However, in order to satisfy the requirement that spectral overlap between signal and masker be minimal, several modifications to the standard method of processing were implemented and are described in detail in the Methods section.

This study was designed to determine: (1) if informational masking can be isolated, at least partially, in speech stimuli and (2) if the size of the spatial release from masking is a function of the type of masking produced and more specifically, if the magnitude of release is different when informational, rather than energetic masking is the primary type of interference present.

II. METHODS

A. Listeners

Listeners were four normal-hearing college students aged 19–23. All had pure-tone thresholds less than or equal to 15 dB HL at octave frequencies from 250–8000 Hz bilaterally. One additional normal-hearing, college-age listener was employed for a brief study of the intelligibility of each of two noise maskers alone, but did not participate in the main study.

B. Stimuli

Speech materials were the CD-recorded sentences of the coordinate response measure (CRM) corpus (Bolia *et al.*, 2000). The corpus contains sentences spoken by eight different talkers—four male and four female. All sentences in the corpus are the same except for three words, which are the callsign (e.g., "Baron," "Charlie," "Ringo"), a color, and a number. Each sentence has the following structure: "Ready [*callsign*] go to [*color*] [*number*] now." Combinations of all eight callsigns, four colors (blue, red, green, and white), and eight numbers (1 through 8) result in 256 different sentences for each talker. The four male talkers of the corpus were used in this study.

The CRM corpus was chosen for several features that make it ideal for this study. One feature is the multiple-choice format that provides a somewhat easier task than open-set materials. The form of the sentences provides three important advantages: (1) linguistic context is not a factor; (2) the same sentence can be presented more than once without concern regarding listener memory for a specific sentence; and (3) the callsign can be used as a marker for the signal sentence. This corpus has been used in several recent studies of speech intelligibility, particularly in relation to multiple talkers, spatial separation, and energetic versus informational masking (Ericson and McKinley, 1997; Brungart, 2001a; Brungart and Simpson, 2001; Brungart *et al.*, 2001; Simpson *et al.*, 1999; Bolia *et al.*, 1999). Some of its basic characteristics and its relation to the articulation index have been reported by Brungart (2001b). Sentences are aligned at the onset of the word "Ready," but speaking rates vary slightly from talker to talker and within talker. Differences in intelligibility of 5 to 20 percentage points have been reported across individual number, color, and talker (Brungart, 2001b, 2001a). However, in the current study these effects were evenly distributed across all conditions and test levels and therefore were not anticipated to have significant impact on the results, except to possibly increase the variability of each individual data point.

Signal sentences were drawn from those with the callsign "Baron," of which 128 were available from the four male talkers. All 128 sentences were used 7 times each to create a pool of 896 processed signal sentences. Each signal sentence was paired with a randomly chosen masker sentence from among those with a different callsign, color, number, and talker than the signal sentence. A different color and number were required to allow for more robust analysis of error/confusion patterns between signal and masker sentences. If color and/or number were the same in the signal and masker, then a correct response could imply that the listener identified either the signal or the masker.

Each sentence was digitally preprocessed in order to produce cochlear implant simulated speech (Shannon *et al.*, 1995), with a total of 15 channels or bands. Software for this processing was provided by House Ear Institute and was modified to use pure-tone rather than narrow-band noise carriers and logarithmically spaced bandpass filters. Dorman *et al.* (1997) showed that speech recognition performance did not differ between narrow-band noise and pure-tone carriers. All sentences were down-sampled from 40 to 20 kHz prior to processing. The preprocessing algorithm included a high-pass Butterworth filter at 1200 Hz with a 6-dB/octave slope, followed by a bank of 15 bandpass filters. The bandpass filters were 1/3 octave, fourth-order Butterworth filters. The center frequencies of the filters were evenly spaced on a log scale from 215 to 4891 Hz with successive center frequencies at a ratio of 1.25. The output of each bandpass filter was half-wave rectified and low-pass filtered at 50 Hz by a fourth-order Butterworth filter to extract the amplitude envelope within each band. The envelope was then used to modulate a pure tone having a starting phase of zero and frequency equal to the center frequency of the corresponding band. The output of preprocessing was a set of 15 speech-envelope-modulated pure tones for each sentence.

The processed signal sentences were generated by summing a set of eight randomly chosen (for each sentence) bands of the 15 available from the signal sentences at the output of the preprocessing. Each signal sentence was paired with one masker sentence, but three different maskers were created from each masker sentence (as described below). Therefore, on each trial the signal sentence was played with one of the three maskers derived from the masker sentence to which it was paired.

1. Different-band sentence masker

The different-band sentence masker was generated by summing six bands of the masker sentence at the output of preprocessing. The six bands were randomly chosen (for each sentence) from the seven bands that were not chosen for the signal sentence with which it was paired. Therefore, the bands comprising a signal/different-band sentence masker pair were mutually exclusive. The top panels of Fig. 1 show

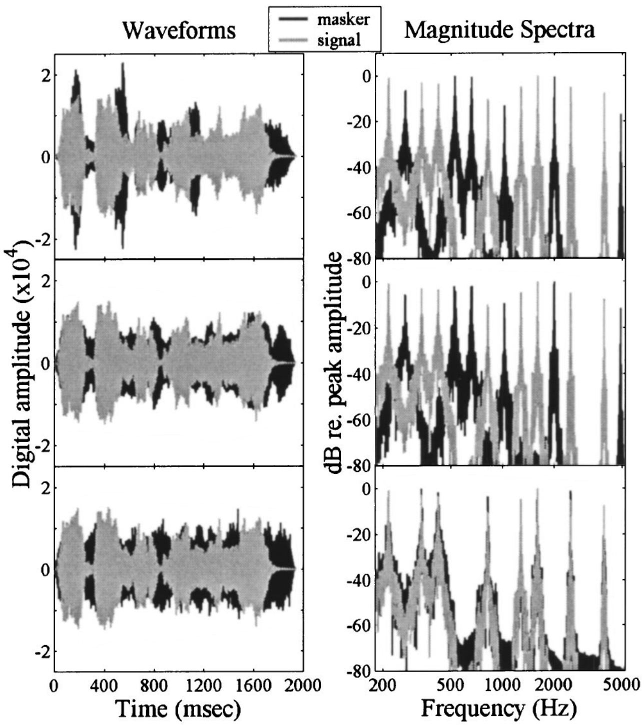

FIG. 1. Time waveforms (left column) and magnitude spectra (right column) for a sample signal sentence and the three types of maskers. The (same) signal sentence is plotted in gray in each panel. The black indicates the masker sentence for the different-band sentence masker (top panels), the different-band noise masker (middle panels), and the same-band noise masker (bottom panels).

the time waveforms and magnitude spectra of a sample pair of signal sentence (gray) and different-band sentence masker (black). The different-band sentence masker was designed to produce primarily informational masking. The magnitude spectra show that the signal and masker peaks did not overlap, minimizing spectral interaction and thus theoretically minimizing energetic masking. Informational masking was theoretically maximized because both signal and masker sentences were intelligible as speech and fundamental frequency information was minimal, making the talkers relatively difficult to discriminate from each other. The harmonic structure (and thus fundamental frequency) of voiced speech was essentially eliminated because the spectrum was represented by sparse, narrow-band, logarithmically spaced frequency bands.

