Role of masker predictability in the cocktail party problem[a)]

Gary L. Jones
Department of Physiology, University of Wisconsin-Madison, Madison, Wisconsin 53705

Ruth Y. Litovsky[b)]
Waisman Center, University of Wisconsin-Madison, Madison, Wisconsin 53705

(Received 25 October 2007; revised 29 August 2008; accepted 6 September 2008)

In studies of the cocktail party problem, the number and locations of maskers are typically fixed throughout a block of trials, which leaves out uncertainty that exists in real-world environments. The current experiments examined whether there is (1) improved speech intelligibility and (2) increased spatial release from masking (SRM), as predictability of the number/locations of speech maskers is increased. In the first experiment, subjects identified a target word presented at a fixed level in the presence of 0, 1, or 2 maskers as predictability of the masker configuration ranged from 10% to 80%. The second experiment examined speech reception thresholds and SRM as (a) predictability of the masker configuration is increased from 20% to 80% and/or (b) the complexity of the listening environment is decreased. In the third experiment, predictability of the masker configuration was increased from 20% up to 100% while minimizing the onset delay between maskers and the target. All experiments showed no effect of predictability of the masker configuration on speech intelligibility or SRM. These results suggest that knowing the number and location(s) of maskers may not necessarily contribute significantly to solving the cocktail party problem, at least not when the location of the target is known.
[DOI: 10.1121/1.2996336]

PACS number(s): 43.66.Lj, 43.66.Pn, 43.71.Gv [RLF] Pages: 3818–3830

I. INTRODUCTION

Cherry (1953) formulated the "cocktail party problem" in terms of how we understand what one person is saying when others are talking at the same time. An important feature of such multitalker environments is that they can be highly unpredictable. While a listener is attending to a target talker, many features of the competing source(s) can vary, including their number, location(s), and content, to name just a few. With so many sources of variability, intuition would suggest that listener expectations about maskers could play an important role in solving the cocktail party problem. However, studies of the cocktail party problem typically use paradigms in which the number and locations of maskers are constant throughout a block of trials, thereby ensuring high predictability of the multitalker environment. There has been little research that specifically addresses how masker predictability affects speech intelligibility in a cocktail party environment.

As noted by Cherry (1953), listeners can make use of differences in the spatial locations of sources to help separate a target signal from competing sources. The decrease in thresholds that results from such spatial separation of target from interferers is known as spatial release from masking (SRM); it is a robust phenomenon that is observed even when azimuthal localization cues are limited to just interaural level differences or just interaural timing differences (ITDs) (Bronkhorst and Plomp, 1988; Culling *et al.*, 2004). Durlach *et al.* (2003a) proposed two different strategies that could enable listeners to extract a target signal from competing sounds: a "max" strategy with enhanced sensitivity to sources from the target location and a "min" strategy in which sensitivity to sources from the masker location(s) is suppressed. It has been shown that knowing the location of the target can improve performance in identification of both speech and nonspeech targets (Arbogast and Kidd, 2000; Ericson *et al.*, 2004; Kidd *et al.*, 2005), a finding consistent with the max strategy. It is important to explore whether responses consistent with the min strategy also contribute significantly to speech intelligibility in multitalker environments.

Masking of a speech target by speech maskers differs in several ways from masking by noise maskers, and this is well illustrated by SRM data. Good examples of this can be seen by considering spatial configurations in which all maskers are at a single location in the horizontal plane 90° away from the target. First, the amount of SRM depends strongly on the number of competing sounds when the maskers are speech or reversed speech but not when the maskers are noise (Peissig and Kollmeier, 1997; Bronkhorst, 2000; Culling *et al.*, 2004; Hawley *et al.*, 2004). Second, when two or three maskers are present, greater SRM is observed with speech maskers than with noise maskers (Culling *et al.*, 2004; Hawley *et al.*, 2004). Third, the mere *perception* of spatial separation has differential effects for speech and noise maskers. The perceived position of a masker that is actually colocated with the target can be displaced by presenting, 60° away from the target, a duplicate

---

[a)] Portions of this work were presented in "Role of masker predictability in the cocktail party effect" at the Midwinter Meeting of the Association for Research in Otolaryngology, New Orleans, LA, February 2005.

