Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns

Gerald Kidd, Jr., Christine R. Mason, Tanya L. Rohtla, and Phalguni S. Deliwala
Department of Communication Disorders, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

(Received 8 April 1997; revised 23 December 1997; accepted 2 March 1998)

A nonspeech pattern identification task was used to study the role of spatial separation of sources on auditory masking in multisource listening environments. The six frequency patterns forming the signal set were comprised of sequences of eight 60-ms tone bursts. Bursts of masking sounds were played synchronously with the signals. The main variables in the study were (1) the difference in spatial separation in the horizontal plane between signals and maskers and (2) the nature of the masking produced by the maskers. Spatial separation of signal and masker ranged from 0–180 degrees. The maskers were of two types: (1) a sequence of eight 60-ms bursts of Gaussian noise intended to produce predominantly peripherally based “energetic masking” and (2) a sequence of eight 60-ms bursts of eight-tone complexes intended to produce primarily centrally based “informational masking.” The results indicated that identification performance improved with increasing separation of signal and masker. The amount of improvement depended upon the type of masker and the center frequency of the signal patterns. Much larger improvements were found for spatial separation of the signal and informational masker than for the signal and energetic masker. This was particularly apparent when the acoustical advantage of the signal-to-noise ratio in the more favorable of the two ears (the ear nearest the signal) was taken into account. The results were interpreted as evidence for an important role of binaural hearing in reducing sound source or message uncertainty and may contribute toward solving the “cocktail party problem.” © 1998 Acoustical Society of America. [S0001-4966(98)03506-1]

PACS numbers: 43.66.Pn, 43.66.Dc, 43.66.Ba [RHD]

INTRODUCTION

In noisy environments binaural cues are thought to provide a significant advantage when the task of the listener is to attend to a particular sound source in the presence of competing sources. When the signal is separated spatially from the masker(s), interaural differences in time of arrival and intensity of the waveforms can provide cues that improve detection or recognition performance relative to the circumstance in which the sounds come from the same direction or when only monaural listening is available (e.g., Hirsh, 1950; Plomp, 1976; Saberi et al., 1991; Gilkey, 1995). One well-known acoustical factor in such conditions is the “head shadow effect.” If the masker and signal originate from the same location, the signal-to-noise ratio is roughly the same in both ears. However, if the signal and masker are spatially separated, the signal-to-noise ratio will be better in the ear nearer the signal source, particularly in the higher frequencies (e.g., Shaw, 1974). The head shadow effect can provide a substantial improvement in performance in detection or recognition tasks. A second factor is “binaural interaction”—as often inferred from studies of the “masking-level-difference” or “intelligibility-level-difference” in which the inputs from the two ears are processed in the central nervous system to improve the effective signal-to-noise ratio of the neural representation reaching the brain. Although masking- and intelligibility-level-differences may be 10–15 dB or more in optimal conditions, studies of binaural analysis applied to free-field listening consistently find relatively small effects (2–5 dB for speech) after accounting for the head shadow effect (e.g., Zurek, 1993; Plomp and Mimpen, 1981; Bronkhorst and Plomp, 1989). This suggests that binaural interaction is relatively unimportant in real-world sound reception. Thus, it is possible to argue that the large subjective advantage of binaural hearing often reported in noisy, multisource environments—sometimes called the “cocktail party effect” (cf. Cherry, 1953; Pollack and Pickett, 1958; Yost, 1991)—in reality is due primarily to head shadow or to nonauditory factors and that binaural mechanisms per se contribute relatively little [Zurek (1993), p. 274, argues that both psychoacoustic and acoustic factors contribute about equally to an average binaural advantage of about 5 dB for typical listening conditions; see also the recent review by Yost (1997)].

An implicit point in studying the processing of sounds in multisource environments is that the sounds interfere with one another. The performance on detection, discrimination, or identification tasks for a signal occurring in isolation is usually better than when it occurs in the presence of competing sounds. One sound may mask another sound because it dominates the neural representation of the stimulus in the auditory periphery. This process has been termed “energetic” or peripheral masking and has been studied extensively. In particular, many studies of binaural processing of sounds in multisource environments have used maskers such as Gaussian noise that produce masking that is thought to be primarily peripheral in origin. Likewise, models of binaural interaction usually assume that central processing of the inputs from the two ears improves the effective signal-to-noise
ratio of signals that are masked in the auditory periphery (cf. Colburn, 1995; Colburn and Durlach, 1978)—in essence, by reducing energetic masking.