2. Different-band noise masker

The different-band noise masker was derived from the different-band sentence masker described above. The long-term, complex spectrum of each different-band sentence masker was multiplied by the long-term, complex spectrum of a broadband Gaussian noise, and the result was inverse fast Fourier transformed (equivalent to convolving the different-band sentence with Gaussian noise). The length of the resultant waveform was truncated to equal the longer of either the different-band sentence masker it was derived from, or the signal sentence with which it was paired. A rise–fall time of 20 ms was imposed with a cosine^2 ramp. The spectra of this masker and the different-band sentence from which it was derived were nearly identical, therefore, they contained approximately the same (minimal) amount of energetic masking. Actual amounts of energetic masking likely varied slightly because the sentence masker contained more amplitude modulation than the different-band noise masker. The different-band noise had no intelligibility as speech and was qualitatively different from the signal sentence. Subjectively, this masker sounded like an amplitude-modulated multitone complex. Therefore, the proportion of informational masking was theoretically low relative to the different-band sentence masker, and energetic masking was approximately the same as for the different-band sentence masker. This masker was included as a control for energetic masking in the different-band sentence masker. The time waveforms and magnitude spectra of a sample different-band noise and signal sentence pair (same signal sentence as in the top panels) are shown in the middle panels of Fig. 1.

3. Same-band noise masker

The same-band noise masker was generated by summing eight bands of the preprocessed masker sentence. The eight bands were the same eight bands that comprised the signal sentence. The long-term, complex spectrum of the result was multiplied by the complex spectrum of a broadband Gaussian noise and inverse fast Fourier transformed (equivalent to convolving the eight-band masker sentence with Gaussian noise). The length of the waveform was truncated to equal the longer of either the processed signal sentence to which it was paired or the different-band masker sentence. A rise–fall time of 20 ms was added with a cosine^2 ramp. The time waveforms and magnitude spectra of a sample same-band noise and signal sentence pair (same signal sentence as in the top/middle panels) are shown in the lower panels of Fig. 1. The magnitude spectra almost completely overlapped, thus maximizing energetic masking. In addition, informational masking was theoretically minimized because the masker had no speech intelligibility and, as with the different-band noise, was not qualitatively similar to the signal sentence. This masker condition was included to contrast the effect of spatial separation for a predominantly informational (different-band sentence) and predominantly energetic (same-band noise) masker.

C. Stimulus presentation

All stimuli were presented in a sound field, located within a single-walled IAC sound booth with the typical metal, perforated surfaces except for the carpeted floor. Two Acoustic Research 215PS speakers were used to present the stimuli, one at 0° azimuth and the other at 90° azimuth to the right of the listener in the horizontal plane. The distance between the listener and the speakers was 5 ft. and the speaker height was approximately ear level for a seated listener. Stimuli were played through Tucker-Davis Technologies (TDT) hardware. The signal and masker were played out through separate TDT system channels. Stimuli were converted at a rate of 50 kHz via a 16-bit, 8-channel D/A converter (DA3-8), low-pass filtered with a 20-kHz cutoff (FT-6), and attenuated via programmable attenuators (PA-4). The signal was sent through a switch (SS-1) to route it to the desired signal speaker. The signal and masker were sent to mixers where they were summed if they were both presented to the 0° speaker location. The stimuli were passed through power amplifiers (Tascam) and then to one or two speakers in the sound field, depending on the experimental condition. The masker level was fixed at 60 dB SPL.

Inverse filters were created for each speaker to compensate for the frequency response. The signal sentences were convolved with the inverse filter of the speaker at 0°. The maskers were convolved with the inverse filter of each speaker (0° and 90°), resulting in two versions of each masker, one appropriate to play from each speaker depending on the spatial separation condition. All processed stimuli were subsequently up-sampled to 50 kHz, scaled to equal rms amplitude, and saved to disk for later use. The difference in overall speaker output was equalized by adjusting the amplifier gain for each channel.

D. Procedures

Identification thresholds were estimated adaptively for the signal alone. Psychometric functions were measured for the signal-plus-masker conditions. Both signal-only and signal-plus-masker conditions were single interval forced-choice tasks with an identification response consisting of a color (four choices) and a number (eight choices). Each response was considered correct only if both the color and number of the signal sentence were identified accurately. Each trial proceeded as follows: warning, stimulus, response interval (color followed by number), and feedback. The listener indicated color response by use of four keys labeled with the written word in a background of the color assigned to it. The number response was by use of separate number keys (1–8). At the end of each block, the listener received threshold or percent-correct feedback for the block just completed.

The tracking rule used to measure signal-only identification thresholds caused a one-step decrease in signal level following a correct response, and a one-step increase in signal level following an incorrect response. This tracking rule targets the 50% correct performance level (Levitt, 1971). The initial 6-dB step size was reduced to 4 dB after the first three reversals. Threshold was taken as the average of a minimum of six reversals, not including the first four or five depending on the total number of reversals. Each block contained 36 trials. The signal sentence was played at the 0° location at 60 dB SPL. The threshold estimates from two blocks of trials were averaged to obtain signal-only identification threshold.

Signal-plus-masker psychometric functions were measured using four to eight signal levels, depending on the condition and the listener's early results. The levels were chosen individually to span percent-correct performance from near chance to near perfect. Signal levels were mixed randomly with replacement within every 50-trial block. Each of the three types of maskers was tested in two spatial separation conditions, 0° (signal and masker at 0°) and 90° (signal at 0° and masker at +90°). Each block of trials contained only one spatial separation condition and one masker type. Blocks of each spatial separation condition were alternated to disperse any learning effects equally across the two separation conditions. A minimum of 100 trials per signal level was obtained for each spatial separation and masker type condition.

Two blocks of signal-only identification trials were collected first. For three of the four listeners, data for the different-band sentence masker were collected next, followed by the different-band and same-band noise maskers intermixed within the same session. The fourth listener generated data with each of the three maskers mixed within all the sessions. Data were collected within six sessions of 2 h each, including several rest breaks. The listeners' heads were not physically restrained, but they were instructed to keep their heads straight, facing the 0° location throughout stimulus presentation.

Additional data were briefly collected to assess the intelligibility of each masker in isolation. The listener's task was to identify the color and number of the masker sentence. One of the four listeners was tested with one 50-trial block of the different-band sentence masker at 60 dB SPL from the 0° location. One additional normal-hearing listener (not one of the original four) performed the identification task for four 60-trial blocks, two for the different-band noise masker and two for the same-band noise masker, presented at 60 dB SPL at the 0° location.