[b)] Electronic mail: litovsky@waisman.wisc.edu of the masker that is time shifted, so it leads the other copy by several milliseconds. This technique takes advantage of the precedence effect, whereby leading sources dominate the perceived locations of paired sounds, and it results in significant SRM with speech maskers but not with noise maskers (Freyman *et al.*, 1999, 2001).[1]

This last example, in which for speech maskers there can be sizable SRM despite *increasing* the overall masker level by 3 dB, illustrates the concept of "informational masking." At present any discussion of this type of masking necessarily involves definitional issues, and there are ongoing efforts to improve existing definitions of informational masking [e.g., see Durlach *et al.* (2003a)]. Informational masking is commonly equated with any masking beyond the "energetic" component of masking, which results from overlapping excitation at the auditory periphery. Thus, it is similar to what was earlier described as "excess" or "perceptual" masking (Carhart *et al.*, 1969). Studies with both speech and nonspeech stimuli have shown that spatial separation of target and maskers is effective at combating informational masking (Kidd *et al.*, 1998; Freyman *et al.*, 1999). Thus, measures of SRM can be particularly useful for identifying conditions in which informational masking is present.

There is a long tradition linking informational masking to stimulus uncertainty, but there has been a recent trend to emphasize the role of similarity between target and masker(s) or even to base the very definition of informational masking on similarity (Watson *et al.*, 1976; Durlach *et al.*, 2003b; Brungart and Simpson, 2004). Generally, however, uncertainty and similarity are both believed to be sources of informational masking (Oh and Lutfi, 1998; Durlach *et al.*, 2003b; Watson, 2005). For the purposes of these experiments, informational masking is defined as any masking beyond energetic masking. Reductions in masker uncertainty can, we believe, reduce the informational component of masking.

Evidence of masking due to masker uncertainty has been found for both nonspeech and speech targets. For example, thresholds for detecting a target tone in the presence of a multicomponent nonspeech masker are elevated under conditions of high uncertainty about the spectral content of the masker (Lutfi, 1993; Oh and Lutfi, 1998). In addition, Fan *et al.* (2008) reported that *spatial* uncertainty of maskers leads to degraded detection of nonspeech stimuli. As regards speech targets, there is evidence of reduced speech intelligibility under conditions of uncertainty about masker content, but the effects reported in the literature have been rather small (Brungart and Simpson, 2004; Freyman *et al.*, 2007). As concerns uncertainty about location, it has been shown that task performance is reduced when there is trial-to-trial uncertainty about the spatial locations of *both* targets and maskers (Shinn-Cunningham and Ihlefeld, 2004; Kidd *et al.*, 2005; Brungart and Simpson, 2007a). An important next step is to isolate the effects, if any, of uncertainty about the number and locations of maskers.

This article presents the results of three experiments that examined speech intelligibility and SRM when listeners were attending to a speech target coming from a known location while the number and locations of competing sounds could vary from trial to trial. Uncertainty about the number or locations of maskers could be a source of informational masking. Thus, decreases in uncertainty could lead to improved speech intelligibility. Furthermore, one might expect that high predictability in the number and locations of maskers would provide listeners with information about the auditory environment, which would allow them to utilize spatial cues more effectively, resulting in greater SRM. Thus, it was hypothesized that as predictability of the number and/or locations of speech maskers is increased, there will be (1) improved speech intelligibility and (2) increased SRM.[2] The first experiment reported here tested whether percent correct scores and SRM at a fixed signal-to-noise ratio (SNR) increase as the predictability of the masker configuration is stepped up from 10% to 80%. The second experiment examined whether speech reception thresholds (SRTs) improve (decrease) and SRM increases when (a) predictability of the masker configuration is increased from 20% to 80% and/or (b) when the complexity of the listening environment is decreased. The third experiment tested whether SRTs decrease and SRM increases when predictability of the masker configuration is increased from 20% up to 100% while minimizing the onset delay between maskers and the target. These experiments did not support the hypothesis; that is, there was no effect of predictability of the number or locations of maskers on speech intelligibility or SRM.

II. EXPERIMENT 1: SPEECH INTELLIGIBILITY AT A FIXED SNR

A. Method

1. Listeners

Participants in the experiment were eight paid university students (eight females), 18–19 years of age. Some subjects had participated in a pilot study for these experiments. Only native English speakers from households in which no language other than English was spoken were recruited. Due to the effects of musical training on performance in informational masking experiments (Oxenham *et al.*, 2003), only subjects who did not have extensive formal musical training were eligible to participate. Audiograms were performed with each potential subject, and pure tone thresholds of 20 dB HL or better at the octave frequencies from 250 to 8000 Hz were required. One potential subject was excluded from this experiment on the basis of tone thresholds.

