Informational and energetic masking effects in the perception of multiple simultaneous talkers

Douglas S. Brungart[a)]
Air Force Research Laboratory, Human Effectiveness Directorate, 2610 Seventh Street, Wright-Patterson AFB, Ohio 45433

Brian D. Simpson
Veridian, 5200 Springfield Pike, Suite 200, Dayton, Ohio 45431

Mark A. Ericson and Kimberly R. Scott
Air Force Research Laboratory, Human Effectiveness Directorate, 2610 Seventh Street, Wright-Patterson AFB, Ohio 45433

(Received 13 February 2001; accepted for publication 13 August 2001)

Although many researchers have examined the role that binaural cues play in the perception of spatially separated speech signals, relatively little is known about the cues that listeners use to segregate competing speech messages in a monaural or diotic stimulus. This series of experiments examined how variations in the relative levels and voice characteristics of the target and masking talkers influence a listener's ability to extract information from a target phrase in a 3-talker or 4-talker diotic stimulus. Performance in this speech perception task decreased systematically when the level of the target talker was reduced relative to the masking talkers. Performance also generally decreased when the target and masking talkers had similar voice characteristics: the target phrase was most intelligible when the target and masking phrases were spoken by different-sex talkers, and least intelligible when the target and masking phrases were spoken by the same talker. However, when the target-to-masker ratio was less than 3 dB, overall performance was usually lower with one different-sex masker than with all same-sex maskers. In most of the conditions tested, the listeners performed better when they were exposed to the characteristics of the target voice prior to the presentation of the stimulus. The results of these experiments demonstrate how monaural factors may play an important role in the segregation of speech signals in multitalker environments.
 [DOI: 10.1121/1.1408946]

PACS numbers: 43.66.Pn, 43.66.Rq, 43.71.Gv [LRB]

I. INTRODUCTION

Many everyday listening situations require the extraction of information from speech signals that are masked by one or more simultaneous competing talkers. In most of these situations, the competing speech signals originate from different locations in the room and the listeners can take advantage of differing inputs to the two ears to spatially segregate the competing messages. This is the classic ''cocktail party'' problem that was first described by Cherry (1953) and has been extensively studied over the past 50 years [see Bronkhorst (2000) and Ericson and McKinley (1997) for recent reviews of the literature in this area]. However, when the target and masking speech originate from the same direction relative to the listener, or when the competing speech signals are presented monaurally or diotically, no binaural segregation cues are available and the listeners must rely on monaural cues to segregate the competing messages. Examples of monaural speech segregation cues include differences in the individual vocal characteristics of the target and masking talkers [vocal tract size, fundamental frequency ($F0$), accent, speaking style, etc.], differences in the prosodic features of the target and masking speech, and differences in the overall levels of the target and masking signals (Darwin and Hukin, 2000; Bregman, 1994).

Previous studies that have examined the diotic or monaural perception of two competing speech signals have shown that differences in the vocal characteristics of the competing talkers, such as differences in target and masker sex, can dramatically improve the intelligibility of the target speech (Brungart, 2001b; Festen and Plomp, 1990). These studies have also shown that listeners can use differences in the levels of the two talkers to selectively attend to the quieter talker in the stimulus (Brungart, 2001b), and that signal-to-noise ratio (SNR) consequently has relatively little influence on the intelligibility of the target talker at SNRs from 0 dB to −10 dB (Egan, Carterette, and Thwing, 1954; Dirks and Bower, 1969).

The results of these 2-talker experiments suggest that differences in the vocal characteristics and the overall levels of the competing talkers would be important in the perception of 3-talker or 4-talker stimuli, but they do not provide any direct quantitative evidence to support this hypothesis. Previous studies that have directly examined the perception of three or more simultaneous talkers have focused primarily on binaural segregation cues (the cocktail-party phenomenon) and not on monaural segregation of speech

a)Author to whom correspondence should be addressed; electronic mail: douglas.brungart@wpafb.af.mil

(Abouchacra *et al.*, 1997; Ericson and McKinley, 1997; Crispien and Ehrenberg, 1995; Drullman and Bronkhorst, 2000; Hawley *et al.*, 1999; Nelson *et al.*, 1999; Peissig and Kollmeier, 1997; Yost *et al.*, 1996). Of the handful of experiments that have examined the perception of three or more monaurally or diotically presented speech signals, the majority have done so only indirectly, either as a control condition in a cocktail-party experiment (Hawley *et al.*, 2000; Drullman and Bronkhorst, 2000; Yost *et al.*, 1996; Ericson and McKinley, 1997) or as a control condition in a dichotic listening experiment (Carhart *et al.*, 1969). Only one early experiment (Miller, 1947) systematically examined the effects of varying both the SNR and the number of competing talkers on the perception of a diotic multitalker stimulus. These studies have shown that the intelligibility of the target talker decreases when additional competing talkers are added to the stimulus and when the level of the target speech is reduced relative to the levels of the competing talkers in the stimulus.

One important factor that has not yet been fully explored is the effect that different configurations of target and masker sex have on the perception of multitalker speech stimuli. Although previous experiments with two talkers have shown that voice characteristics play an important role in speech segregation (Brungart, 2001b), very little effort has been made to systematically examine the effects of target and masker sex in listening environments with more than two competing talkers.

This paper describes a series of experiments examining the effects that differences in the vocal characteristics and overall levels of the competing talkers have on the perception of a target phrase in a multitalker speech signal. These experiments were based on a previous experiment that used the coordinate response measure (CRM) to examine the effects of target and masker sex and target-to-masker ratio (TMR) on the perception of two simultaneous talkers (Brungart, 2001b). The CRM method, which requires listeners to identify one of eight numbers and one of four colors in each target phrase, was selected both to allow direct comparison with the earlier 2-talker results and to emphasize the effects of informational masking in the multitalker results. Note that the term ''informational masking'' refers to listening situations where the target and masker signals are clearly audible but the listener is unable to segregate the elements of the target signal from the elements of the similar-sounding distracters (Freyman *et al.*, 1999; Doll and Hanna, 1997; Kidd *et al.*, 1995; Kidd *et al.*, 1994; Watson *et al.*, 1976). This differs from traditional ''energetic masking,'' where competing signals overlap in time and frequency in such a way that portions of one or more of the signals are rendered inaudible. The results of an earlier 2-talker experiment (Brungart, 2001b) showed that speech perception with the CRM was dominated by informational masking: the listeners were generally able to hear both competing speech messages, but they had difficulty segregating the content of the target phrase from the content of the masking phrase. Although it is reasonable to expect the effects of energetic masking to increase when more talkers are added to the stimulus, the strong informational masking effects found in that 2-talker experiment suggest that informational masking may also influence performance in the CRM task when the target phrase is masked by more than one competing talker.

Three different experiments were conducted. The first experiment, which was essentially a 3-talker and 4-talker extension of an earlier 2-talker experiment (Brungart, 2001b), examined the effects of TMR and target and masker voice characteristics when all of the masking voices were presented at the same level relative to the target phrase. The second experiment examined 3-talker listening situations where all three talkers were presented at different levels. The third experiment examined how providing *a priori* information about the vocal characteristics of the target talker affected performance in the CRM task.