Listener training was minimal. Each listener initially received a 50-trial block of identification practice. The signal sentences were presented without a masker at a fixed level of 60 dB SPL from the 0° speaker. If performance was at least 90% correct, the listener was given four 50-trial practice blocks with the signal presented in the different-band sentence masker, two blocks for each separation condition. One fixed signal level was used in each block. The signal-to-masker ratio was +15 dB for the first block, but was decreased to +10 dB for the second practice block. These data were not included in any analyses or plots.

E. Data analyses

Psychometric functions were constructed showing percent-correct identifications as a function of signal level. The psychometric functions were fit by the Nelder–Mead simplex method (FMINS function in MATLAB) with a logistic of the form

$$p(c) = \alpha + ((1.0-\alpha)/(1.0+e^{-k(x-m)})), \tag{1}$$

where $\alpha = 0.03$ (chance performance for 1 out of 32), k is the slope of the function, m is the level of the midpoint (about 51.5% correct), and x is the signal sentence level.

All statistical analyses were performed with repeated-measures analysis of variance (ANOVA) with listener as the error term. *Post hoc* testing was performed with Tukey's Studentized Range (HSD) Test, which controls for type I errors, at the $\alpha = 0.05$ level of significance. Two difference measures were calculated: Δm and Δk. These values were calculated from the midpoint and slope parameters of the fitted logistic functions described above, and represent the difference in spatial separation condition (0°–90°), not masker type.

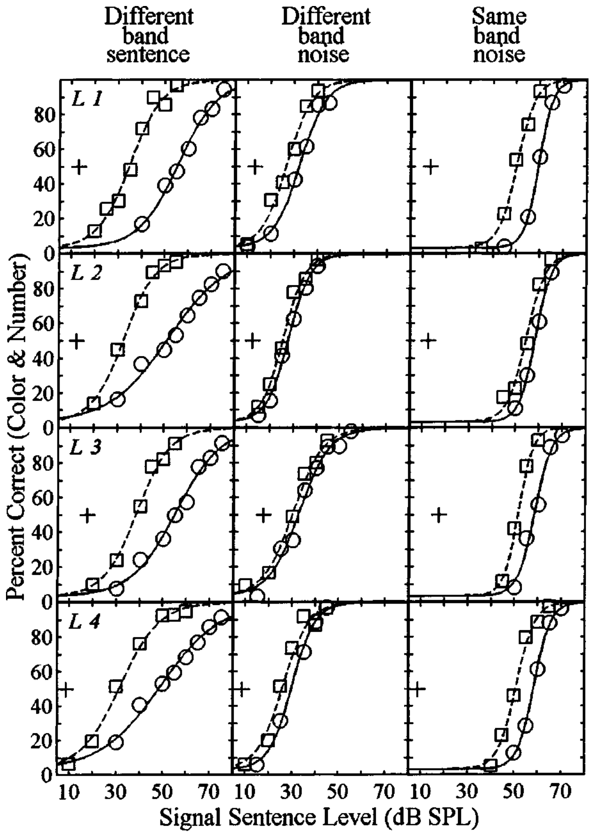

FIG. 2. Psychometric functions for each listener (rows L1 to L4) and each masker type (columns). The plus symbol marks signal-only adaptive identification threshold for 50% correct (plotted three times for each listener as a reference). The circles are the data points and the solid line is the logistic fit for 0° separation, and the squares are the data points and the dashed line is the logistic fit for 90° separation.

III. RESULTS

The signal sentences by themselves (without a masker) were very intelligible at a normal conversational level (60 dB SPL). All listeners achieved 96% to 98% correct for the initial 50-trial practice block, and all reported that the task was easy to perform. Signal-only adaptive identification thresholds ranged from 8.7 to 17.7 dB SPL. The intelligibility of the different-band sentence masker by itself was also very high (98% correct), while the different-band and same-band noise maskers were each identified near chance (3%)—4.2% for the different-band noise and 3.3% for the same-band noise.

The signal-plus-masker data and fitted logistic functions are plotted in Fig. 2. Each panel contains the data for one listener and one masker type. Percent correct for identifying both color and number correctly in the same trial is plotted as a function of signal sentence level. Zero dB signal-to-masker ratio corresponds to a signal level of 60 dB SPL. Each row displays the data for one listener (denoted L1 to L4). The different-band sentence masker results are located in the first column, the different-band noise in the middle column, and the same-band noise in the last column. The circles represent 0° and squares represent 90° separation. The plus symbol indicates the signal-only 50%-correct mean threshold from the adaptive procedure and is plotted three times as a reference, once for each masker type and listener. The fitted logistic functions are indicated by the solid line for 0° and by the dashed line for 90°.

The first important feature of the data is the consistent monotonicity of the psychometric functions across the four listeners and three maskers used in the experiment. In all three masker types, the function for 90° lies to the left of that for 0° along the abscissa, and more importantly, the magnitude of the shift changes as a function of masking condition. The location and slope of the psychometric functions also change with masker type independent of spatial separation. Table I contains m (midpoint) and k (slope) values and an estimate of the goodness of fit (proportion of variance accounted for) from the fitted psychometric functions, as well as the difference measures Δm and Δk, for each listener and condition. The midpoint parameter was used as a summary measure of the effect each masker had on performance and for comparison between spatial conditions. The midpoint was chosen because it is not influenced by ceiling or floor effects, and is readily available from the logistic fits. However, it should be kept in mind that the midpoint may not fully describe the effects, particularly when the function slopes are not parallel, in which case the size of the effect depends on the performance level.

Figure 3 plots the mean m (top panel) and k (bottom panel) values from Table I as a function of masker type with spatial separation as the parameter. The circles are for 0° and the squares for 90° separation. The dashed line at the bottom of the top panel indicates the mean signal-only identification threshold in dB SPL.

A. Effect of masker type

For 0° spatial separation, the mean m value was highest for the same-band noise at 58.7 dB, followed closely by 52.9 dB for the different-band sentence, while the different-band noise resulted in a lower value of 30.8 dB. An ANOVA confirmed a significant main effect of masker type $[F(2,6) = 311.6,\ p < 0.0001]$, and *post hoc* testing revealed that the same-band noise value was significantly higher than the different-band sentence masker, and both were higher than the different-band noise midpoint value. Midpoint values in the 90° separation condition followed a similar pattern with a mean of 51.8 dB for the same-band noise, 34.5 dB for the different-band sentence, and 27.2 dB for the different-band noise. The ANOVA on the 90° midpoint values revealed a significant main effect of masker $[F(2,6) = 256.7,\ p < 0.0001]$ and *post hoc* testing showed that each midpoint value was significantly different from the others.