During the past 20 years, many studies have demonstrated large masking effects that do not appear to be due to peripherally based energetic masking. Interference in detection, discrimination, or identification tasks can be produced when the signal is embedded in a sequence of masker tones (e.g., Watson et al., 1975, 1976; Leek and Watson, 1984) or is played simultaneously with the masker (e.g., Kidd et al., 1986; Neff and Green, 1987; Kidd et al., 1994) under conditions of high uncertainty. The effects on performance may be quite large, exceeding a 40-dB threshold shift in some conditions of high uncertainty. The effects on performance may vary from moment to moment in the degree to which they overlap in the frequency domain. The only attempt we are aware of to quantify the relative amounts of energetic and informational masking was for the tone-in-noise detection task (Lutfi, 1990).

In the present study, we examined the hypothesis that the large binaural advantage commonly thought to occur in noisy listening environments is due in part to a reduction in informational masking. This could be the case if the perception that a desired sound source is spatially distinct from competing sources reduces uncertainty along one or more stimulus dimensions. In order to test this hypothesis, we constructed maskers designed to produce large amounts of informational masking with little concomitant energetic masking in a nonspeech pattern identification task (Kidd et al., 1995b). The main experimental variables were the degree of spatial separation in the horizontal plane of signal and masker and whether the masker was the type causing peripherally based energetic masking or centrally based informational masking. Based on previous work, our expectation was that the advantage in identification performance due to spatial separation of sound sources would result primarily from the head shadow effect for the Gaussian noise masker. However, there might be an additional advantage for the informational masker if spatial separation of sources reduces listener uncertainty.

I. METHODS

A. Subjects

Three listeners with prior experience in psychoacoustic experiments, including author TR, served as subjects. Their ages ranged from 24–29 years. All three listeners had normal hearing as determined from audiometric assessment.

B. Stimuli

The stimuli were generated digitally and played through 16-bit digital-to-analog converters (Tucker-Davis Technology) at a rate of 50 000 samples per second, then low-pass filtered at 20 000 Hz. The signals were sequences of tones arranged in six frequency patterns. The signal-tone sequences were comprised of eight 60-ms bursts, including 10-ms cosine-squared rise/decay times with no delay between bursts, resulting in a total pattern length of 480 ms. The six frequency patterns were: (1) constant; (2) rising; (3) falling; (4) alternating; (5) one-step up; and (6) one-step down and are illustrated schematically in Fig. 1.

The signal patterns (and informational maskers, as described below) were generated from 16 eight-frequency bands spaced equally on a logarithmic scale. The center frequencies of the bands ranged from 215–6112 Hz. The width of the frequency bands was 14% of the center frequency. On each trial of the 16 possible bands was selected at random to contain a signal pattern. The pattern—one of the six shown in Fig. 1—was then generated from the eight frequencies available within that band. Signal level was equal for tones in a pattern. The overall level was selected randomly on each trial among a set of levels chosen to generate identification-level functions.

The masker sequences of eight 60-ms bursts played synchronously with the signal bursts. There were two types of maskers: independent samples of Gaussian noise bandpass filtered from 200–6500 Hz at a spectrum level of 32 dB and multitone complexes generated to have features similar to the signal patterns in that they were comprised of sets of narrow-band tone sequences. We chose this multitone masker because earlier work from our laboratory (Kidd and Mason, 1997) indicated that it produced large amounts of informational masking for identification, significantly greater masking, in fact, than that produced by the “multiple-bursts different” masker used in the earlier pattern-identification experiment (Kidd et al., 1995b). We have designated this sequence to be quite large, exceeding a 40-dB threshold shift in some conditions of high uncertainty. The effects on performance may vary from moment to moment in the degree to which they overlap in the frequency domain. The only attempt we are aware of to quantify the relative amounts of energetic and informational masking was for the tone-in-noise detection task (Lutfi, 1990).