II. EXPERIMENT 1: EFFECTS OF TARGET AND MASKER SEX AND TARGET-TO-MASKER RATIO

A. Methods

1. Stimuli

The target and masking phrases used in this experiment were taken directly from the publicly available CRM speech corpus for multitalker communications research (Bolia *et al.*, 2000). This corpus, which is based on a speech intelligibility test first developed by Moore (1981), consists of phrases of the form ''Ready (call sign) go to (color) (number) now'' spoken with all possible combinations of eight call signs (''Arrow,'' ''Baron,'' ''Charlie,'' ''Eagle,'' ''Hopper,'' ''Laker,'' ''Ringo,'' ''Tiger''); four colors (''blue,'' ''green,'' ''red,'' ''white''); and eight numbers (1–8). Thus, a typical utterance in the corpus would be ''Ready Baron go to blue five now.'' Eight talkers (four male, four female) were used to record each of the 256 possible phrases, so a total of 2048 phrases are available in the corpus.

In the speech-on-speech masking conditions of this experiment, each stimulus presentation consisted of three or four simultaneous phrases from the CRM corpus: a target phrase with the call sign ''Baron'' and two or three masker phrases with different randomly selected call signs other than ''Baron.'' The CRM phrases in each trial were randomly selected with the restriction that all of the target and masker phrases contained different color coordinates and different number coordinates.

The overall level (RMS power) of each masker phrase was set to the same level (approximately 60–70 dB SPL). The overall level (RMS power) of the target phrase was adjusted relative to the levels of the masker phrases to produce 1 of 10 randomly chosen TMRs ranging from −12 dB to +15 dB in 3 dB steps. The target and masker phrases were added together, and the combined signal was randomly roved over a 6 dB range (in 1 dB steps) before being presented to the listener diotically over headphones. This roving prevented the listeners from using level to differentiate between the target and masking talkers, and ensured that the results were not dependent on any one particular overall stimulus level. Note that throughout this paper target-to-masker ratio (TMR) refers to the ratio of the target talker to one individual masking talker, while signal-to-noise ratio (SNR) refers to the ratio of the target talker to the total combined masking signal. Thus, when the levels of all three talkers are the same

TABLE I. Number of trials collected in each experimental condition.

Condition	Total trials	Trials per TMR
TD[a]	8828	883
TS[a]	6884	688
TT[a]	2288	289
TM[a]	2400	480
TN[a]	9020	902
TDD	3542	354
TSD	4792	479
TSS	2825	283
TTT	2161	216
TMM	2500	357
TDDD	2053	205
TSDD	2208	221
TSSD	2053	205
TSSS	2196	220
TTTT	2277	228
TMMM	2520	360

[a]Data collected in previous 2-talker experiment (Brungart, 2001b).

in the 3-talker condition of this experiment, the TMR of each masking talker would be 0 dB and the overall SNR would be approximately −3 dB.

The different voice configurations used in the experiment are referred to by alphabetic codes describing the relative similarities of the target and masking voices. The first letter in the code is always a T, representing the target phrase. Additional letters are added to represent each masking phrase in the stimulus: a T represents a phrase spoken by the same talker used in the target phrase; an S represents a different talker of the same sex as the target talker; and a D represents a talker who was different in sex than the target talker. For example, in the 3-talker conditions, four different target and masker voice configurations were tested (Table I). In the TDD configuration, the two masking talkers were both different in sex than the target talker. In the TSD configuration, one of the masking talkers was the same sex as the target talker, and the other masking talker was a different sex than the target talker. In the TSS configuration, the masking phrases were both spoken by talkers who were the same sex as the target talker. And in the TTT configuration, the same talker was used in the target phrase and in both masking phrases.

In addition to these speech-on-speech masking conditions, two additional conditions were tested in which the masking talkers were replaced by two (TMM) or three (TMMM) envelope-modulated speech-shaped noise maskers. The speech-shaped noise used for these masking signals was produced by spectrally shaping Gaussian noise with a FIR filter matching the average long-term RMS spectrum of the 2048 sentences in the CRM corpus (Brungart, 2001b). This spectrally shaped noise was then modulated with the envelopes of two or three randomly selected competing speech phrases from the CRM corpus (calculated by convolving the absolute value of the speech waveform with a 7.2 ms rectangular window) and added together in order to produce a noise signal with roughly the same temporal distribution of energy as the 2- or 3-talker competing speech signals. The level of each modulated noise masker was determined from the overall RMS power of the noise waveform, and set to the same level used for each competing talker in the speech-on-speech masking conditions. The modulated noise conditions were tested at 7 different target-to-masker ratios ranging from −12 dB to 6 dB in 3 dB steps.

2. Listeners

Nine paid listeners, five male and four female, participated in the experiment. All had normal hearing (15 dB HL from 500 Hz to 6 kHz) and their ages ranged from 21 to 55. Each had participated in previous auditory experiments, and all had previous experience in an earlier experiment using the same speech materials with two competing talkers (Brungart, 2001b).

3. Procedure

The listeners participated in the experiment while seated at a control computer in a quiet listening room. On each trial, the speech stimulus was generated by a sound card in the control computer (Soundblaster AWE-64) and presented to the listener diotically over headphones (Sennheiser HD-520). Then an eight-column, four-row array of colored digits corresponding to the response set of the CRM was displayed on the CRT, and the listener used the mouse to select the colored digit corresponding to the color and number used in the target phrase containing the call sign ''Baron.''

The trials in the speech-on-speech masking conditions were divided into blocks of 120, each taking approximately 12 minutes to complete. Within each block of trials, each talker in the corpus was used as the speaker of the target phrase in exactly 15 trials. The masking talkers were selected randomly in order to produce a roughly even balance across the different possible configurations of target sex and masker sex shown in Table I. The masking talker selection in the 3-talker condition was initially done randomly, which produced a large number of instances of the TSD configuration (roughly 48%) and only a small number of instances of the TTT configuration (roughly 5%). In order to produce a more even distribution across the talker configurations in the overall results, the selection process was changed approximately 2/3 of the way into the data collection process to select the TTT configuration in 40% of the trials and the TSD configuration in 10% of the trials. This manipulation resulted in the overall distribution of trials shown in Table I. On each trial of the 4-talker configuration, the target and maskers were equally likely to occur in any of the five talker-masker configurations shown in Table I. All other variables in the 3- and 4-talker presentations, including the masking call signs, the numbers and colors of the target and masker phrases, and the target-to-masker ratio, were chosen randomly with replacement on each trial. Each of the nine listeners first participated in approximately 2000 trials in the 3-talker condition, resulting in 1480 valid trials per subject which were used in the subsequent analysis.[1] Each listener then participated in 1200 trials in the 4-talker condition.

The modulated-noise trials were collected after the speech-on-speech masking trials were completed. The trials in each of the modulated noise conditions were divided into

two blocks of 140 trials collected on different days. The TMM configuration was collected first, followed by the TMMM configuration. The total number of trials collected in each condition, as well as the number of trials for each target-to-masker ratio tested in each condition, is shown in Table I.

4. Comparison of 3-talker and 4-talker results to 2-talker results

In order to provide additional insight into the 3-talker and 4-talker conditions examined in this experiment, the results of these conditions are compared directly to the results of an earlier experiment that examined 2-talker speech-on-speech masking with the same nine listeners and the same CRM speech materials used in this experiment (Brungart, 2001b). A total of three 2-talker configurations were tested in that experiment (denoted by the superscript ''a'' in Table I): a different-sex masking condition (TD), a same-sex masking condition (TS), and a same talker masking condition (TT). In addition, a TM configuration was tested with a single modulated speech-spectrum-shaped noise masker of the same type used in the TMM and TMMM configurations of this experiment, and a TN configuration was tested in which a continuous speech-spectrum-shaped noise masker was rectangularly gated to the same length as the target phrase. Because all of the procedures used in these 2-talker conditions were essentially the same as those used in the 3- and 4-talker conditions of this experiment, the results of the two experiments are directly comparable.