2. Stimuli

Stimuli for this experiment were recorded by a male talker, trained to speak at constant levels and rates. The same voice was used for both targets and maskers in order to facilitate uncertainty-based masking effects, if any, through high target-masker similarity [e.g., see Durlach *et al.* (2003b)]. The targets were from a closed set of 40 spondees with equivalent intelligibility as determined by pilot testing. This stimulus corpus was chosen because it can be used for repeated testing on numerous conditions. The maskers were concatenated recordings of Harvard IEEE sentences (Rothauser *et al.*, 1969). Thus, in this experiment a relatively brief spondee target could be presented with a substantial onset

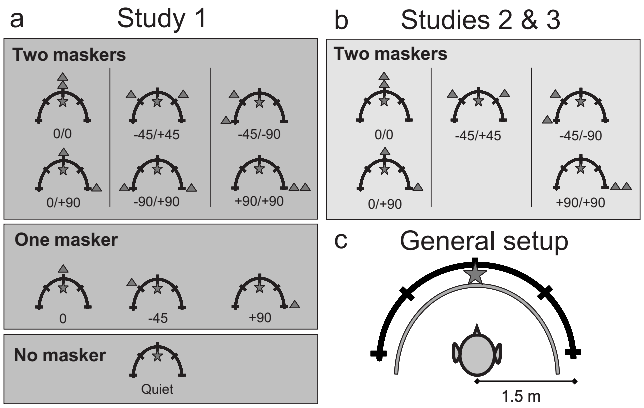

FIG. 1. Schematic view of the masker configurations used in these experiments. (a) The ten masker configurations with two maskers, one masker, or no masker shown at the left were used in experiment 1. (b) The five two-masker configurations shown at the right were used in experiments 2 and 3. The target (shown by a star) was presented from 0° in all experiments. In each diagram, the triangles indicate masker positions. A schematic view of the setup used in these experiments is shown in (c). Subjects were seated at the center of a 1.5 m loudspeaker array with loudspeakers located every 45° from −90° to +90°. The positions of the loudspeakers were concealed by a curtain.

delay in the midst of a lengthy masker. This is similar to an approach in which Freyman *et al.* (1999) used an onset delay to provide a basis for the subject to attend to the target in conditions in which target and masker were presented from the same location. In many experiments that measure masking of speech targets by speech interferers a large proportion of errors result from the subject misidentifying an interferer as the target [e.g., see Brungart (2001)]. One motivation for using spondees as targets, as opposed to sentences from the same corpus as the maskers, was to minimize such errors based on confusing the *content* of the target with the content of the maskers.

The maskers were edited in two ways to limit opportunities for "listening in the gaps" between individual words or sentences in the masker. First, silent gaps between consecutive words were edited out manually. Second, both ends of each recorded sentence were edited using an iterative automated algorithm. The algorithm works from each end of the sentence and removes 10 ms segments until it finds a segment whose root mean square (rms) amplitude is within 12 dB of the rms of the whole sentence, at which point it identifies a zero crossing near the end of the sentence and makes the final cut there. The automated editing caused six of the recorded sentences to sound unnatural, and these six sentences were discarded. The remaining sentences were concatenated to create three-sentence maskers with a mean duration of 5.4 ± 0.1 s. Some sentences occurred in multiple maskers, but no sentence was repeated in any masker and two maskers containing the same sentence were never presented on the same trial. The resulting maskers had a rapid cadence of about 250 words/min but were judged by both experimenters and by the listeners to be natural and were readily understandable. Although the maskers still had considerable amplitude modulation, the editing of the recordings could be expected to reduce opportunities for listening in the gaps.

3. Procedures

A schematic overview of the experimental setup is shown in Fig. 1(c). The experiments were conducted in a small room ($2.9\ \text{m} \times 3.6\ \text{m}$) with reverberation time (RT_{60}) of approximately 250 ms. The subject was seated at the center of a hemispheric loudspeaker array with a radius of 1.5 m. Subjects were seated on a chair on a platform that could be raised or lowered so that the opening of the ear canal was within 4 cm of the horizontal plane. The platform, walls, and ceiling were covered by 8 cm acoustically absorbent foam. The positions of the loudspeakers were concealed from the subjects by a visually opaque, acoustically transparent curtain. The experiment used five loudspeakers at 0° elevation spaced every 45° from −90° (left) through 0° (straight ahead) to +90° (right).