In the present study, we examined the hypothesis that the large binaural advantage commonly thought to occur in noisy listening environments is due in part to a reduction in informational masking. This could be the case if the perception that a desired sound source is spatially distinct from competing sources reduces uncertainty along one or more stimulus dimensions. In order to test this hypothesis, we constructed maskers designed to produce large amounts of informational masking with little concomitant energetic masking in a nonspeech pattern identification task (Kidd et al., 1995b). The main experimental variables were the degree of spatial separation in the horizontal plane of signal and masker and whether the masker was the type causing peripherally based energetic masking or centrally based informational masking. Based on previous work, our expectation was that the advantage in identification performance due to spatial separation of sound sources would result primarily from the head shadow effect for the Gaussian noise masker. However, there might be an additional advantage for the informational masker if spatial separation of sources reduces listener uncertainty.
masker ‘multiple-bursts different-narrow-band’ (MBDN). The MBDN masker was constructed from the same set of 16 bands as the signals. On each trial, the band containing the signal and the two adjacent bands were excluded from selection as masker bands. Eight of the 13 remaining bands were randomly chosen to contain masker tones. For that burst sequence, all of the masker tones fell within the selected bands. Within each masker band, the frequency was a random sample on each burst from among the eight possible frequencies. This resulted in a masker comprised of eight narrow-band tone sequences. The level of the tones in a given masker band was constant and chosen at random from a uniform range from 50–70 dB SPL. These two maskers—Gaussian noise (referred to as ‘BBN’) and MBDN—are illustrated schematically in Fig. 2 in spectrogram form.

C. Procedures

The stimuli were presented through seven loudspeakers located 5 ft from the listener roughly at head height (when seated) and separated by 30 degrees azimuth extending from directly in front of the listener (0 degrees) to the listener’s right (+90 degrees) and left (−90 degrees). The listeners were asked to maintain a head position directly facing 0 degrees but were not constrained otherwise. The impulse responses of the individual speakers were measured and digital corrections applied to the stimuli to obtain flat frequency responses. The listening environment was quiet but not anechoic (e.g., the floor was carpeted and the subject was surrounded by cloth panel dividers and foam-covered walls; the measured reverberation time was roughly 4 s).

In the identification task, a one-interval six-alternative forced-choice (1I-6AFC) procedure was used in which the six stimulus alternatives were the six patterns. Thus, chance performance was about 16.7%. The signals were chosen randomly, with replacement, on each trial throughout a block of trials. Signal level was mixed throughout each trial block, which consisted of 50–80 trials depending on the number of levels that were tested in that condition. Response feedback was given following each trial. An illustration similar to Fig. 1 identifying the stimulus set was positioned above the response keys. The locations of the signal and, when present, the masker were chosen pseudorandomly on every presentation from among the seven speakers. The pseudorandom selection procedure chose signal and masker locations according to equal probability of occurrence for angular separation (i.e., a separation of 180 degrees was as likely as that of 30 degrees, although there were only 2 possibilities for signal and masker for 180 degrees and there were 14 for 30 degrees). The locations of the masker and signal were cued by a visual map on a computer screen prior to each trial.

The listeners were highly practiced (a minimum of several hundred trials per condition) prior to collection of the data presented here. There were three experimental conditions: identification in ‘quiet’ (no added masker) measured for each speaker location, and the two masked conditions—BBN and MBDN—in which signal and masker location/spatial separation was varied. The data plotted represent roughly 100 trials per point for the identification-level functions for each condition.

II. RESULTS

In order to evaluate the effects of signal frequency, the 16 frequency bands from which the signals were drawn were divided into four groups or ‘bins’ of four bands each. For post hoc analysis, the data were pooled within each bin. To compare performance across different conditions, the decibel values corresponding to the 60% correct point on the identification-level functions were estimated. We chose that point because it is roughly the middle of the range between chance and perfect performance and because it allows more direct comparison with the results of our previous binaural identification work (cf. Kidd et al., 1995b). The estimates were obtained by fitting straight lines to the rising portions of the middle of the functions.

A. Performance in quiet

In the quiet condition, no systematic effect of speaker location was observed, so the results were averaged across locations. The sound pressure levels corresponding to the
60% correct points on the identification-level functions, and the slopes of the fitted functions vary with frequency as would be expected from the form of the audibility curve measured in the sound field (e.g., Robinson and Dadson, 1956). For the lowest frequency region, the 60% correct points of the identification-level functions are about 31–36 dB SPL and decline to about 6–10 dB SPL in the highest frequency region. On average, the slopes of the fitted functions range from about 3%/dB in the lowest frequency bin to about 5%/dB in the highest frequency bin. Subjects achieved near-100%-correct performance in this task in quiet at moderate levels.

B. Performance in BBN and MBDN maskers

The results from the masked conditions also take the form of identification-level functions. The slopes and orderliness of the functions vary markedly according to the type of masker. This point is illustrated in Fig. 3. The parameter plotted in the figure is degree of spatial separation between masker and signal. Because of the number of functions displayed in each panel, we did not attempt to identify the angular separations. In general, though, the functions are ordered from left to right according to decreasing angular separation. A detailed analysis of the functions is given below.