B. Results and discussion

Figure 1 shows the percentage of trials where the listener correctly identified both the color and the number in the target phrase as a function of target-to-masker ratio for each of the different target-masker configurations in the experiment. The 2-talker configurations are shown in the top panel of the figure, the 3-talker configurations are shown in the middle panel of the figure, and the 4-talker configurations are shown in the bottom panel of the figure. The data have been averaged across the nine listeners used in the experiment, and the error bars in the figure represent the 95% confidence interval of each data point. The results in the figure indicate that overall performance depended on both the relative similarities between the target and masking voices and the TMR. This was verified by conducting three separate within-subject ANOVAs on the factors of TMR and target-masker configuration for the 2-talker, 3-talker, and 4-talker data shown in each panel of Fig. 1. In these analyses, the percentages of correct responses were calculated separately for each listener for each combination of target-masker configuration and TMR. These percentages were then transformed with the arcsine transform and used as the dependent variables in the within-subject ANOVAs. The results of these analyses showed that the main effect of TMR and target-masker configuration and the interactions between TMR and target-masker configuration were significant at the $p < 0.001$ level for each of the three panels of Fig. 1.

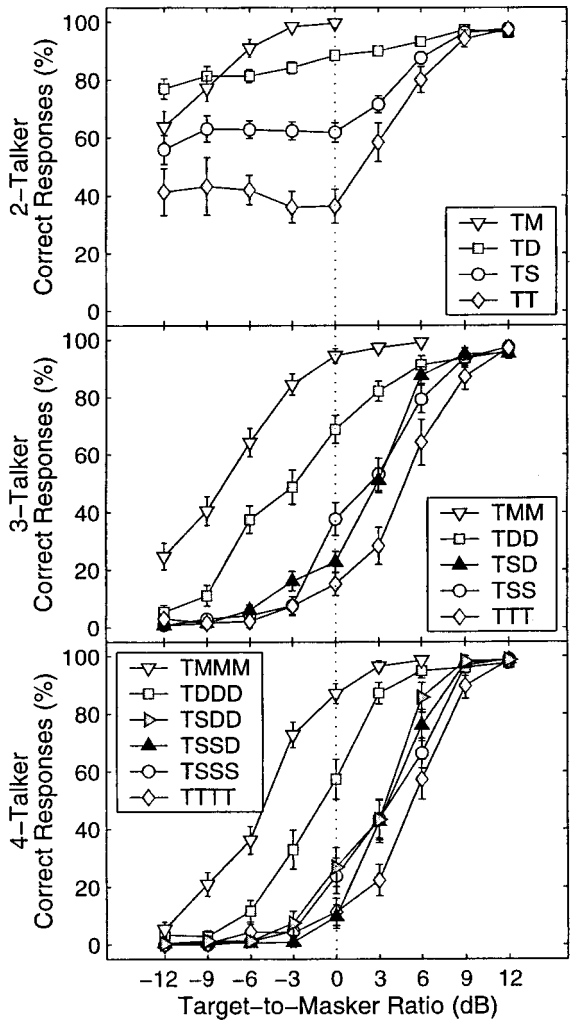

FIG. 1. Percentage of trials in Experiment 1 in which the listeners correctly identified both the color and number coordinates in the target phrase as a function of the target-to-masker ratio. The top panel shows results from an earlier 2-talker experiment that used the same speech materials and listeners as the 3-talker and 4-talker conditions tested in this experiment (Brungart, 2001a, 2001b). The middle panel shows results for the 3-talker conditions tested in this experiment, and the bottom panel shows results for the 4-talker conditions tested in this experiment. The legends indicate the different target-masker voice configurations used: T=target voice; S=voice of different talker of the same sex as the target voice; D=voice of a talker of the opposite sex of the target voice; M=envelope-modulated noise masker. The error bars represent 95% confidence intervals (±1.96 standard errors) in each condition.

The effects of the different voice configurations were relatively consistent across the 2-, 3-, and 4-talker conditions. In general, performance was best when the target voice was qualitatively different than the masking voices, and performance was worse when the target voice was qualitatively similar to the masking voices. The percentage of correct responses was consistently highest when the target talker was a different sex than any of the masking talkers (TD, TDD, and TDDD, shown by the squares in the figure) and the percentage of correct responses was generally lowest when the same voice was used in the target and masking phrases (TT, TTT, and TTTT, shown by the diamonds in the figure). Performance in the other voice configurations varied systematically with the qualitative similarity of the target and masking voices only when the target phrase was presented at a sub-

stantially higher level than the masking phrases. At a TMR of +6 dB, for example, performance decreased systematically with the number of masking talkers who were the same sex as the target talker in both the 3-talker configurations (TDD>TSD>TSS) and the 4-talker configurations (TDDD>TSDD>TSSD>TSSS). When the TMR was reduced to +3 dB, however, the performance differences between the same-sex and mixed-sex masking configurations were drastically reduced (TSD≈TSS; TSSS ≈TSDD≈TSSD) and, when the TMR was reduced to 0 dB (indicated by the vertical dotted line in the figure), performance was actually substantially lower in some of the mixed-sex masking configurations than in the same-sex masking configurations. Specifically, in those 3- and 4-talker configurations where one masking talker was a different sex than the other talkers in the stimulus (TSD, TSSD), performance was lower than in the corresponding same-sex masking configurations (TSS, TSSS) when the TMR was near 0 dB. This effect, which we refer to as ''odd-sex distraction,'' cannot be explained by traditional theories of energetic or informational masking. Energetic masking would be expected to cause more interference with a same-sex masker than with a different-sex masker because of the greater spectral overlap of two same-sex speech signals. One might also expect more informational masking to occur when the masker is qualitatively similar to the target than when the masker is qualitatively different from the target. Odd-sex distraction appears to be a special form of informational masking in which a particularly salient masker causes the listener's attention to be drawn away from the target phrase. This occurs only when the overall levels of the talkers in the stimulus are similar enough (TMRs near 0 dB) that the listener must rely entirely on vocal characteristics to segregate the competing talkers. The strength of this distracting effect is illustrated by the number of trials where the listeners' responses matched both the color and number spoken by the odd-sex masking talker. In the TSD configuration, the listeners were nearly as likely to respond with both the color and number spoken by the odd-sex masking talker as they were to respond with both the color and number spoken by the target talker (21% vs 24% at a TMR of 0 dB). In the TSSD configuration, they were actually more likely to respond with the coordinates spoken by the odd-sex talker than with the coordinates spoken by the target talker (17% versus 10%). Odd-sex distraction is examined in more detail in Experiments 2 and 3.

Further insights into the differences between the 2-talker, 3-talker, and 4-talker conditions can be obtained by examining the results as a function of the overall decreases in SNR that occur when additional masking talkers are added to the stimulus. At a TMR of 0 dB, where all the talkers are speaking at the same level, the total power present in the masking speech (relative to the target speech) is 0 dB in the 2-talker condition, approximately 3 dB in the 3-talker condition, and approximately 4.8 dB in the 4-talker condition. Figure 2 plots performance in the different-sex (top row), same-sex (second row), same-talker (third row), and noise-masker (bottom row) conditions of the experiment as a function of TMR (left column) and as a function of SNR (right column). As in Fig. 1, the data have been averaged across the 9 listeners in the experiment and the error bars represent 95% confidence intervals for each data point.