The mean slope values for both the 0° and 90° separation conditions increased from approximately 0.1 for the different-band sentence masker to nearly 0.3 for the same-band noise, with the different-band noise values intermediate. An ANOVA on the slope values revealed a significant main effect of masker type in both the 0° $[F(2,6) = 67.03,\ p < 0.0001]$ and 90° conditions $[F(2,6) = 39.03,\ p < 0.001]$. *Post hoc* testing revealed that in the 0° condition, all three slopes were significantly different from each other. In the 90° condition, the slope for the same-band noise was significantly steeper than for the other two maskers, but the

TABLE I. Results of the logistic fits applied to the data for each masker (subtables A, B, and C). The *m*- and *k*-values and amount of variance accounted for (var) by the fit (in percent) for each listener, as well as means and standard deviations across listener are displayed in each subtable for each spatial separation condition. The final two columns of each table give the Δm and Δk difference values between the 0° and 90° separation conditions, in addition to the means and standard deviations across listener.

	0° *separation*			90° *separation*			0°–90°	
Listener	*m*	*k*	Var	*m*	*k*	Var	Δm	Δk
				(A) Different-band sentence masker				
1	55.58	0.1195	98.9	34.91	0.1462	98.0	20.67	−0.027
2	51.72	0.0799	97.9	32.70	0.1487	99.4	19.02	−0.069
3	55.31	0.1064	98.6	38.54	0.1436	99.3	16.77	−0.037
4	48.98	0.0784	98.4	31.71	0.1203	99.3	17.27	−0.042
Average	**52.90**	**0.0961**	⋯	**34.47**	**0.1397**	⋯	**18.43**	**−0.044**
(s.d.)	**(3.15)**	**(0.02)**		**(3.03)**	**(0.01)**		**(1.78)**	**(0.02)**
				(B) Different-band noise masker				
1	32.42	0.1772	99.3	26.83	0.1849	98.5	5.59	−0.008
2	27.93	0.2150	99.4	25.68	0.2151	99.3	2.25	−0.0001
3	32.81	0.1550	98.3	30.24	0.1586	99.2	2.57	−0.004
4	30.01	0.2044	99.7	26.12	0.1824	97.3	3.89	0.022
Average	**30.79**	**0.188**	⋯	**27.22**	**0.185**	⋯	**3.58**	**0.003**
(s.d.)	**(2.28)**	**(0.03)**		**(2.07)**	**(0.02)**		**(1.52)**	**(0.01)**
				(C) Same-band noise masker				
1	59.53	0.3279	100	50.30	0.2560	99.3	9.23	0.072
2	58.35	0.2967	99.8	54.53	0.2479	97.8	3.82	0.049
3	58.49	0.2777	98.3	51.52	0.3223	99.7	6.97	−0.045
4	58.39	0.2839	99.9	50.82	0.2533	99.3	7.57	0.031
Average	**58.69**	**0.2966**	⋯	**51.79**	**0.2699**	⋯	**6.90**	**0.027**
(s.d.)	**(0.56)**	**(0.02)**		**(1.90)**	**(0.04)**		**(2.26)**	**(0.05)**

different-band sentence slope at 0.14 was not different from the 0.19 slope for the different-band noise.

B. Effect of spatial separation

For the different-band sentence masker, the mean difference in *m* value (0°–90°), or Δm, is much larger at 18.4 dB than for the other maskers. The different-band noise masker produced a mean Δm of 3.6 dB and the same-band noise masker resulted in a slightly larger Δm of 6.9 dB. All three mean Δm's are statistically larger than zero on paired-sample *t*-tests [$t(3)=20.8$, $p<0.001$ for different-band sentence; $t(3)=4.7$, $p<0.05$ for different-band noise; and $t(3)=6.1$, $p<0.01$ for same-band noise]. An ANOVA revealed a significant main effect of masker type on Δm [$F(2,6)=143.1$, $p<0.001$]. *Post hoc* testing revealed that the mean Δm for each masker was significantly different from all the other maskers.

The effect of masker type on the difference in slope, Δk (0°–90°), was less obvious because it was very small. The different-band sentence masker Δk was consistently negative across listener and had a mean of about −0.04 (the 0° functions were shallower than the 90° functions). This slope change was significantly larger than zero on a paired-sample *t*-test [$t(3)=-4.9$, $p<0.05$]. The average Δk for the different-band noise masker was essentially zero [$t(3)=0.4$, $p=0.716$] and for the same-band noise masker it was less than 0.03 and not significant [$t(3)=1.1$, $p=0.368$], both on paired-sample *t*-tests. An ANOVA on the Δk's with masker type as the factor approached significance [$F(2,6)=4.85$, $p=0.056$].

C. Error analysis

Examination of the identification errors revealed some striking patterns across masker and spatial separation conditions. Figure 4 plots the mean percentage of incorrect responses that matched the color or number from the masker sentence as a function of condition.

Percentages were calculated separately for color and number by dividing the number of responses corresponding to the masker, by the total number of incorrect responses for a limited subset of signal levels. In order to obtain a clearer picture of error patterns, trials presented with signal levels significantly above or below the level required for 50% correct were eliminated. Few errors would occur at levels significantly above 50% correct. For signal levels significantly below that for 50% correct in the different-band sentence masker, the listener clearly heard the masker sentence but the signal sentence was barely intelligible, if at all. Listeners were aware that the signal sentence could not contain the same color and number as the masker sentence, and they reported that at very low signal-to-masker ratios they purposely chose a random color and number other than that heard clearly from the masker sentence. Therefore, it was primarily the signal-to-masker ratios near that required for 50% correct performance that were of most interest for examination of confusions between signal and masker. For each masker/spatial separation condition, trials presented with signal levels equal to or within $\pm j$ dB of the midpoint of the fitted psychometric function, *m*, were included in the error calculations, with $j=10$ for the different-band sentence

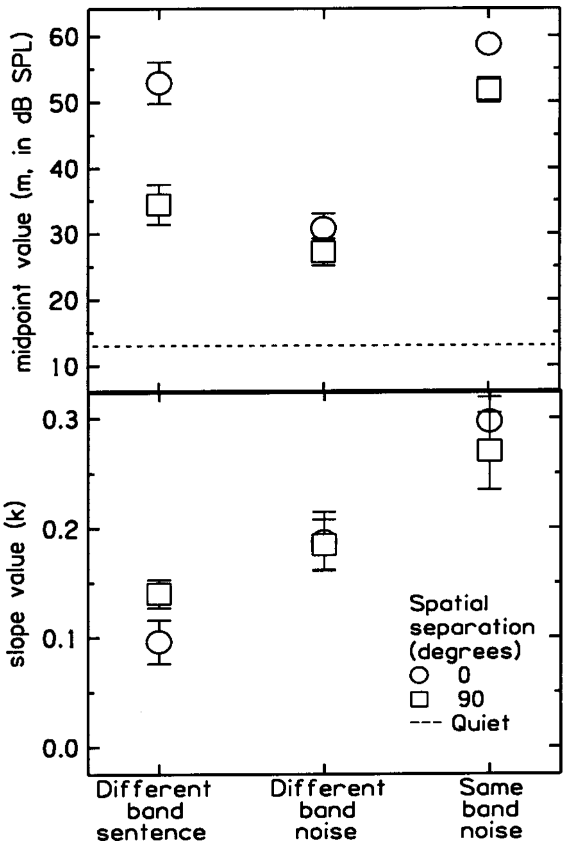

FIG. 3. Midpoint (top panel) and slope (lower panel) values from the fitted logistic curves as a function of masker type, with spatial separation as the parameter. Data are plotted as the means across four listeners and the errors bars are ± 1 standard deviation. The circles represent 0° separation and the squares are for 90° separation. The dashed line at the bottom of the top panel indicates the mean signal-only sentence identification threshold.