Software for stimulus presentation and data collection was written in MATLAB (Mathworks Inc.). Target and masker stimuli, which had been recorded from a trained male talker, low-pass filtered at 10 kHz and saved at a sampling rate of 44.1 kHz, were upsampled, summed (where necessary), and played at a sampling rate of 48.8 kHz using 16 bit digital-to-analog conversion. Speaker switching and amplification were controlled through Tucker Davis Technologies (TDT) System III hardware (RP2, PM2, SA1) in conjunction with a PC host. The overall masker level was 57 dB(A) on all masked trials. On two-masker trials, the levels of the two maskers were equal to one another. Target words were presented at a fixed level of 38 dB(A). The −19 dB SNR was chosen based on piloting to avoid floor effects in the most difficult masker configurations and ceiling effects in the least difficult configurations. A cosine-squared window with 2 ms rise and fall times was applied to each stimulus; this was done prior to upsampling to reduce resampling distortion. Target words were presented with a 2 s delay relative to the onset of the masker(s). The onset of the target word always occurred after the start of the second sentence in the masker, and no part of the target word ever overlapped with the transition from one sentence to the next.

The 40 spondee targets were displayed in alphabetical order, reading down the columns of an 8×5 grid, and subjects were asked to verbally identify the target word on each trial using words from the list in a 40-alternative forced choice task (40AFC). They were instructed to guess if they could not identify the target word. Responses were entered into a computer by a tester who was monitoring the experiment via closed-circuit TV. Subjects were informed that the target would always be presented from straight ahead, and the 0° location was visually marked on the curtain. No other information was given about the positions of the loudspeakers. Subjects were instructed to face straight ahead, with the nose pointing at the 0° loudspeaker, and they were monitored throughout each testing session to ensure that they maintained proper head position.

Ten masker configurations were used in this experiment. As shown in Fig. 1(a), there were six two-masker configurations, three one-masker configurations, and one no-masker configuration. The two-masker configurations were of three types: (1) one or both maskers at the target location, (2) symmetrical arrays with maskers bilaterally distributed about the target, and (3) both maskers separate from the target and in a single hemifield. In the three one-masker configurations, separation between target and masker was 0°, 45°, or 90°. All masker configurations occurred in all blocks of trials, and the order of masker configurations within a block was randomized. Thus, on any given trial, the target word could be masked by 0, 1, or 2 maskers.

The experiment, consisting of 3600 trials per subject, was organized into 34 blocks of trials divided over seven 2 h testing sessions in an incomplete Latin square design. A within-subject design was used in which each subject participated in all blocks of trials. In "equal-probability" blocks of trials, each of the ten masker configurations occurred on 10% of the trials. In order to test the effect of predictability of the masker configuration on speech intelligibility and release from masking, there were also blocks of trials in which one of the masker configurations occurred with greater than 10% frequency. In these "unequal-probability" blocks, one of the masker configurations occurred on 20%, 40%, 60%, or 80% of the trials, and the remaining trials were divided equally among the other nine masker configurations. In all, 520 trials were needed to test a single masker configuration in all four types of unequal-probability blocks; thus, it was not practical to test all possible combinations of the ten masker configurations and the five levels of predictability. Rather, six of the ten masker configurations were tested at all levels of predictability, including two one-masker configurations (0 and +90) and four two-masker configurations (0/0, 0/+90, −90/+90, and +90/+90).

In order to familiarize subjects with the task, there was a 10 min practice session at the start of the first testing session and a 5 min practice session at the start of each subsequent testing session. Within each block, the first 20 trials were used to familiarize the subjects with the general frequency of occurrence of the masker configurations; e.g., in an 80% block, the tested configuration occurred on 16 of the 20 familiarization trials. The familiarization trials were not included in the data analysis. The number of trials in each block was selected so as to have 40 analyzed trials of the tested masker configuration. For example, each 80% block was 70 trials long, with 20 familiarization trials followed by 50 experimental trials.[3] Specific features of the incomplete Latin square design include the following: there could be at most one equal-probability block, one 40% block, one 60% block, and one 80% block during a single testing session, and any given masker configuration could be tested at greater than 10% predictability only once during a single testing session. The twelve 20% blocks were divided over the seven testing sessions as equally as possible.

Blocks were limited to no more than 120 trials, including familiarization trials, to avoid possible effects of fatigue on performance. Thus, in order to have 40 analyzed trials in the equal-probability blocks and the 20% blocks, there were four equal-probability blocks and there were two separate 20% blocks for each tested masker configuration. Although subjects could not leave the testing booth during the course of a block of trials, they were given a brief break of 30–60 s every 6–7 min during a block and longer breaks between blocks.