The TMR data plotted in the left column of the figure show that, within a given voice configuration, there was a substantially larger degradation in performance when the number of talkers was increased from two to three than when the number of talkers was increased from three to four (Fig. 2, left column). This occurred because overall performance degraded substantially more with decreasing TMR in the 3-talker and 4-talker conditions than in the 2-talker condition. In the 2-talker condition, performance was essentially independent of TMR at TMR values lower than 0 dB. In the 3-talker and 4-talker conditions, however, performance decreased monotonically with decreasing TMR in every target-masker configuration tested. Other multitalker experiments have also shown that the addition of a second interfering talker produces a much larger decrease in performance than the addition of the initial interfering talker (Hawley *et al.*, 2000; Drullman and Bronkhorst, 2000; Miller, 1947).

The SNR data plotted in the right column of the figure show that the number of talkers present in the stimulus had different effects on performance at positive and negative SNRs. At positive SNRs, performance generally increased with the number of competing talkers. This may have occurred because the difference in level between the target talker and any one of the masking talkers (TMR) increased as the number of competing talkers increased at a fixed SNR. If the listeners were using differences in the overall levels of the talkers as a means to segregate the target speech from the masking speech, this could explain why they performed better as the number of competing talkers increased at a fixed SNR.

At negative SNRs, the number of talkers had a somewhat different effect on the intelligibility of the target phrase. The listeners were apparently able to focus their attention on the less intense voice in the 2-talker stimulus even at SNRs as low as −12 dB. However, in the 3- and 4-talker stimuli, there is no indication that they were able to use differences in level to segregate the less intense target speech from the more intense masking speech. Performance in the 3–4 talker conditions dropped off rapidly with decreasing SNR, and was in all cases near chance when the SNR was less than −9 dB (Fig. 2, right panels). It is interesting to note that performance at negative SNRs was almost identical in the 3-talker and 4-talker conditions in each of the three speech-on-speech voice configurations shown in Fig. 2. This suggests that performance at negative SNRs with three or more competing talkers depends primarily on the voice characteristics and total power of the masking speech, and not on the number of masking talkers in the stimulus.

The results from the modulated noise masking conditions indicate that energetic masking was probably not a controlling factor in any of the listening conditions tested. Performance in the conditions with two or three modulated noise maskers (triangles in Fig. 1) was substantially better than in any of the conditions with a corresponding number of speech maskers. The modulated noise maskers contained roughly the same temporal distribution of energy as the

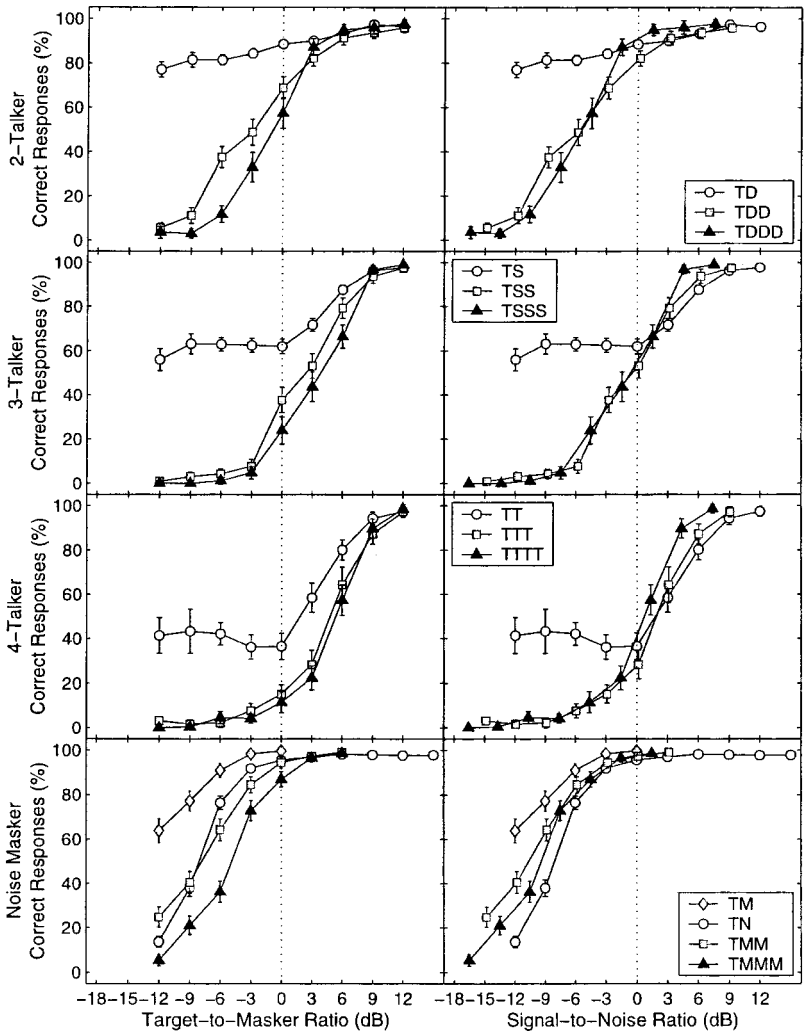

FIG. 2. Correct color and number identifications in the 2-talker, 3-talker, and 4-talker conditions of Experiment 1 for different-sex maskers (top row), same-sex maskers (second row), same-talker maskers (third row), and modulated noise maskers (bottom row). The left column shows the results as a function of target-to-masker ratio, which is defined as the level of the target voice relative to the levels of the individual masker voices. The right column shows the results as a function of SNR, which is the level of the target voice (RMS power) relative to the summed level of all of the masker voices (RMS power). The error bars represent 95% confidence intervals (± 1.96 standard errors) in each condition.

speech signals, but their performance curves were consistently shifted 3 dB to the left of the easiest speech masker configurations tested (TDD, TDDD), and as much as 12–15 dB to the left of the more difficult speech masker configurations (TTT, TTTT). In other words, the modulated noise signal had to be 3–15 dB more intense (in terms of overall RMS power) than a multitalker speech signal to produce the same amount of masking. Although the time-frequency energy distributions in the modulated noise maskers were not identical to those in the speech maskers, it does not seem likely that they were different enough to account for these differences in masking effectiveness. Thus, there is reason to believe that some form of nonenergetic masking was occurring even in the performance curves at negative SNRs in the 3-talker and 4-talker conditions. The roles of informational and energetic masking in multitalker speech perception are further explored in Experiments 2 and 3.

When the number of modulated noise talkers was increased at a fixed SNR, overall performance systematically decreased (Fig. 2, bottom right panel). It is likely that this occurred because the temporal distribution of energy was more uniform when the number of modulated noise maskers was increased and the listeners were less able to listen to the target phrase between ''the gaps'' of the masking signal (Bronkhorst and Plomp, 1992). The performance levels in the TMM and TMMM configurations fell between those for the single modulated noise masking condition (TM) and the continuous (unmodulated) noise masking conditions (TN). This is essentially the same pattern of performance reported by Bronkhorst and Plomp (1992) for speech intelligibility as the number of speech-modulated noise maskers increased.

III. EXPERIMENT 2: MASKING EFFECTS OF TWO COMPETING TALKERS WITH DIFFERENT TMRs

In the first experiment, all of the masking voices in the stimulus were presented at the same level relative to the target speech. In many real-world listening situations, however, the target and masking voices are all at different levels. In order to determine how intelligibility is affected by differences in the TMRs of the different masking talkers, a second experiment was conducted in which the levels of two masking talkers were varied independently. In half of the trials, both masking talkers were the same sex as the target; in the remaining trials, one masking talker was the same sex as the target and the other was a different sex.