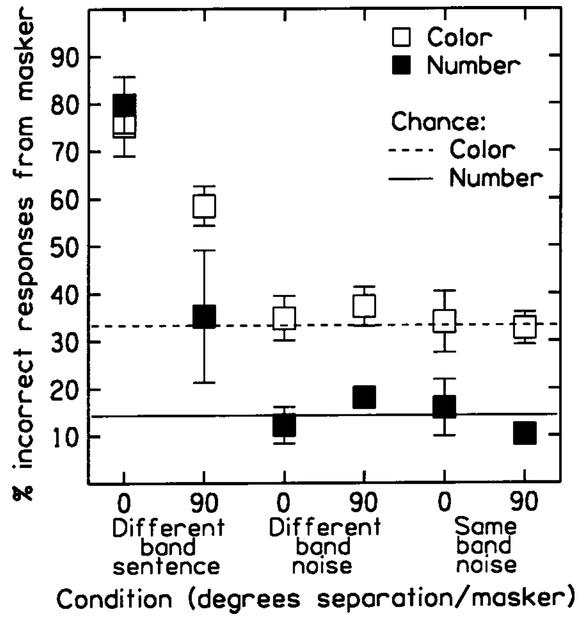

FIG. 4. Percent of incorrect color and number responses that matched color and number in the masker sentence. Results are for a limited range of signal sentence levels near the 50%-correct point on the psychometric functions (see the text for more detailed explanation), and are plotted as a function of spatial separation and masker type. The data are means across the four listeners and the errors bars are ± 1 standard deviation. The open squares represent color response errors and the closed squares are number response errors. Chance is indicated for color by the dashed line (33.3%) and for number by the solid line (14.3%).

masker and $j=5$ for the two noise maskers due to the steeper functions.

Each point in Fig. 4 represents the mean of the four listeners and includes at least 360 incorrect trials. The solid squares are number errors and the solid line indicates expected performance for random guessing or a random distribution of errors for number (14.3%). The open squares are for color errors and the dashed line represents expected performance for a random distribution of color errors (33.3%). Errors for the different-band sentence masker were greater than that expected for randomly distributed responses. For 0° separation, 75% (color) to 80% (number) of errors were from the masker sentence. In the 90° configuration, these numbers dropped to 35% for number and 58% for color, but both remained above their respective random distribution levels. ANOVAs were performed separately for color and number errors in the different-band sentence masker, with spatial separation as the factor. The results confirmed a significant main effect of spatial separation on both color [$F(1,3) = 163.0$, $p < 0.001$] and number errors [$F(1,3) = 59.72$, $p < 0.01$]. In the remaining masker/spatial separation conditions, percentages approximated that expected for a random distribution of responses.

IV. DISCUSSION

The psychometric functions obtained in all three maskers were monotonic. Psychometric functions for a speech signal masked by a single talker have been known to contain nonmonotonicities (plateaus/small dips between -10- and 0-dB signal-to-masker ratio), particularly when the characteristics of the two talkers are similar (Egan *et al.*, 1954; Dirks and Bower, 1969; Brungart, 2001a; Freyman *et al.*, 1999; Brungart *et al.*, 2001). The presumed origin of these nonmonotonicities is segregation by level. It is difficult to segregate two similar talkers when the signal-to-masker ratio is near 0 dB. However, when the signal level drops just below that of the masker, the listener can segregate by level and extract or assign information from/to the softer talker. There is no evidence in the current data to suggest nonmonotonic functions. The level cue is probably a weak cue used only when all else fails to segregate concurrent talkers. Here, the stronger cue may have been timbre/pitch differences between the two talkers arising from the particular set of bands chosen during processing for the signal and masker. However, the timbre/pitch cue varied from trial to trial because the bands were picked randomly on each trial. Therefore, although this cue was possibly more salient than level differences, it was probably not stable enough to be a very strong cue for segregating the signal and masker sentences.

A. Masker type—Relative proportions of informational and energetic masking

The influence of masker type, independent of spatial separation, is interesting to consider with a few caveats. The different-band sentence masker had greater amplitude modulations than the two noise maskers. The larger amplitude peaks in the different-band sentence masker may have contributed to greater nonsimultaneous masking than either of the noise maskers. However, the fact that, in general, amplitude-modulated noise actually results in less masking than an equal-rms flat-amplitude noise (e.g., Festen and Plomp, 1990) suggests that forward and backward masking do not produce more total energetic masking in an amplitude-modulated noise than in a flat-amplitude noise. Based on this argument, the different-band sentence masker possibly contained slightly less energetic masking than the different-band noise. In addition, the fact that speech is dominated by slower envelope fluctuations (and therefore longer dips) relative to a noise, also suggests that the different-band sentence masker produced slightly less energetic masking than the different-band noise masker. Therefore, the different-band noise masker was a conservative control for energetic masking in the different-band sentence masker.

A second caveat is that the midpoints of the psychometric functions were used for comparison between maskers. Due to the different slopes found across masker type, the size of the difference between maskers will change depending on the performance level one looks at. However, the main comparison made below is between the different-band sentence masker and the different-band noise. The psychometric functions in these maskers were significantly separated, in the same direction, at all performance levels above chance, but it should be noted that the difference between these maskers actually increased with performance level.

With full knowledge of the above, the effect of masker type on the 0° separation psychometric functions leads to interpretations regarding the type of masking that each causes. Of particular interest is the difference in midpoint values between the different-band sentence and different-band noise masker functions—about 22 dB. The different-band noise was intended to provide an approximate measure of how effectively the construction of both different-band maskers (sentence and noise) minimized energetic masking. These two maskers contained energy that was spectrally quasi-interleaved with the signal energy to the same extent. Therefore, they should have produced about the same amount of energetic masking based on their spectra, but the design of the stimuli emphasized informational masking in the different-band sentence masker and minimized informational masking in the different-band noise masker. The different-band noise resulted in a mean of 18 dB of (primarily) energetic masking, and the different-band sentence masker likely resulted in about the same amount of energetic masking. However, the different-band sentence masker produced a mean of 40-dB overall masking. This difference is too large to be accounted for by the temporal differences between the different-band sentence and different-band noise maskers. Therefore, some other type of masking contributed to the total amount of interference present in the different-band sentence masker.