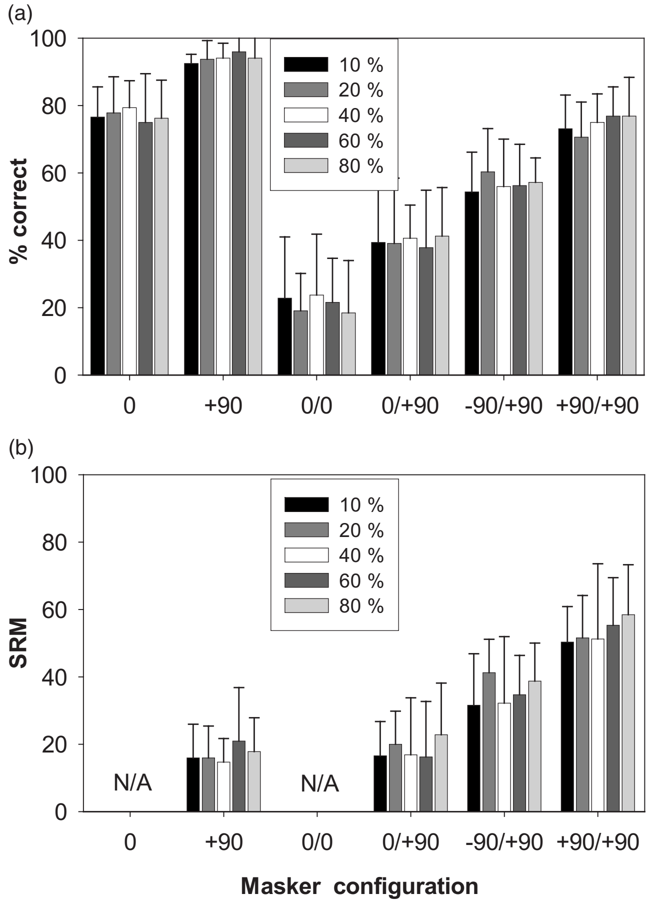

FIG. 2. Group means (±SD) for the six masker configurations in experiment 1 that were tested at several levels of predictability are shown as (a) percent correct and (b) SRM.

4. Data analysis

Percent correct scores were calculated for each of six masker configurations at 10%, 20%, 40%, 60%, and 80% predictability. Within each level of predictability, SRM was calculated as the difference in percent correct between each masker configuration with masker(s) away from 0° and the corresponding configuration with masker(s) at the target location. Two-way analyses of variance (ANOVAs) were performed on percent correct and SRM, with masker configuration and predictability as factors. Following each ANOVA, *post hoc* comparisons were made between different levels within each factor via Scheffe analysis (criterion: $p < 0.05$).

B. Results

Group means (±SD) for the masker configurations that were tested at multiple levels of predictability are shown in Fig. 2 as percent correct [Fig. 2(a)] and as SRM [Fig. 2(b)]. There was a significant main effect of masker configuration [$F(5,35) = 151.49$; $p < 0.0001$] on percent correct scores, but no main effect of predictability and no interaction [Fig. 2(a)]. Scheffe analysis of percent correct scores collapsed across predictability showed that all pairs of masker configurations, except 0 and +90/+90, were significantly different from each other. For similar location configurations, percent correct scores declined as the number of maskers was increased, with greater masking for two maskers (0/0 and +90/+90) than for one (0 and +90), despite the overall level equivalency $[F(5,7)=59.96;\ p<0.0001]$.

Mean SRM values were positive for all tested masker configurations [Fig. 2(b)]. There was a significant main effect of masker configuration $[F(3,21)=77.51;\ p<0.0001]$ on SRM, but there was no main effect of predictability and no interaction. Scheffe analysis of SRM data collapsed across predictability showed that all masker configurations were significantly different from each other, except +90 and 0/+90.

C. Discussion

These results showed greater masking with two maskers than with one and significant SRM for all masker configurations in which at least one masker was spatially separated from the target. Within the two-masker configurations, SRM was lowest for the 0/+90 configuration and highest for the +90/+90 configuration. This is broadly consistent with previous results with noise maskers (Bronkhorst, 2000) and particularly with speech maskers (Hawley *et al.*, 2004). Contrary to the hypothesis of higher percent correct scores and greater SRM with increasing predictability, these results showed no effect of predictability on either measure. Although these data suggest that predictability of the number or locations of maskers does not influence speech intelligibility or SRM, alternative explanations also need to be explored.

First, there are inherent limitations of the fixed SNR approach used in this experiment. Pilot testing had suggested that the one- and two-masker configurations would not be subject to floor or ceiling effects at the target level used in this experiment; however, a ceiling effect in the +90 configuration and a floor effect in the 0/0 configuration are apparent in the individual data (not shown). While it is worth noting that in Fig. 2(a) there is also no indication of an effect of predictability on speech intelligibility in the other four masker configurations (0, 0/+90, −90/+90, +90/+90), the findings of experiment 1 would certainly be strengthened by eliminating floor and ceiling effects.