A. Methods

The procedures used in the second experiment were nearly identical to those used in the first experiment. The

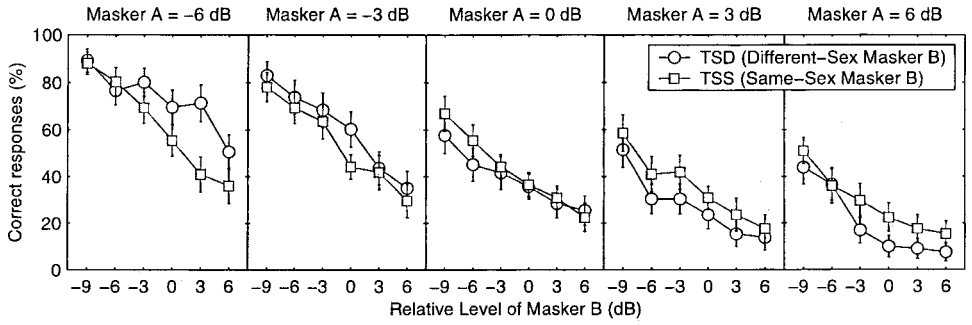

FIG. 3. Percentage of correct color and number identifications in Experiment 2 as a function of the relative levels of a same-sex masker phrase (Masker A) and a same- or different-sex masking phrase (Masker B). The five panels represent different relative levels of Masker A. In the leftmost panel, Masker A was presented at a level 6 dB lower than the target talker, and in the rightmost panel Masker A was presented at a level 6 dB higher than the target phrase. Within each panel, each symbol represents a different relative level of Masker B. The two curves in each panel represent same-sex and different-sex voices for Masker B, as indicated in the legend. The error bars represent 95% confidence intervals (±1.96 standard errors) in each condition.

speech signals were selected randomly from the CRM speech corpus, adjusted to the correct TMR according to the RMS powers of the target and masking speech signals, electronically combined, output through a DAC (Tucker–Davis DD1), and presented to the listeners diotically over headphones. The listeners then responded with the color and number coordinates contained in the target phrase addressed to the call sign ''Baron.''

The target-masker configurations used in the second experiment were somewhat different than those used in the first experiment. Every stimulus presentation contained three different talkers: the target phrase; a same-sex masking phrase (Masker A) presented at a level (relative to the target) of −6 dB, −3 dB, 0 dB, 3 dB, or 6 dB; and a same-sex or different-sex masking phrase (Masker B) presented at a level (relative to the target) of −9 dB, −6 dB, −3 dB, 0 dB, 3 dB, or 6 dB. In all, a total of 45 unique target-masker configurations were tested in the second experiment.

Eight paid volunteer listeners (3 males, 5 females) participated in the experiment. All eight were also participants in the first experiment. Each listener participated in a total of 1200 trials, divided into six blocks of 180 trials each plus one additional block of 120 trials. Within each block, the trials were randomly distributed across the 45 different possible target-masker configurations. The total experiment was completed over a two-week period, with each listener participating in 1-2 blocks of trials each day.

B. Results and discussion

Figure 3 shows the percentage of correct color and number identifications as function of the relative level of Masker B for each of the five different relative levels of Masker A tested in Experiment 2. The results have been averaged across the eight listeners used in the experiment, and they are plotted separately for the conditions with a same-sex Masker B (circles) and those with a different-sex Masker B (squares). These results show that the percentage of correct responses systematically decreased when the relative level of either of the two masking voices was increased. This can be seen from the decrease in the percentage of correct responses that occurred when the relative level of Masker A was increased at a fixed level of Masker B (moving from the left panel to the right panel of Fig. 3) and when the relative level of Masker B was increased at a fixed level of Masker A (moving from left to right within each panel of Fig. 3). A three-factor within-subject ANOVA conducted on the arcsine-transformed individual listener results for the factors of Masker A TMR, Masker B TMR, and Masker B Sex verified that the main effects of Masker A TMR and Masker B TMR were both significant at the $p<0.001$ level.

In general, performance was degraded more by the addition of a same-sex Masker B when the level of the same-sex Masker A was lower than the level of the target voice (left panel of Fig. 3), and was degraded more by the addition of a different-sex Masker B when the level of the same-sex Masker A was higher than the level of the target voice (right panel of Fig. 3). The three-factor ANOVA conducted on the results showed that this interaction between Masker A TMR and Masker B sex was significant at the $p<0.001$ level ($F_{(4,420)}=3.454$). Traditional theories of energetic masking, based on the spectral overlap of the target and masking signals, and informational masking, based on the qualitative similarity of the target and masking signals, would predict better performance in the TSD condition at all levels of Masker A. The superior performance that occurred in the TSS condition when Masker A was 6 dB more intense than the target phrase seems to be directly related to the odd-sex distraction that occurred at TMRs near 0 dB in Experiment 1. Apparently listeners are more susceptible to distraction from an odd-sex talker than to interference from a same-sex talker when they are listening to the quieter of two same-sex voices. This increased vulnerability to distraction may be related to the intense concentration required to selectively focus attention on the quieter of two same-sex talkers, which is inherently more difficult than listening to the more intense talker in the stimulus. Apparently the presence of a more salient different-sex voice makes it difficult to remain focused on the quieter of two same-sex talkers.

It is not clear why the performance advantage of the TSS configuration over the TSD configuration did not extend to the case where all three talkers were presented at the same level (center panel of Fig. 3), as it did in Experiment 1. One possible explanation is that the listeners were able to learn that the target phrase was never spoken by the odd-sex talker in the TSD configuration of the second experiment (as it was in the TDD configuration of the first experiment) and that

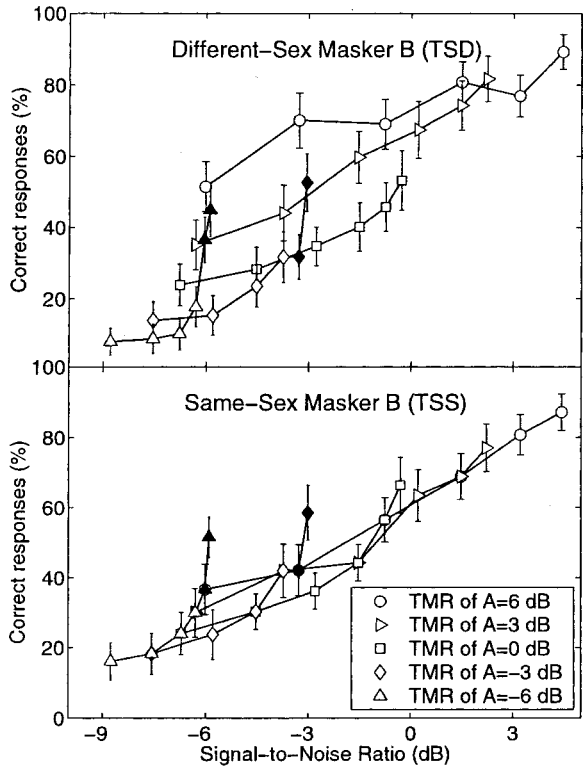

FIG. 4. Performance in Experiment 2 plotted as a function of SNR. The top panel shows the data from the TSD configuration, and the bottom panel shows the data from the TSS configuration. Within each panel, the different curves represent nine different levels of Masker B (relative to the target) at the fixed level of Masker A (relative to the target) shown in the legend. The filled symbols represent points where Masker A was presented at a level that was 12 dB or more higher than the level of Masker B (see text for details). The error bars represent 95% confidence intervals (± 1.96 standard errors) in each condition.

they adopted a strategy of selectively ignoring the odd-sex talkers in the TSD configurations of Experiment 2. The results of a third experiment, which is discussed later, will show that listeners are able to selectively ignore the odd-sex talker when they have *a priori* knowledge about the sex of the target talker.