Both maskers shifted the identification-level functions to the right along the level axis relative to the quiet condition (the midpoint of the function in quiet is shown by the asterisk at the left of each panel). However, the slopes of the identification-level functions were shallower—often much shallower—for the MBDN masker (solid lines) than for the BBN masker (dashed lines). This finding is consistent with that reported in the earlier study by Kidd et al. (1995b) although it is even more striking here, possibly because of the greater uncertainty in this task due to randomization in frequency of both the signal and the masker. In some cases, particularly for the highest frequency bin, performance appeared to reach an asymptote at percent correct values well below 100%. In the plots that follow the comparisons of performance are based on the decibel values corresponding to the 60% correct points. Because the slopes of the functions are so different for the two masker types the magnitude of the improvement may vary considerably depending on the level of performance at which a comparison is made. In some cases such comparisons are not even possible (e.g., high percent correct values were not obtained in the MBDN masker even at the highest signal level). There were a few
TABLE I. Sound pressure levels corresponding to 60% correct from the straight-line fits used to summarize the data (see text) for the BBN and MBDN conditions. The slopes of the fitted functions are also given in %/dB SPL. The values are for angular separations of signal and masker in the azimuthal plane. The 16 signal frequency bands were pooled to form four bins ranging from low (bin 1) to high (bin 4) frequencies (see text).

<table>
<thead>
<tr>
<th>Bin No.</th>
<th>Listener 1</th>
<th>Listener 2</th>
<th>Listener 3</th>
<th>Average (s.e.)</th>
<th>Listener 1</th>
<th>Listener 2</th>
<th>Listener 3</th>
<th>Average (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>55.1</td>
<td>56.4</td>
<td>55.4</td>
<td>52.8 (0.12)</td>
<td>57.0 (0.71)</td>
<td>5.50 (0.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>55.1</td>
<td>56.4</td>
<td>55.4</td>
<td>52.8 (0.12)</td>
<td>57.0 (0.71)</td>
<td>5.50 (0.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>52.9</td>
<td>53.5</td>
<td>52.5</td>
<td>51.8 (0.89)</td>
<td>53.3 (0.71)</td>
<td>5.21 (0.39)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>52.5</td>
<td>53.5</td>
<td>52.5</td>
<td>51.8 (0.89)</td>
<td>53.3 (0.71)</td>
<td>5.21 (0.39)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>51.8</td>
<td>53.0</td>
<td>51.4</td>
<td>51.8 (0.89)</td>
<td>53.3 (0.71)</td>
<td>5.21 (0.39)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>50.8</td>
<td>51.3</td>
<td>50.6</td>
<td>50.9 (0.89)</td>
<td>52.8 (0.67)</td>
<td>5.11 (0.39)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>49.9</td>
<td>50.3</td>
<td>49.0</td>
<td>49.9 (0.89)</td>
<td>51.8 (0.67)</td>
<td>5.00 (0.39)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bin No.</th>
<th>Listener 1</th>
<th>Listener 2</th>
<th>Listener 3</th>
<th>Average (s.e.)</th>
<th>Listener 1</th>
<th>Listener 2</th>
<th>Listener 3</th>
<th>Average (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>59.9</td>
<td>63.7</td>
<td>61.6</td>
<td>59.9 (1.12)</td>
<td>61.7 (0.71)</td>
<td>5.76 (0.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>60.5</td>
<td>63.7</td>
<td>61.8</td>
<td>62.0 (0.93)</td>
<td>63.9 (0.71)</td>
<td>5.76 (0.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>58.9</td>
<td>62.3</td>
<td>58.4</td>
<td>59.9 (1.12)</td>
<td>61.7 (0.71)</td>
<td>5.76 (0.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>53.4</td>
<td>58.9</td>
<td>57.4</td>
<td>56.6 (1.64)</td>
<td>57.6 (1.40)</td>
<td>5.58 (0.15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>52.3</td>
<td>54.1</td>
<td>54.0</td>
<td>53.5 (0.58)</td>
<td>53.5 (0.58)</td>
<td>5.37 (0.12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>50.3</td>
<td>53.7</td>
<td>52.5</td>
<td>52.1 (0.99)</td>
<td>53.5 (0.58)</td>
<td>5.37 (0.12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>49.1</td>
<td>53.9</td>
<td>51.8</td>
<td>51.6 (1.40)</td>
<td>53.5 (0.58)</td>
<td>5.37 (0.12)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
instances where we encountered difficulties in deciding which points to include in the straight-line fits. However, the various rules that were tried for including points yielded only slight differences in midpoints that did not affect the conclusions drawn from the data.