Second, the paradigm used in this experiment may not have fully exploited differences in stimulus uncertainty. Although the predictability of the tested masker configuration covered a large range, one invariant in this experiment was that ten different masker configurations occurred in each block. It is possible that speech intelligibility and SRM are affected by the complexity of the entire listening environment and not just by the frequency of occurrence of any particular masker configuration. This can be tested by limiting the number of loudspeaker locations or the number of different masker configurations that occur in a block of trials. Furthermore, it is possible that the familiarization period (20 trials) was not long enough for subjects to recognize and adjust to the predictability of the masker configuration.

Third, learning effects played a role in these results. After factoring out differences in baseline performance between masker configurations, there was a significant correlation $(0.37, p<0.0001)$ between performance in each block and which testing session that block occurred in. For all of these reasons, additional experiments were undertaken to further evaluate the original hypothesis. In order to limit the role of learning effects, an experiment with fewer testing sessions was developed.

III. EXPERIMENT 2: VARIABLE SNR, EFFECT OF PREDICTABILITY AND ARRAY COMPLEXITY

A. Method

1. Listeners

Listeners were nine university students (eight females and one male), 19–30 years of age, who were paid for their participation. None had taken part in the previous experiment or pilot testing. Requirements for participation were the same as for the previous experiment.

2. Stimuli

Target words and masker sentences were the same as in experiment 1.

3. Procedures

Several changes were made to the procedures to address the issues raised in the discussion of experiment 1. First, in order to measure SRM in decibel units and to eliminate floor and ceiling effects, SRTs were determined by using an adaptive procedure with four reversals. This paradigm has been used reliably with both pediatric and adult populations (Litovsky, 2005; Johnstone and Litovsky, 2006; Garadat and Litovsky, 2007). The data were then fitted to a logistic function by a constrained maximum-likelihood estimation (MLE) method (Wichmann and Hill, 2001a, 2001b). The data fitting yields an estimate of the underlying psychometric function, which makes it possible to analyze the effect of predictability on both threshold and slope. Second, in order to give subjects more time to recognize and adjust to the predictability of the masker configuration than in experiment 1, the familiarization period was increased from 20 trials to 100 trials.

As in experiment 1, the overall masker level was 57 dB(A). SRTs were estimated using a method with slight modifications to an adaptive algorithm that has been described previously (Litovsky, 2005). An adaptive tracking method was used to vary the level of the target signal from an initial SNR of +6 dB, such that correct responses result in level decrement and incorrect responses result in level increment. The algorithm includes the following rules: (1) level is initially reduced in steps of 8 dB following each correct response. (2) Following the first incorrect response, a 3-down/1-up rule is used, whereby level is decremented following three consecutive correct responses and incremented following a single incorrect response. (3) Following each reversal, the step size is halved. (4) A step size that has been used twice in a row in the same direction is doubled. For instance, if the level was decreased from 40 to 36 dB (step size =4 dB) and then again from 36 to 32 dB, a set of three consecutive correct responses at 32 dB would result in an 8 dB drop to 24 dB. (5) Testing is terminated following four reversals. The 3-down/1-up method converges to the 79.4% correct point (Levitt, 1971). Masker unpredictability was generated by presenting adaptive tracks for two or more masker configurations in interleaved fashion. In each block of trials, thresholds for all tested masker configurations were obtained concurrently [e.g., see Leek (2001)].

Data were collected with the 40-AFC task used in experiment 1. The lower bound of the psychometric function was set to the chance level of performance, 0.025. The sampling scheme and lapses in listener attention can introduce biased estimates of threshold. The bias introduced by attentional lapses can be overcome by confining the upper bound of the psychometric function to within a narrow range of $p(c)=1$ (Wichmann and Hill, 2001b). In these experiments a constrained MLE fitting procedure was used in which the lambda parameter, which determines the upper asymptote of the psychometric function, was confined to the range [0, 0.05]. Threshold was calculated as the 80% point on the fitted psychometric function as described previously (Litovsky, 2005; Johnstone and Litovsky, 2006; Garadat and Litovsky, 2007).

Five masker configurations were tested in experiment 2 as opposed to ten in experiment 1 [see Fig. 1(b)]. Without this change, the extended familiarization period and the use of adaptive tracking methods in experiment 2 would have greatly increased the size of the test blocks. In addition, the decision to test fewer masking configurations reduced the number of testing sessions and thus limited opportunities for learning effects. As a result of these choices, experiment 2 examined possible effects of predictability of the *locations* of maskers only. The *number* of maskers was 2 on every trial.