The results shown in Fig. 3 are replotted in Fig. 4 as a function of overall SNR. In other words, each symbol in Fig. 4 represents one of the symbols shown in Fig. 3 plotted as a function of the ratio of the RMS energy in the target speech to the RMS energy in the combined two-talker masking signal. Different symbols have been used for each level of Masker A, and the data for the TSS and TSD conditions have been plotted in separate panels. In both the TSS and TSD configurations, the resulting performance curves increased abruptly with respect to the SNR whenever Masker A was presented at a level that was 12 dB or more higher than the level of Masker B (filled symbols in Fig. 4). It is likely that these abrupt increases occurred because the combined signal from the target talker and Masker A was intense enough to render Masker B inaudible, effectively reducing the TSS and TSD listening tasks to the much easier 2-talker TS listening task. Thus, the curves in these configurations may reflect a transition from the relatively low level of performance that occurs in 3-talker listening at negative SNRs to the relatively high level of performance that occurs in 2-talker listening at negative SNRs (see Fig. 2). The large improvements in performance that occurred with almost negligible increases in SNR in these curves provide strong evidence that informational masking plays an important role in 3-talker speech perception—energetic masking alone would not be expected to generate such large changes in performance with such small changes in SNR. These curves also suggest that our ability to attend to the quieter talker in a 2-talker stimulus is ''fragile'': performance in this task appears to drop off precipitously as soon as a third talker becomes audible, even when that talker has a negligible effect on overall SNR of the stimulus.

In conditions where there was no abrupt increase in performance with respect to SNR (open symbols in Fig. 4), there was a clear distinction between the performance curves for the TSD configuration shown in the top panel of Fig. 4 and the performance curves for the TSS configuration shown in the bottom panel of Fig. 4. In the TSD configuration, the vertical positions decrease systematically as the relative level of Masker A was increased. These curves show that performance in the TSD configuration depended on both the individual levels of the masking talkers and the overall SNR of the stimulus. However, the curves in the lower panel of Fig. 4 suggest that performance in the TSS configuration depended almost exclusively on the overall SNR of the stimulus. The individual levels of the masking talkers had little impact on the results, and the data from the TSS configuration were clustered into a tight distribution, suggesting an almost linear relation between overall SNR and the percentage of correct identifications (linear correlation coefficient $r = 0.98$). A seemingly related result was found in the 3- and 4-talker same-sex configurations of Experiment 1. Specifically, a given SNR led to similar performance for both the 3-talker and 4-talker same-sex configurations (see Fig. 2, second panel, right column). These results suggest that masking signals consisting of two or more same-sex talkers are effectively grouped together into a single same-sex, multiple-talker masking signal. When the overall SNR of the stimulus is less than 0 dB, the impact of this multitalker masking noise on performance is primarily determined by its total power and not by the number or levels of its component talkers.

IV. EXPERIMENT 3: SELECTIVE AND DIVIDED ATTENTION WITH 2, 3, OR 4 COMPETING TALKERS

When listening to multiple competing speech messages, a distinction must be made between divided-attention tasks, where the listeners must simultaneously monitor all the speech channels for pertinent information that might come from any of the competing talkers, and selective-attention tasks, where the listeners know *a priori* which talker they should listen for and they are attempting to focus their attention on the target talker while ignoring the masking talkers (Yost, 1997; Yost *et al.*, 1996; Abouchacra *et al.*, 1997). The CRM-based speech perception test that was used in Experi-

ments 1 and 2 is, in effect, a combination of these two types of tasks: initially, the listeners must divide their attention across all of the competing talkers to determine which one is directly addressing the call sign ''Baron;'' then, they must selectively attend to this target talker in order to extract the color and number coordinates from the target phrase. It is, however, possible to change this CRM listening task into a selective-attention task by providing the listeners with *a priori* information about the vocal characteristics of the target talker. In order to examine the effects of selective and divided attention on multitalker speech perception, a third experiment was conducted that compared performance in a task where the listeners knew the vocal characteristics of the target talker *a priori* (the ''selective-attention'' condition) and a task where the listeners had no knowledge about the identity of the target talker prior to hearing the stimulus (divided attention plus selective attention; the ''divided-attention'' condition).

A. Methods

The procedures used in the third experiment were very similar to the procedures used in the first and second experiments. The target and masker phrases in each trial were randomly selected from the CRM corpus, equalized, electronically summed, and presented to the listeners diotically over headphones. The listeners then identified the color and number coordinates used in the target phrase containing the call sign ''Baron'' by moving the mouse to the appropriate colored number on the CRT of the control computer.

Six paid volunteer listeners participated in the third experiment. All six had previously participated in Experiment 1, and five of the six were also participants in Experiment 2.

A total of nine different target-masker configurations were used in the third experiment: two 2-talker configurations (TS and TD); three 3-talker configurations (TSS, TSD, and TDD); and four 4-talker configurations (TSSS, TSSD, TSDD, and TDDD). In all of these configurations, the overall RMS power of each competing talker was normalized to the same level as the RMS power of the target talker (TMR=0 dB). The target-masker configurations were randomly selected (with replacement) prior to each trial of the experiment. Note that the divided attention conditions replicated a subset of the conditions collected in Experiment 1 when the target-to-masker ratio was 0 dB.

The experiment was divided into two different conditions. In the divided-attention condition, the target talkers in each block of 180 trials were selected randomly from the eight talkers in the corpus. In the selective-attention condition, the same target talker was used in all of the 180 trials in each block, and, prior to beginning data collection in each block, 10 training trials were provided in which only the target talker was presented. This enabled the listeners to become familiar with the characteristics of the target voice. Each listener participated in eight blocks of 180 trials in the selective-attention condition (one for each of the eight talkers in the corpus), and four blocks of 180 trials in the divided-attention condition. The ordering of the two conditions (and of the different talkers in the selective-attention conditions) was randomized across the six different listeners used in the experiment.

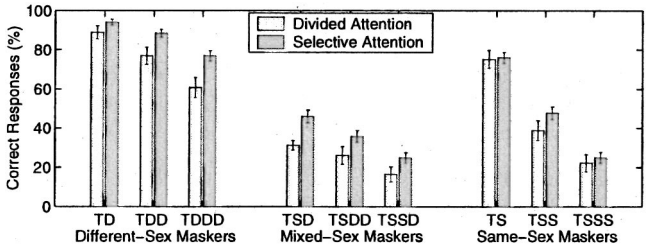

FIG. 5. Percentage of correct color and number identifications in the selective and divided attention conditions of Experiment 3. All data were collected with a TMR of 0 dB. The error bars represent 95% confidence intervals (±1.96 standard errors) in each condition.

B. Results

The results of the experiment are shown in Fig. 5. Each pair of bars compares performance in the divided attention task to performance in the selective attention task for a different target-masker configuration. The data have been grouped separately for configurations with different-sex maskers, configurations with mixed-sex maskers, and the configurations with same-sex maskers.