The sound pressure levels corresponding to the 60% points from the fitted functions and the slopes of the functions are contained in Table I. For the BBN masker, the 60% points of the functions fell within a 12-dB range for all listeners and conditions. The 180-degree separation produced the lowest levels, which were about the same for all four frequency bins, and the highest levels were obtained for the smaller separations for the highest frequency bin. The slopes of the functions were steepest for the middle two bins and the smaller spatial separations with identification performance increasing about 6%–7%/dB, and were shallowest for most separations of the high-frequency bin with performance increasing about 3%–5%/dB.

For the MBDN masker, there was a much greater range of 60% points—about 40 dB—and generally shallower slopes. Averaged across the three listeners, the highest midpoint was about 64 dB SPL for the lowest frequency bin and 0-degree separation while the lowest midpoint was about 24 dB SPL for the highest frequency bin and 180-degree separation. On average, the slopes of the functions ranged from about 0.78% to 2.66%/dB. Although these three listeners’ were quite similar for quiet and BBN conditions, large individual differences were apparent for the MBDN condition.

The improvement in performance measured at the 60% correct point on the identification-level functions is plotted in Fig. 4 as a function of the angular separation between signal and masker. The reference condition is 0-degree separation. For the BBN masker the results were quite consistent across listeners. Small improvements were observed with increasing signal-masker spatial separation through the lower three signal bins. On average, the largest improvement across the three listeners was less than 5 dB in all cases. For the highest frequency bin, more of an effect was observed with improvement increasing steadily with separation up to about 10 dB.

For the MBDN masker, much larger improvements were found in general and there were greater differences in performance between listeners. The largest improvement due to spatial separation for listener 1 occurred for the lowest frequency bin, while for listeners 2 and 3 the largest improvement was found for the highest frequency bin. Some of the advantages were quite large: for the lowest frequency bin the average improvement for the 180-degree separation was just over 20 dB, while listeners 2 and 3 demonstrated advantages in excess of 30 dB for the larger separations in the highest frequency bin. For all three subjects, the least improvement with spatial separation of sources occurred for frequency Bin #3. This result is likely due to the poorer localization accuracy typically found in the mid-frequencies where neither time of arrival nor intensity differences are particularly strong (e.g., Mills, 1972).

For the 30-degree separation condition, the data were examined for differences in performance as a function of the absolute azimuth of the signal. For the BBN masker, no difference was found for any frequency bin except for subject 1 who showed a small advantage for 0 degree azimuth at the highest frequency. For the MBDN masker, performance was generally better when the signal was played from the front than from the sides, although the functions were noisy and the differences between subjects were substantial. This general trend is also consistent with expectations based on the dependence of the minimum audible angle on azimuth [cf. Perrott (1984) for estimates of the minimum audible angle for concurrent stimuli].
To summarize, the listening advantage provided by spatial separation of masker and signal varied according to the type of masker. For Gaussian noise maskers, the obtained advantages were less than 5 dB for signal frequencies below about 3000 Hz, increasing to as much as 10 dB for higher signal frequencies. For the MBDN masker much larger advantages were found in most cases, often exceeding 20 dB for some subjects, and were apparent at high and low signal frequencies.

C. Role of head shadow effect

Part of the improvement found with increasing separation between signal and masker is due to the favorable signal-to-noise ratio in the ear nearest the signal. To estimate the influence of this factor on the improvements found in the preceding section, acoustic measurements of the “head shadow effect” were obtained.

Golay Code sequences of 512 20-μs digitally generated clicks were presented from each loudspeaker [the general procedures are described in more detail by Kulkarni (1992)]. A Convolutron signal processing system (Crystal River Engineering) was used for digital-to-analog conversion, analog-to-digital-conversion, and computation of impulse responses. With the exception of the D/A source, the clicks were played through the instrumentation used in the experiments. The peak-equivalent sound pressure level was 80 dB.