In each block of trials, there were up to five parallel adaptive tracks. These adaptive tracks were presented in interleaved fashion through the use of custom software written in MATLAB, which allowed the user to specify how frequently each masker configuration would occur within a given block of trials. In equal-probability blocks, each of the five masker configurations in Fig. 1(b) occurred on 20% of trials, and in unequal-probability blocks the tested masker configuration, either 0/0 or +90/+90, occurred on 80% of trials. There were two types of unequal-probability blocks in this experiment. In blocks with the "full array," the tested masker configuration occurred on 80% of trials, and the remaining 20% of trials were divided equally among the other four masker configurations. In blocks with a "simple array," the 0/0 and +90/+90 configurations occurred in an 80/20 or 20/80 ratio. There were five blocks of trials divided over three testing sessions of up to 2 h each. The order of blocks was balanced across subjects in an incomplete Latin square design.

In order to minimize learning effects, subjects completed a practice session during an initial visit for hearing screening. Collection of analyzed data began with the second visit. Due to the extended familiarization period in this experiment, the blocks of trials were quite long, and it was important to avoid fatigue effects on performance. Subjects did not leave the testing booth while a block was underway, but they were given a brief break of 30–60 s every 6–7 min and a break of about 2 min every 15 min. There was a 5 min practice session before the start of each testing session.

4. Data analysis

SRTs for the 0/0 and +90/+90 configurations were calculated at 20% and 80% predictability for both the full array and the simple array. During a block of trials, subjects would often complete multiple adaptive tracks of a masker configuration (e.g., the configuration tested at 80% predictability in blocks with the simple array). In such cases the reported threshold is the average of the thresholds from all completed adaptive tracks of that configuration after the familiarization period.

Within each combination of predictability and array complexity, SRM for the +90/+90 configuration was calculated as SRT for the 0/0 configuration minus SRT for the +90/+90 configuration. Three-way ANOVAs were computed for SRTs and for the slopes of the psychometric functions of the 0/0 and +90/+90 masker configurations with masker configuration, predictability, and array complexity as factors. A two-way ANOVA was calculated on SRM values for the +90/+90 masker configuration with predictability and array complexity as factors.

B. Results

Group means (±SD) for the +90/+90 configuration across predictability and across loudspeaker array are shown in Fig. 3 as SRT [Fig. 3(a)] and SRM [Fig. 3(b)]. There is very little difference in SRT or SRM with changes in either predictability of the masker configuration or the complexity of the loudspeaker array. A three-way ANOVA on SRT data showed a significant main effect of masker configuration [$F(1,8)=1604.92; p<0.0001$], but there was no effect of either predictability or the complexity of the loudspeaker array, and there were no interactions among the factors. A three-way ANOVA on the slopes of the psychometric functions at threshold showed no significant main effects or interactions. A two-way ANOVA on SRM data for the +90/+90 configuration across predictability and complexity of the loudspeaker array showed no significant main effects and no interaction. Consistent with the SRM data from experiment 1, the data for all five masker configurations in the equal-probability blocks (data not shown) exhibit increasing SRM as (1) one masker was placed away from the target, (2) the two maskers were bilaterally distributed about the target, and (3) both maskers were displaced from the target and confined to a single hemifield.

C. Discussion

This second experiment was motivated primarily by concerns that the methodology of the first experiment may not have fully exploited differences in predictability. The three major changes from experiment 1 to experiment 2 were as follows: the length of the familiarization period was increased, an adaptive procedure was used to avoid the floor and ceiling effects seen in experiment 1, and possible effects of the complexity of the loudspeaker array on SRT and SRM were also tested. Despite these changes, the basic result of