Three factors suggest that this other type of masking was informational: (1) the design of the stimuli; (2) the psychometric function slopes; and (3) the error analysis. The first factor was discussed in the Methods section. The different-band sentence masker was designed to maximize informational masking of the signal sentence. The processing of the stimuli resulted in nonoverlapping spectral peaks that reduced energetic masking while preserving the intelligibility and qualitative similarity of the two sentences. The second factor, psychometric function slopes, provides further evidence of masking type. In studies comparing energetic and informational masking, psychometric functions in noise maskers tend to be fairly steep, while informational maskers usually result in relatively shallow psychometric functions (Kidd *et al.*, 1995, 1998, 2002). Psychometric functions for speech masked by a single talker are shallower than those masked by a noise (Dirks and Bower, 1969; Festen and Plomp, 1990, Freyman *et al.*, 1999; Wilson *et al.*, 1990), which might be explained by the larger proportion of informational masking present when speech is masked by a speech masker than a noise masker (Freyman *et al.*, 1999, 2001; Brungart, 2001a; Brungart *et al.*, 2001). At 0° separation, the different-band sentence masker functions had a mean slope of 0.10 (2.3% per dB[1]), the different-band noise 0.19 (4.5% per dB) and the same-band noise 0.30 (7.1% per dB); this is a factor of 2 increase in slope from the different-band sentence to the different-band noise, and a factor of 3 increase from the different-band sentence to the same-band noise.

The reason for shallower psychometric functions in the presence of an effective informational masker is not entirely clear, although several possibilities seem reasonable to consider. Psychometric functions obtained for a predominantly energetic masker are inherently strongly dependent upon signal-to-masker ratio. However, performance in a predominantly informational masker may depend much less on signal-to-masker ratio because level difference between signal and masker is likely a weak segregation cue. Other factors such as signal-masker similarity and/or auditory scene analysis cues (other than level) might have more influence in this circumstance.

Psychometric function slopes probably cannot be used in isolation to determine masking type (informational/energetic). Shallow slopes may also be accounted for by inattention or inappropriately focused attention on some proportion of trials (e.g., Allen and Wightman, 1994). Performance at one signal-to-masker ratio might result from an average of trials for which attention was appropriately focused and trials for which it was not. For the different-band sentence masker it is likely that attention was incorrectly focused on the masker sentence during some proportion of the trials (this is supported by the error analysis discussed below). This is less likely to have occurred for the noise maskers because they were perceptually dissimilar from the signal. However, a similar averaging can be imagined without invoking inattention. When maskers or signals are not homogeneous in terms of effectiveness or recognition, the psychometric function produced by a mixture of these stimuli is shallower than the individual stimulus psychometric functions because the overall function is an average of the individual functions (e.g., Wilson and Carter, 2001). Some combinations of signal and different-band sentence masker were more easily segregated than others (due to spectral and/or temporal differences). For the different-band noise masker, variations in masker effectiveness arose because of differences in the spread of excitation for various combinations of frequency bands. The slopes of the psychometric functions for the different-band sentence masker may be shallower because they were influenced by all these factors (weak dependence on level, misfocused attention, and stimulus nonhomogeneity), whereas the different-band noise masker may have only been susceptible to the influences of nonhomogeneity. Other factors including the design of the stimulus, the error analysis (see below), and the size of the spatial release from masking (e.g., Freyman *et al.*, 1999; Kidd *et al.*, 1998) imply that informational masking was dominant for the different-band sentence masker and energetic masking was dominant for the different-band noise (although less energetic than the same-band noise masker).

The error analysis is the third factor that supports a high informational to energetic masking ratio in the different-band sentence masker. This analysis is similar to that used by Brungart (2001a) in which it was argued that if incorrect responses were due to primarily energetic masking (i.e., the signal was inaudible), the distribution of responses on incorrect trials should be random among the remaining alternatives. When the signal was inaudible, there was no particular reason (and in fact it would have been detrimental) to pick the masker color/number combination because it clearly belonged to the masker. In addition, in the current study, the listeners were aware that a masker–signal pair always contained a different color–number set. Therefore, a significant proportion of masker responses indicates difficulty correctly labeling each sentence as ''signal'' or ''masker,'' which implies that the listener confused the masker and signal and not that the masker necessarily degraded the representation of the signal in the auditory system. Brungart (2001a) reported that the percentage of errors from the masker sentence was significantly larger than expected from random error distribution for a CRM sentence masked by another CRM sentence (both unprocessed), particularly when the masker was the same talker or a same-sex talker. In the current study, listeners responded with the masker color and number no more often than was expected by chance (33.3% for color; 14.3% for number) for both of the noise maskers. In the different-band sentence masker, the proportion of responses that corresponded to the masker was much larger at 75%–80% for the 0° separation and in the 90° separation condition, 35% for number and 58% for color. This suggests that when the signal-to-masker ratio was near the 50%-correct point, the listeners often confused the signal and masker, implying the presence of informational masking.

B. Spatial separation

The different-band sentence masker resulted in the largest spatial release from masking at 18.4 dB. The different-band noise release was 3.6 dB and the same-band noise release was 6.9 dB. The nearly 7-dB release obtained for the same-band noise is consistent with that expected for 90° separation of speech and noise, based on binaural interaction and headshadow cues (Zurek, 1993), and is similar to that reported for 90° separation in other studies of speech masked by noise (e.g., Bronkhorst and Plomp, 1990; Duquesnoy, 1983; Peissig and Kollmeier, 1997; Hawley, 2000; Bronkhorst and Plomp, 1992; Nilsson *et al.*, 1992). The smaller release that occurred for the different-band noise was presumably due primarily to a release (via binaural interaction and headshadow) from the small amount of energetic masking that was present. It was smaller than the release for the same-band noise because the effective level of the noise was very low within auditory filters containing the signal, and binaural interaction is less effective at low noise levels (e.g., Hirsh, 1948; Dolan and Robinson, 1967; Dolan, 1968; McFadden, 1968). An alternative measure of spatial release is the difference in identification performance at one signal level. When calculated with the midpoint of the 0° function as the reference, the spatial advantage is 17, 37, and 43 percentage points for the different-band noise, same-band noise, and different-band sentence, respectively. The difference in spatial advantage between the different-band sentence and same-band noise becomes less striking, but the overall trend remains the same. Spatial release could have also been calculated at a different performance level, in which case the exact magnitude of the advantage in each masker would change, but again, the same trend would be apparent.