The stimuli were measured by 3⁄4 Etymotic microphones mounted at the entrance of the ear canals of each listener. The microphones were held in place by custom-fit ear molds that occluded the canals. The outputs of the microphones were sent to a preamplifier and antialiasing filter, then to the inputs of two A/D channels of the Convolutron. Each measurement consisted of 32 averages of the Golay sequences from which 1024-tap impulse responses were computed. The head shadow estimate was found by taking the difference in the log-magnitude spectra from the two ears. Measurements were made for each speaker location. Note that the microphone placement was at the entrance of the blocked ear canal and thus does not include the ear canal transfer function.

Results of these measurements are shown in Fig. 5. The figure plots the difference in level between the two microphones as a function of frequency for all three listeners for the ±90-degree speaker locations averaged together. The magnitude of the differences found for the other speaker locations decreased as the angle approached 0 degrees.

It is relatively straightforward to apply these measurements to the data obtained in the BBN masker. The curves reflect the improvement in signal-to-noise ratio as a function of frequency expected in the nearer of the two ears of the listener. Consider the case when both the signal and the masker originate from the same speaker: ±90 degrees, for example. Regardless of the signal-to-noise ratio, it is about the same in each ear within the “critical band” containing the signal. Now, if the signal remains in the speaker at +90 degrees and the noise is moved to the speaker at −90 degrees, the noise is attenuated by the amount plotted in Fig. 5 at each frequency in the ear nearest +90 degrees while the signal is attenuated the same amount for its frequency in the opposite ear. The identification-level function should shift along the abscissa by roughly the amount of the attenuation of the masker at the signal frequency. Using that approximation, it appears that most, if not all, of the advantage observed with increasing spatial separation of signal and BBN masker can be attributed to head shadow.

For the MBDN masker, the same frequency-dependent attenuation of the masker relative to the signal in the near ear has a more complicated effect. This is primarily because the masking is not due to the signal-to-noise ratio in the frequency region around the signal but rather to interference produced by components remote in frequency from the signal. Consider first when the signal is in the higher frequencies. Most of the components of the masker would be lower in frequency than the signal (although some components could be two bands away on either side of the signal) and thus would fall in frequency regions where there is less of a head shadow effect. Thus, it is hard to account for the large improvements found for listeners 2 and 3 in the high frequency bin based on head shadow. For the same reasons, it could be argued that the head shadow effect contributed to the large advantages found at the low signal frequencies because most of the masker components would fall higher in frequency than the signal and would be subject to larger amounts of attenuation at the nearer ear. However, even for the low frequency bin, it is unclear whether head shadow has any significant effect. This is because we do not know much about the effects of proximity of masker components in producing informational masking, or, especially, how the amount of masking is influenced by the shape of the masker spectrum or the number of components/MBDN bands. We do know that informational masking increases with increasing number of masker tones in a detection task up to 10–20 components (Neff and Green, 1987; Neff, 1995), although large amounts of masking are found in some conditions for as few as 2–4 masker tones. Also, Neff et al.’s work (1993) examining the effect of widening the “protected region” (our term meaning the frequency band from which masker tones were excluded) around the signal suggests that, for listeners demonstrating large amounts of informational masking, proximity of masker components is not a critical
factor. Moreover, when they restricted the masker components to the frequency region above, or below, the signal, the amount of masking was not greatly reduced. In the context of the present study, where frequency components remote from the signal frequency were attenuated due to head shadow, Neff et al.’s results are relevant and may be viewed as supporting a minor role of head shadow, although clearly there were important procedural differences between studies (e.g., detection versus identification; fixed versus random signal frequency). Further, because the slopes of the identification-level functions in the present study were so shallow for the informational masker, changes on the order of 10–15 dB typically do not produce large changes in performance (see Fig. 3 and Table I). Thus, we conclude that the head shadow effect was not primarily responsible for the advantages found for spatial separation of sources for the MBDN masker, particularly at the lower frequencies.

III. DISCUSSION

The results described above suggest that the magnitude of the listening advantage for the task of identification provided by spatial separation of sound sources depends on the type of masking the listener experiences. When one sound masks the other sound principally by dominating the pattern of excitation such that the peripheral neural representation of the signal is obscured by that of the masker, the listening advantage is largely a consequence of acoustical factors. The signal-to-noise ratio is improved due to the filtering of the masker by the head as it is received in the ear nearest the signal source. Overlaid on that effect is binaural analysis, which, for these conditions at least, probably contributed only slightly to the observed advantage and likely was a factor only for the lower signal frequencies. However, it should be noted that even a small signal-to-noise improvement for a Gaussian noise masker may produce substantial improvements in percent correct identification performance based on the relatively steep slopes observed for the identification-level functions.