In general, performance improved when the listeners were provided with *a priori* information about the characteristics of the target voice. A two-factor within-subject ANOVA on the arcsine-transformed individual data for the factors of *a priori* information and target-masker configuration showed that the main effect of *a priori* exposure to the target voice was significant at the $p<0.001$ level ($F_{(1,90)}=32.278$). This improvement was, on average, substantially larger in the different-sex and mixed-sex masking configurations (+12% and +10%, respectively, averaged across all the different-sex and mixed-sex configurations) than in the same-sex masking configurations (+4%). The improvement was also substantially larger in the 3-talker and 4-talker configurations (+12% and +9%, respectively) than in the 2-talker configurations (+4%). The only two voice configurations in which *a priori* information did not produce a substantial improvement in performance were the same-sex 2-talker and 4-talker conditions (TS and TSSS). The relative ineffectiveness of the *a priori* voice information in these same-sex conditions can be explained by a closer examination of the distribution of incorrect responses in the experiment. The data plotted in Fig. 6 show the proportion of trials in which the listeners incorrectly responded with both the color and number coordinates spoken by one of the masking talkers in the experiment. The top panel shows the proportion of responses that matched one of the masking talkers of a different sex than the target talker, and the bottom panel shows the proportion of responses that matched one of the masking talkers of the same sex as the target talker. These data show that the *a priori* voice information provided in the selective attention conditions led to a substantial reduction in the number of different-sex confusions, but had no meaningful effect on the number of same-sex confusions. This result indicates that most of the useful information the listeners were able to obtain from *a priori* exposure to the target talker

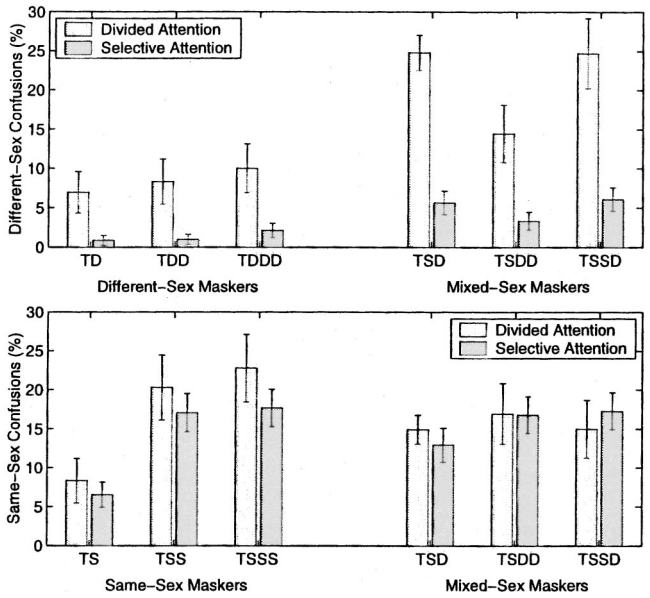

FIG. 6. Same-sex confusions and different-sex confusions in Experiment 3. Same-sex confusions occurred when the listener's responses included both the color and number spoken by one of the same-sex masking talkers in the stimulus. Different-sex confusions occurred when the listeners' responses included both the color and number spoken by one of the different-sex masking talkers in the stimulus. The error bars represent 95% confidence intervals (± 1.96 standard errors) in each condition.

was the sex of the target talker. They were not very good at exploiting the subtle variations in the voices of the same-sex talkers used in the experiment. This would explain why the performance improvements in the selective attention condition were much smaller in the same-sex configurations than in different-sex and mixed-sex configurations.

There is also a relatively simple explanation for why the *a priori* voice information produced less improvement in the 2-talker conditions than in the 3- and 4-talker conditions. When a multitalker speech stimulus contains only two talkers, listeners really only need to attend to one of the two call signs in order to successfully identify the color and number in the target phrase. If they happen to initially focus their attention on the correct talker, they can simply maintain their focus on that talker for the remainder of the carrier phrase. If they happen to initially focus their attention on the wrong talker, then by elimination they know that the target call sign was spoken by the unattended talker and they simply need to switch their attention to the other voice for the remainder of the carrier phrase. Alternatively, listeners may be able to attend to the incorrect voice for the entire stimulus and then, realizing it was the wrong voice, retrieve the coordinates of the correct voice from their short-term auditory memories. In either case, the use of the process of elimination to reduce the 2-talker divided-attention task into the much easier selective-attention task could explain why performance with two talkers was not much different in the selective and divided-attention conditions.

One important difference between the selective- and divided-attention conditions is that the *a priori* voice information provided in the selective-attention condition essentially eliminated the odd-sex distraction that occurred in Experiment 1 and in the divided-attention conditions of Experiment 3. In the divided-attention condition, performance was significantly worse in the TSD configuration than in the TSS configuration, and significantly worse in the TSSD configuration than in the TSSS configuration ($p < 0.05$, two-tailed *t*-tests). In the TSSD configuration, odd-sex distraction was so strong that the listeners were significantly more likely to respond with the color and number spoken by the odd-sex masking talker (25%) than with the color and number spoken in the target phrase (16%) ($p < 0.05$, two-tailed *t*-test). In the selective-attention condition, however, there was no meaningful difference between the TSSS and TSSD configurations, or between the TSS and TSD configurations. Odd-sex distraction is apparently only a factor when the listener is forced to simultaneously monitor all of the competing call signs to find the target call sign, and it does not occur when the listeners know they should ignore the odd-sex talker prior to being exposed to the stimulus. This may indicate that odd-sex distraction occurs because the listeners tend to initially focus their attention on the more salient odd-sex talker in the stimulus, which frequently causes them to miss the information presented in the target phrase. When they know *a priori* that they should not listen to the odd-sex talker, they perform no worse in the odd-sex masking configurations than in the same-sex masking configurations.

The distribution of incorrect responses in Experiment 3 can also be used to evaluate the influence of energetic masking in multitalker listening. Mutual energetic masking between the competing talkers in the stimulus should be greatest when all of the simultaneous talkers are speaking at the same level, as they were in Experiment 3. If this mutual energetic masking were powerful enough to make the competing speech messages inaudible, one would expect the distribution of incorrect responses to be random across all of the different incorrect colors and numbers in the response set. In this experiment, however, the incorrect responses were not randomly distributed: virtually all of the listeners' color and number responses were present in one of the masking phrases in the stimulus. The percentage of responses containing either a color or a number not contained in the stimulus never exceeded 4% in any of the configurations tested in Experiment 3. These percentages are almost negligible in comparison to the percentages of randomly selected responses one would expect to contain a color or number not in the stimulus, which would range from 50% in the 4-talker configurations to more than 85% in the 2-talker configurations. This suggests that at least some, and perhaps all, of the competing voices were audible in the multitalker stimuli of this experiment when the TMR was 0 dB, and that most of the incorrect responses resulted from the listeners' inability to segregate the target phrase from the masking phrases (informational masking) and not from the listeners' inability to hear the target phrase (energetic masking).

V. SUMMARY AND CONCLUSIONS

The results of this study may help to reveal some of the strategies listeners use to process monaural speech signals

with two or more competing talkers. One of the most important findings is that there are fundamental differences in the multitalker speech perception task when the target talker is more intense than the masking talkers and when one or more of the masking talkers is more intense than the target talker.

Target level higher than masker level: When the target-to-masker ratio is positive, performance is generally best when the target voice is qualitatively different than the masking voices. Different-sex maskers degrade performance less than same-sex maskers, and same-sex maskers degrade performance less than same-talker maskers. Performance also generally increases with the number of competing talkers when the overall SNR of the stimulus is fixed.

Target level at or below masker level: When the target-to-masker ratio is negative, performance is much worse when there are three or four competing talkers than when there are only two competing talkers. Overall performance is generally less dependent on target-masker similarity than it is at positive TMRs: performance is still best in the different-sex conditions, but there is little difference between the mixed-sex and same-sex masking conditions, and, in some cases, listeners can be severely distracted by a single masking talker that is different in sex than the other talkers in the stimulus. This odd-sex distraction effect is strongest when all of the competing talkers are presented at the same level, and probably occurs because listeners have a natural tendency to initially focus their attention on the most salient talker in a stimulus. When the maskers are all the same sex as the talker, performance depends almost exclusively on the SNR, and not on the number or individual levels of the masking talkers.