The 18.4-dB release obtained in the different-band sentence masker is much larger than can be explained by both headshadow and binaural interaction cues. The size of the release is also a bit larger than found in much of the previous speech-on-speech masking research. Freyman *et al.* (1999) found a 14-dB advantage in an anechoic chamber for a spatial separation of 60° between the signal sentence and a single, same-sex speech masker. In real and virtual rooms, Hawley (2000) found spatial advantages between 5 and 11 dB for one to two same-sex maskers separated from the signal by 60° to 90°. However, the content of the masker sentence(s) was held constant and was well-known to the listener. Therefore, stimulus uncertainty was not high and the informational portion of masking was probably relatively low. Both Duquesnoy (1983) and Peissig and Kollmeier (1997) found spatial advantages between 5 and 7 dB for a 90° separation between signal and speech masker. However, both studies employed stimuli or procedures that likely reduced stimulus uncertainty or signal–masker similarity, therefore limiting informational masking (Duquesnoy used a different-sex masker, and Peissig and Kollmeier used only two signal sentences in continuous discourse masking speech). Comparisons to the current study are also complicated by the fact that the studies above tested some unknown combination of energetic and informational masking of speech. The 18-dB advantage found here was also in the presence of both energetic and informational masking, but the ratio of informational to energetic masking was probably substantially larger than in the above studies. Ericson and McKinley (1997) found large advantages in a same-sex masker that are comparable or larger than those found for the predominantly informational masker in the current study. However, as discussed in the Introduction, direct comparison to the current results is difficult due to significant methodological and procedural differences. Nonetheless, the functions were striking in terms of the size of the spatial advantage and perhaps provide some further support for the existence of substantial advantages due to perceptual factors.

The large spatial release found in the current study may be a consequence of the greater isolation of informational masking due to signal–masker similarity and processing of the signal and masker sentences into mutually exclusive frequency bands. Informational masking may be subject to a significantly large spatial release as a result of perceptual segregation of the two sources, as suggested by Freyman *et al.* (1999) and Freyman *et al.* (2001). They demonstrated that the presence of a perceptual separation between signal and masker sentences produced a 4–10-dB advantage for 60° of separation. This may seem small compared to the current study, but it did not include binaural interaction and headshadow cues, and was the result of an informational to energetic masking ratio which was probably smaller than that created here. The 18-dB advantage found in this study is comparable to the advantages found by Kidd *et al.* (1998) using nonspeech signal patterns in a multitone informational masker. Informational masking was maximized relative to energetic masking, similar to the current study.

The effect here is believed to be mainly due to perceptual factors, not acoustic. The small contribution due to acoustic factors can be estimated by the 3.6-dB advantage found for the different-band noise masker; the remainder of the effect is presumably due to perceptual segregation by spatial location. When the signal is played with a predominantly informational masker, this perceptual effect can improve signal recognition. For 0° separation, the necessary information about the signal is present, it is audible, but the listener cannot separate it from the informational portion of the masker because there are not many salient differences between the signal and masker. When the masker is moved away from the signal, the perceptual effect of spatial separation enables the listener to disentangle the signal and masker, or to correctly assign "signal" and "masker" to the appropriate sentences. The perception of the signal and masker at two separate locations effectively reduces the similarity and confusion between signal and masker. A primarily energetic masker does not benefit as much from this same effect because it is mostly affected by acoustic, not perceptual factors. The signal and masker sentences were perceptually segregated spatially, but probably by other factors as well, such as small differences in speaking rate, continuity of envelope variations over time, and timbre/pitch differences created by using mutually exclusive frequency bands within each signal–masker pair. These additional factors were also present in the 0° separation condition, but the spatial separation may have helped make these additional cues more salient.

Masker type influenced the change in psychometric function slope with spatial separation. For the different-band and same-band noises, the slopes of the functions did not change significantly with spatial separation. However, for the different-band sentence masker, the slope was steeper for the 90° function (0.14 or 3.3% per dB[1]) than for the 0° function (0.10 or 2.3% per dB); a factor of 1.4 steeper. The difference in slope between the two functions resulted in diverging functions as signal level increased and therefore, the advantage of spatial separation increased (below 90% correct) with signal level. This result is not consistent with the masking level difference literature in which the psychometric function slopes are parallel or converge for N_oS_o and N_oS_Π for suprathreshold tasks such as loudness matching, frequency discrimination, or intensity discrimination (see Yost, 1997 for a review) and for nonspeech pattern identification (Kidd *et al.*, 1995). This difference is important because it illustrates the difference in mechanisms causing the release from masking. The measurement of masking level differences is dominated by energetic masking and the release from masking occurs due to a decrease in energetic masking as a result of binaural interaction. In contrast, the different-band sentence is dominated by informational masking for the 0° separation condition, and the release from masking for the 90° condition occurs due to a decrease in informational masking as a result of perceptual factors. The masking therefore becomes proportionally more energetic, resulting in a steeper psychometric function slope. The release from informational masking causes the divergence of the functions.

An interesting comparison is across both masker type and spatial separation at the same time, specifically the different-band noise masker functions (either the 0° or 90°) to the different-band sentence at 90°. The 90° separation between the signal and different-band sentence masker shifted the midpoint close to that for the different-band noise functions. The midpoint difference between the different-band sentence and different-band noise maskers shrank from 22 dB at 0° to 4–7 dB at 90°, and the slope of the different-band sentence masker function increased to 0.14 at 90° separation, toward the value of 0.19 for the different-band noise functions. Although the function for the different-band sentence masker at 90° was distinguishable from the different-band noise functions, a substantial amount of the informational masking active in the 0° separation condition was apparently eliminated by moving the different-band sentence masker 90° to the right.

The results of the error analysis for the different-band sentence masker were also affected by spatial separation. Listeners confused the signal and different-band sentence masker more often when they originated from the same location (75%–80% of errors) than when the masker was moved 90° to the right (35%–58% of errors). This suggests, along with the psychometric function slopes and magnitude of the spatial advantage, that the spatial separation between signal and masker reduced informational masking in the different-band sentence masker.

V. CONCLUSIONS

(1) The results suggest that the ratio of informational to energetic masking can be maximized in speech stimuli using a modification of cochlear implant simulation processing. This allows for the examination of informational masking in the cocktail party problem more directly by reducing the potentially confounding effects of energetic masking.

(2) For 0° spatial separation, the masking caused by the different-band sentence masker was primarily informational, as evidenced by: (a) approximately 22 dB more masking than for the different-band noise masker con-

trol; (b) shallow psychometric function slopes; and (c) high percentage of confusions with the masker sentence.

(3) A robust spatial separation effect was obtained with all types of maskers tested, but the size of the spatial release from masking differed significantly with masker type. The different-band sentence masker (mainly informational) produced the largest spatial advantage (18 dB). The same-band noise (mainly energetic) produced a significantly smaller release (7 dB), in line with previous research of speech masked by noise. The different-band noise (minimally energetic and informational) produced the smallest spatial release (4 dB).

(4) A relatively large ratio of informational to energetic masking (due to signal–masker similarity and/or stimulus uncertainty) is likely the key to producing such a large spatial release from masking.

(5) The 90° spatial separation condition reduced informational masking caused by the different-band sentence masker, as evidenced by: (a) 14 dB more spatial release from masking than for the different-band noise masker control; (b) steeper psychometric function slopes for the 90° separation condition than for 0° separation; and (c) significant reduction in the percentage of confusions with the masker sentence when the masker was moved to 90°.