When the masking one sound produces on another sound is informational in nature, large advantages due to spatial separation of sources are possible. These advantages do not appear to be based purely on the acoustic filtering of the head or on within-channel binaural analysis mechanisms (e.g., equalization-cancellation, cf. Colburn and Durlach, 1978).

Informational masking is usually attributed to “uncertainty” (e.g., Watson, 1987; Leek et al., 1991) and occurs despite a sufficiently robust representation of the signal in the auditory periphery for solving the task (e.g., Kidd et al., 1994; also Neff and Deteleffs, 1995). Presumably, the observer is uncertain about the exact content of the stimulus even though the critical elements of the stimulus are audible.

Because the signals and MBDN maskers used in this study were drawn from the same set of frequencies, the first tone in each narrow-band sequence could be either the beginning of a signal pattern or part of the masker. As the burst sequence progresses, the likelihood of the hypothesis that the elements falling within a given frequency band forms a signal may change, with some hypotheses becoming more likely and others less likely. It would happen very infrequently that a masker band actually contained a signal pattern because there are 8 possible combinations of frequencies in each sequence per band. However, if we assume that there is some noise in the coding of the elements in the periphery and that the listener is tolerant of approximations to patterns (cf. Kidd et al., 1995a), it is not surprising that confusions occur, i.e., that a masker band sounds enough like a signal pattern to elicit an incorrect response. Further, if the listener adopts the strategy of focusing attention on only one frequency region, performance would be at chance, so attention must rapidly switch or be divided among frequency bands. In the attempt to adequately evaluate more than one plausible-sounding narrow-band stimulus, the listener may be unable to discern enough of the signal pattern to correctly identify it. Spatial separation of sources can greatly reduce the uncertainty of the listener about where in frequency attention should be focused. In these experiments, whenever a single narrow-band sound originates from a different spatial location than the remaining frequency sequences, it is always a signal.

Studies of the effects of spatial separation of sources in speech recognition tasks have sometimes used speech as both signals and maskers. On the one hand, these studies are clearly relevant to the current study in that the listener must attend to and extract information from a target source while ignoring one or more other sources that may be highly distracting or confusing to the listener, potentially producing large amounts of informational masking. Depending on the procedures used, the maskers may be very similar to the targets which are distinguishable by various factors that are part of the experimental design (e.g., voice characteristics of the talker, the message being conveyed, etc.). On the other hand, as mentioned in the Introduction, the broadband nature of speech likely also causes large amounts of energetic masking which makes it difficult to attribute the observed effects to one process or the other and the magnitude of the advantage observed may depend on the exact procedures used.

Yost et al. (1996) studied the effect of spatial separation of sources on word recognition and localization accuracy in a sound-field listening environment. They measured performance in three conditions: “normal listening,” in which subjects were seated in the center of a semicircle of loudspeakers and made judgments about sounds played from the speakers; a condition in which recordings of the stimuli were made through the two channels of KEMAR located in the center of the sound field and subsequently played via headphones to subjects for judging; and a one-headphone condition in which a recording was made using a single microphone located in the sound field and later played monaurally to listeners. They found a small but significant improvement in word recognition for spatially separated sources in the two conditions providing binaural information (‘‘normal’’ and KEMAR conditions) relative to the one-headphone condition—especially when there were three, rather than two, simultaneous talkers. For example, the amount of the advantage in identification of letters and numbers (the ‘‘let’’ condition in Yost et al.’s analysis) for the normal condition compared to the one-headphone condition was less than 5%
for single talkers, slightly less than 10% for two talkers, but more than 35% for three talkers. Localization of sources was at chance for the one-headphone condition and much better than chance for the binaural conditions.

Yost et al. concluded that “there is not a large advantage provided by spatial listening when there are two sound sources...” but that “when the sound field becomes more complex with three concurrent sounds, spatial cues appear to play a greater role...” (p. 1034).

In the Yost et al. study the fact that speech was masking speech probably means that energetic masking contributed significantly to the results so, from that perspective, relatively small improvement due to spatial separation of sources would be expected except at the higher frequencies where head shadow would be substantial. However, it should also be noted that uncertainty was minimal: the listener was allowed to listen repeatedly to the stimulus until ready to register a response. Also, the listener was permitted to move his or her head as desired. It is possible that larger advantages due to spatial separation might have resulted if the two-source listening task was more uncertain. It is tempting to explain the larger binaural advantages found for three simultaneous sources by asserting there was greater uncertainty for the listeners than that which occurred for two sources. However, we have no directly relevant data or way of measuring uncertainty that would test that speculation.