When the SNR of the stimulus is fixed and the listeners are given *a priori* information about the voice characteristics of the target talker, performance generally improves only in mixed-sex and different-sex listening conditions. Most of this improvement results from a reduction in the number of responses matching the different-sex masking voices in the stimulus. The number of same-sex confusions is unaffected by the *a priori* exposure to the target voice. Prior information about the voice characteristics of the target does, however, effectively eliminate odd-sex distraction.

There is also evidence that most of these effects result from informational masking rather than energetic masking. Performance in these experiments was substantially worse in the speech-masking conditions than in the conditions with a corresponding number of modulated noise maskers. In addition, only a small number of the incorrect responses at a TMR of 0 dB contained words which were not present in the stimulus. Although energetic masking certainly played some role at the lowest TMRs used in this experiment, in most cases it appears that performance was most influenced by informational masking. In light of these results, it is important to note that the particular speech test used in these experiments (the CRM) is likely to be substantially more sensitive to informational masking than other types of speech tests that have been used in previous multitalker experiments. At least four factors contribute to this sensitivity: (1) In contrast to most other multitalker speech tasks, which use a target talker that is easily discriminated from the masking talkers on the basis of vocal characteristics (Festen and Plomp, 1990; Drullman and Bronkhorst, 2000), location (Drullman and Bronkhorst, 2000; Hawley *et al.*, 1999; Crispien and Ehrenberg, 1995), or onset time (Freyman *et al.*, 1999), the CRM speech task used in Experiments 1 and 2 of this study uses a randomly selected target talker that is identifiable only by the use of the call sign ''Baron'' in the target phrase. This introduces a substantial amount of uncertainty about the identity of the target talker both in the call sign portion of the carrier phrase and in the color-number portion of the carrier phrase. (2) Although the listeners know the positions of the call sign and coordinate words within the CRM phrases, the phrases themselves contain no contextual information. Thus, all three critical words (call sign, color, and number) must be correctly identified to correctly respond in the CRM talk. This is in direct contrast to natural speech, which contains contextual clues that can be used to reconstruct sentences even when some of the words in the sentences are unintelligible. (3) All of the critical words in the CRM are aligned to occur almost simultaneously, which would almost never happen in natural speech. (4) The key color and number words in the CRM are drawn from small vocabularies of relatively dissimilar words. This allows listeners to correctly guess the right color or number even when they are only able to hear a small portion of the phonetic information in the word. For example, the color ''green'' could be distinguished from the other colors in the CRM corpus either by the initial consonant sound ⟨gr⟩, the vowel ⟨ee⟩, or the final consonant sound ⟨n⟩. Thus, listeners are able to correctly identify the color green from any one of these three phonetic components even if the other two are obscured by energetic masking. This phonetic redundancy makes it possible to measure informational masking effects with the CRM in listening configurations where the target speech would be rendered completely unintelligible by energetic masking in many other speech perception tests. Because the CRM is so highly tuned to measuring informational masking, the results of this experiment should be viewed not as a general indicator of the role informational masking plays in all multitalker speech tasks, but rather as an isolated study of the effects of informational masking in a context where energetic masking is relatively unimportant.

Perhaps the most intriguing aspect of these findings is the fragility of our ability to attend to the quieter of two simultaneous talkers. It is apparent from these results that the mechanisms humans use to focus their attention on the quieter of two talkers are severely disrupted as soon as a third talker becomes audible in the stimulus. Further research is needed to determine how listeners focus their attention on the quieter talker in a 2-talker signal and why they are no longer able to do so when a third talker is added to the stimulus. A thorough understanding of this process might facilitate the development of algorithms for unmixing speech signals at negative SNRs that would be tremendously beneficial in the development of advanced hearing aids and speech recognition systems.

ACKNOWLEDGMENTS

The authors would like to thank Chris Darwin, Adelbert Bronkhorst, and Kim Abouchacra for their helpful comments in the paper. This work was supported by AFOSR Grant No. 01-HE-01-COR.

[1]The other trials were collected early in the experiment and contained invalid configurations including trials where the target talker was used for one of the two masking talkers and trials where more than one talker used the same color or number coordinates. These trials were eliminated from the data analysis.

Abouchacra, K., Tran, T., Besing, J., and Koehnke, J. (**1997**). ''Performance on a selective attention task as a function of stimulus presentation mode,'' Proceedings of the Midwinter Meeting of the Association for Research in Otolaryngology.

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

Bregman, A. S. (**1994**). *Auditory Scene Analysis* (MIT, Cambridge).

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

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

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

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

Carhart, R., Tillman, T., and Greetis, E. (**1969**). ''Perceptual masking in multiple sound backgrounds,'' J. Acoust. Soc. Am. **45**, 694–703.

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

Crispien, K., and Ehrenberg, T. (**1995**). ''Evaluation of the 'Cocktail Party Effect' for multiple speech stimuli within a spatial audio display,'' J. Audio Eng. Soc. **43**, 932–940.

Darwin, C., and Hukin, R. (**2000**). ''Effectiveness of spatial cues, prosody, and talker characteristics in selective attention,'' J. Acoust. Soc. Am. **107**, 970–977.

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

Doll, T., and Hanna, T. (**1997**). ''Directional cueing effects in auditory recognition,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ).

Drullman, R., and Bronkhorst, A. (**2000**). ''Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation,'' J. Acoust. Soc. Am. **107**, 2224–2235.

Egan, J., Carterette, E., and Thwing, E. (**1954**). ''Factors affecting multichannel listening,'' J. Acoust. Soc. Am. **26**, 774–782.

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

Festen, J., 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., Helfer, K., McCall, D., and Clifton, R. (**1999**). ''The role of perceived spatial separation in the unmasking of speech,'' J. Acoust. Soc. Am. **106**, 3578–3587.

Hawley, M., Litovsky, R., and Culling, J. (**2000**). ''The 'cocktail party' effect with four kinds of maskers: Speech, time-reversed speech, speech-shaped noise, or modulated speech-shaped noise,'' Proceedings of the Midwinter Meeting of the Association for Research in Otolaryngology, p. 31.

Hawley, M., Litovsky, R., and Colburn, H. (**1999**). ''Speech intelligibility and localization in a multi-source environment,'' J. Acoust. Soc. Am. **105**, 3436–3448.

Kidd, G. J., Mason, C., Deliwala, P., Woods, W., and Colburn, H. (**1994**). ''Reducing informational masking by sound segregation,'' J. Acoust. Soc. Am. **95**, 3475–3480.

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

Miller, G. (**1947**). ''Sensitivity to changes in the intensity of white Gaussian noise and its relation to masking and loudness,'' J. Acoust. Soc. Am. **191**, 609–619.

Moore, T. (**1981**). ''Voice communication jamming research,'' *AGARD Conference Proceedings 331: Aural Communication in Aviation* (Neuilly-Sur-Seine, France), pp. 2:1–2:6.

Nelson, W. T., Bolia, R. S., Ericson, M. A., and McKinley, R. L. (**1999**). ''Spatial audio displays for speech communication. A comparison of free-field and virtual sources,'' Proceedings of the 43rd Meeting of the Human Factors and Ergonomics Society, pp. 1202–1205.

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. **35**, 1660–1670.

Watson, C., Kelly, W., and Wroton, H. (**1976**). ''Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty,'' J. Acoust. Soc. Am. **60**, 1176–1185.

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

Yost, W., Dye, R., and Sheft, S. (**1996**). ''A simulated 'cocktail party' with up to three sources,'' Percept. Psychophys. **58**, 1026–1036.