(6) The results of this study provide support for the notion that spatial release from informational masking and energetic masking are produced by different mechanisms. The spatial release from an informational masker is mainly produced by a perceptual effect, whereas the spatial release from an energetic masker is primarily mediated by binaural interaction and headshadow.

ACKNOWLEDGMENTS

This work was supported by NIH/NIDCD Grant Nos. DC04545, DC04663, and DC0100, and by the Boston University Hearing Research Center. The authors are grateful to Andrew Brughera for critical technical assistance, Robert Shannon and Qian-Jie Fu of House Ear Institute for providing the cochlear implant simulation algorithm, and Richard Freyman for early discussions regarding this project. We are also grateful to Douglas Brungart and an anonymous reviewer for providing helpful comments on an earlier draft of the paper.

[1]Slope in percent per dB is from a straight-line fit between 40% and 60% correct performance.

Allen, P., and Wightman, F. (**1994**). ''Psychometric functions for children's detection of tones in noise,'' J. Speech Hear. Res. **37**, 205–215.

Bolia, R. S., Ericson, M. A., Nelson, W. T., McKinley, R. L., and Simpson, B. D. (**1999**). ''A cocktail party effect in the median plane?'' J. Acoust. Soc. Am. **105**, 1390–1391.

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.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (M.I.T. Press, Cambridge, MA).

Bronkhorst, A. W., and Plomp, R. (**1990**). ''A clinical test for the assessment of binaural speech perception in noise,'' Audiology **29**, 275–285.

Bronkhorst, A. W., and Plomp, R. (**1992**). ''Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing,'' J. Acoust. Soc. Am. **92**, 3132–3139.

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

Brungart, D. S. (**2001b**). ''Evaluation of speech intelligibility with the coordinate response measure,'' J. Acoust. Soc. Am. **109**, 2276–2279.

Brungart, D. S., and Simpson, B. D. (**2001**). ''Contralateral masking effects in dichotic listening with two competing talkers in the target ear,'' J. Acoust. Soc. Am. **109**, 2486.

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.

Cherry, E. C. (**1953**). ''Some experiments on the recognition of speech with one and two ears,'' J. Acoust. Soc. Am. **25**, 975–979.

Dirks, D. D., and Bower, D. R. (**1969**). ''Masking effects of speech competing messages,'' J. Speech Hear. Res. **12**, 229–245.

Dolan, T. R. (**1968**). ''Effect of masker spectrum level on masking-level differences at low signal frequencies,'' J. Acoust. Soc. Am. **44**, 1507–1512.

Dolan, T. R., and Robinson, D. E. (**1967**). ''An explanation of masking level differences that result from interaural intensive disparities of noise,'' J. Acoust. Soc. Am. **42**, 977–981.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). ''Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs,'' J. Acoust. Soc. Am. **102**, 2403–2411.

Duquesnoy, A. J. (**1983**). ''Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons,'' J. Acoust. Soc. Am. **74**, 739–743.

Egan, J. P., Carterette, E. C., and Thwing, E. J. (**1954**). ''Some factors affecting multi-channel listening,'' J. Acoust. Soc. Am. **26**, 774–782.

Ericson, M. A., and McKinley, R. L. (**1997**). ''The intelligibility of multiple talkers separated spatially in noise,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ), pp. 701–724.

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., 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.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). ''Spatial release from informational masking in speech recognition,'' J. Acoust. Soc. Am. **109**, 2112–2122.

Gelfand, S. A., Ross, L., and Miller, S. (**1988**). ''Sentence reception in noise from one versus two sources: Effects of aging and hearing loss,'' J. Acoust. Soc. Am. **83**, 248–256.

Hawley, M. L. (**2000**). ''Speech intelligibility, localization and binaural hearing in listeners with normal and impaired hearing,'' Ph.D. dissertation, Biomedical Engineering, Boston University, Boston, MA.

Hirsh, I. J. (**1948**). ''The influence of interaural phase on interaural summation and inhibition,'' J. Acoust. Soc. Am. **20**, 536–544.

Kidd, G., Jr., Mason, C. R., and Arbogast, T. L. (**2002**). ''Similarity, uncertainty and masking in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **111**, 1367–1376.

Kidd, G., Jr., Mason, C. R., and Rohtla, T. L. (**1995**). ''Binaural advantage for sound pattern identification,'' J. Acoust. Soc. Am. **98**, 1977–1986.

Kidd, G., Jr., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). ''Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **104**, 422–431.

Leek, M. R., Brown, M. E., and Dorman, M. F. (**1991**). ''Informational masking and auditory attention,'' Percept. Psychophys. **50**, 205–214.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Lutfi, R. A. (**1989**). ''Informational processing of complex sound. I. Intensity discrimination,'' J. Acoust. Soc. Am. **86**, 934–944.

McFadden, D. (**1968**). ''Masking-level differences determined with and without interaural disparities in masker intensity,'' J. Acoust. Soc. Am. **44**, 212–223.

Neff, D. L. (**1995**). ''Signal properties that reduce masking by simultaneous, random-frequency maskers,'' J. Acoust. Soc. Am. **98**, 1909–1920.

Neff, D. L., and Green, D. M. (**1987**). ''Masking produced by spectral uncertainty with multi-component maskers,'' Percept. Psychophys. **41**, 409–415.

Nilsson, M., Gelnett, D., Sullivan, J., Soli, S. D., and Goldberg, R. L. (**1992**). ''Norms for the hearing in noise test: The influence of spatial separation, hearing loss, and English language experience on speech reception thresholds,'' J. Acoust. Soc. Am. **92**, 2385.

Peissig, J., and Kollmeier, B. (**1997**). ''Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners,'' J. Acoust. Soc. Am. **101**, 1660–1670.

Pollack, I. (**1975**). ''Auditory informational masking,'' J. Acoust. Soc. Am. Suppl. 1 **57**, S5.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.

Simpson, B. D., Bolia, R. S., Ericson, M. A., and McKinley, R. L. (**1999**). ''The effect of sentence onset asynchrony on call sign detection and message intelligibility in a simulated 'cocktail party,' '' J. Acoust. Soc. Am. **105**, 1024.

Watson, C. S. (**1987**). ''Uncertainty, informational masking and the capacity of immediate auditory memory,'' in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 267–277.

Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). ''Factors in the discrimination of tonal patterns: II. Selective attention and learning under various levels of stimulus uncertainty,'' J. Acoust. Soc. Am. **60**, 1175–1181.

Wilson, R. H., and Carter, A. S. (**2001**). ''Relation between slopes of word recognition psychometric functions and homogeneity of the stimulus materials,'' J. Am. Acad. Audiol **12**, 7–14.

Wilson, R. H., Zizz, C. A., Shanks, J. E., and Causey, G. D. (**1990**). ''Normative data in quiet, broadband noise, and competing message for Northwestern University Auditory Test No. 6 by a female speaker,'' J. Speech Hear Disord. **55**, 771–778.

Yost, W. A. (**1997**). ''The cocktail party problem: Forty years later,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. A. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ), pp. 329–348.

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