A recent study by Koehnke et al. (1997) also examined the advantage due to spatial separation of sources in a multisource environment. In their study, two or three simultaneous talkers were presented to listeners via headphones. One talker was the “target” and the other talker(s) were “jammers.” The spatial locations of the talkers were simulated by applying transfer functions obtained from microphones located in the two ears of KEMAR. Both the test materials and the maskers were sentences. They found improvements in intelligibility at the 50% correct point of up to 15 dB for two sources separated by 180 degrees and 8 dB for three sources separated by 90 degrees relative to listening conditions where the stimuli were mixed and played monaurally. They noted that both head shadow and binaural analysis contributed to the listening advantages, but did not specify how much of the improvement was due to each factor separately.

Yost (1997) has reviewed the substantial body of literature concerning the “cocktail party problem” (Cherry, 1953). He points out that the binaural advantage generally declines with increasing signal-to-noise ratio for a variety of tasks. Specifically, the improvement in frequency and intensity discrimination [e.g., Gebhardt et al., 1972; Henning, 1973], speech recognition (Levitt and Rabiner, 1967; Carhart et al., 1968)], and loudness perception (Townsend and Goldstein, 1972) all decline to essentially no advantage once the level of the stimulus is 10–20 dB above masked threshold. Here, though, the advantage due to spatial separation of sources does not decline appreciably with level for the informational masker over a wide range of levels. Further, most identification-level functions for signals embedded in Gaussian noise extend over a relatively narrow range of 10–20 dB unless the signals are complex and have information distributed over a broad range of frequencies. Even then, though, for any given narrow frequency region performance asymptotes quickly as signal-to-noise ratio increases. In the conditions tested in this study performance increases with increasing signal-to-noise ratio much more gradually, suggesting that informational masking, and the factors that cause release from informational masking, operate well above threshold (cf. Kidd et al., 1995b). As Yost points out, little of the work done on the cocktail-party problem has been designed to mimic real-world listening environments. We certainly would not claim that the current experiments are any less artificial or “laboratory-like” than earlier work or that there is a simple extrapolation from the identification of nonspeech patterns to the recognition of speech. However, to the extent that uncertainty is a factor in communication situations like those Cherry and others envisioned when framing the cocktail-party problem, the present study may indicate one possible binaural solution.

IV. SUMMARY AND CONCLUSIONS

(1) Spatial separation of signal and masker in the horizontal plane improves the identification of nonspeech auditory patterns.

(2) Larger improvements were found for informational maskers than for energetic maskers, especially after the head shadow effect was taken into account. The magnitude of the improvement for the informational masker in some cases was greater than 30 dB, although the individual differences between subjects also were much greater.

(3) The binaural advantage for informational masking, unlike that for energetic masking, is present over a wide range of levels above masked threshold.

(4) The mechanism of masking for the two types of maskers is assumed to be different; thus, the reason for the improvement due to spatial separation is probably different as well. For the energetic masker (BBN), after accounting for the head shadow effect, the improvement due to binaural analysis is relatively small, consistent with studies of free-field intelligibility-level-differences for speech (e.g., Bronkhorst and Plomp, 1988; Zurek, 1993). However, the large advantage found for the informational masker likely is due to a reduction in uncertainty due to perceptual segregation of the sound sources. This large effect may be responsible for the powerful subjective impression of a binaural advantage in the “cocktail party effect.”

ACKNOWLEDGMENTS

This work was supported by Grant No. DC00100 from NIH/NIDCD, by the ONR-managed MURI Grant No. Z883402, by the Dudley Allen Sargent and Sargent Accelerated Research Award funds, and by the Boston University Hearing Research Center. The authors wish to express their thanks to Peter Chiu, Monica Hawley, Abhijit Kulkarni, and Susan Turney for their assistance with various aspects of this research.
We have recently measured identification-level functions for MBDN maskers shaped according to the head-shadow curves shown in Fig. 5 and presented together with the signals at 0 degree azimuth. The preliminary results suggest that the amount of improvement in performance due to the head shadow effect may be roughly comparable for the BBN and MBDN maskers. That is, very small improvements are found at low frequencies increasing to improvements of 10–15 dB at high frequencies.


Plomp, R. (1976). "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)," Acustica 34, 200–211.