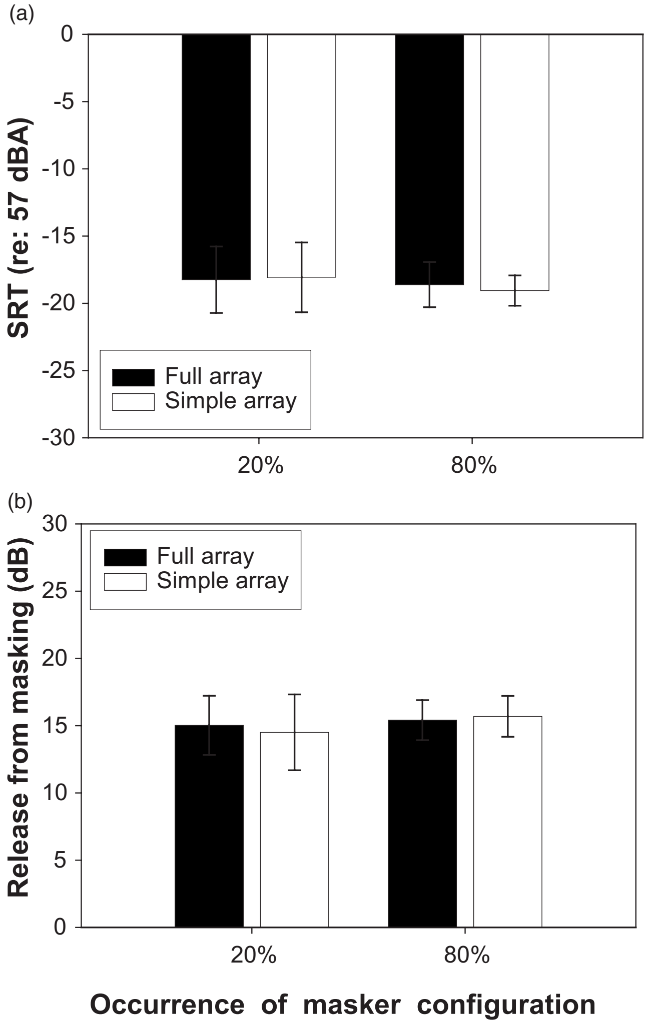

FIG. 3. Group means (±SD) for the +90/+90 masker configuration in experiment 2 at two levels of predictability and two levels of complexity of the loudspeaker array are shown as (a) SRTs and (b) SRM.

experiment 2 is the same as the result of experiment 1; namely, changes in the predictability of masker locations had no effect on speech intelligibility or SRM. Moreover, the additional tests that were added in experiment 2 also offered no evidence to support the hypothesis. That is, there was no effect of predictability on the slope of the psychometric function at threshold, and simplifying the loudspeaker array had no effect on SRT, SRM, or slope.

These results suggest that in a cocktail party environment, there may not necessarily be added masking due to uncertainty about the locations of maskers. This is consistent with the findings of experiment 1. However, it is possible that even the modified paradigm used in experiment 2 did not fully exploit differences in the predictability of the masker configuration. First, 80% predictability may not be high enough to eliminate informational masking effects of uncertainty about the masker configuration. Second, the onset delay between maskers and the target provided listeners with an opportunity to focus on the masker configuration prior to target onset and may have allowed them to overcome any effect of uncertainty. Thus, one further experiment was conducted to test for effects of predictability of masker locations on speech intelligibility and SRM.

IV. EXPERIMENT 3: EFFECT OF 100% CERTAINTY WITH MINIMAL ONSET DELAY

A. Method

1. Listeners

Ten university students (ten females), 18–22 years of age, participated in this experiment. Subjects were paid for their participation, and none had taken part in either of the previous experiments or pilot testing. Requirements for participation were the same as for the previous experiments.

2. Stimuli

Maskers with a mean duration of 3.6±0.0 s were created by concatenating recordings of pairs of Harvard IEEE sentences.[4] These two-sentence maskers were identical to the three-sentence maskers used in the previous two experiments, but with the third sentence removed. The targets were the same 40 recorded spondees as in the previous two experiments.

3. Procedures

Before the start of each trial, the subject was prompted with the word "Ready?" from the 0° loudspeaker. The prompt was recorded by the same talker as the targets and maskers, and there was a silence of 0.6±0.1 s between the prompt and masker onset. The delay between the onset of the maskers and the onset of the target was reduced from 2 s in experiments 1 and 2 to a variable delay of 0–0.2 s in experiment 3. There were no familiarization trials in this experiment; rather, subjects were told in advance whether the masker locations in the upcoming block of trials would be fixed or variable.

The five two-masker configurations shown in Fig. 1(b) were included in this experiment. In equal-probability blocks, there were five interleaved adaptive tracks, and each of the five combinations of masker locations occurred on 20% of the trials. In *fixed* blocks, there was a single adaptive track, and the same masker configuration occurred on 100% of trials. The equal-probability blocks were run until the subject had completed at least one adaptive track for each of the five masker configurations. Each fixed block was run until the subject had completed a single adaptive track. There were two repetitions of each block, always in separate testing sessions, and the order of blocks was balanced across subjects in an incomplete Latin square design. Overall, there were 12 blocks of trials, and these blocks were run concurrently with an experiment from a separate study not reported here over three approximately 2 h testing sessions.

There was a 5 min practice session before the start of each testing session. Subjects did not leave the testing booth during a block of trials. During equal-probability blocks, there was a short break of 30–60 s every 6–7 min and a longer break of about 2 min every 15 min. No breaks were given during fixed blocks, which only lasted a few minutes. Subjects were familiarized with the task during an initial visit for hearing screening.

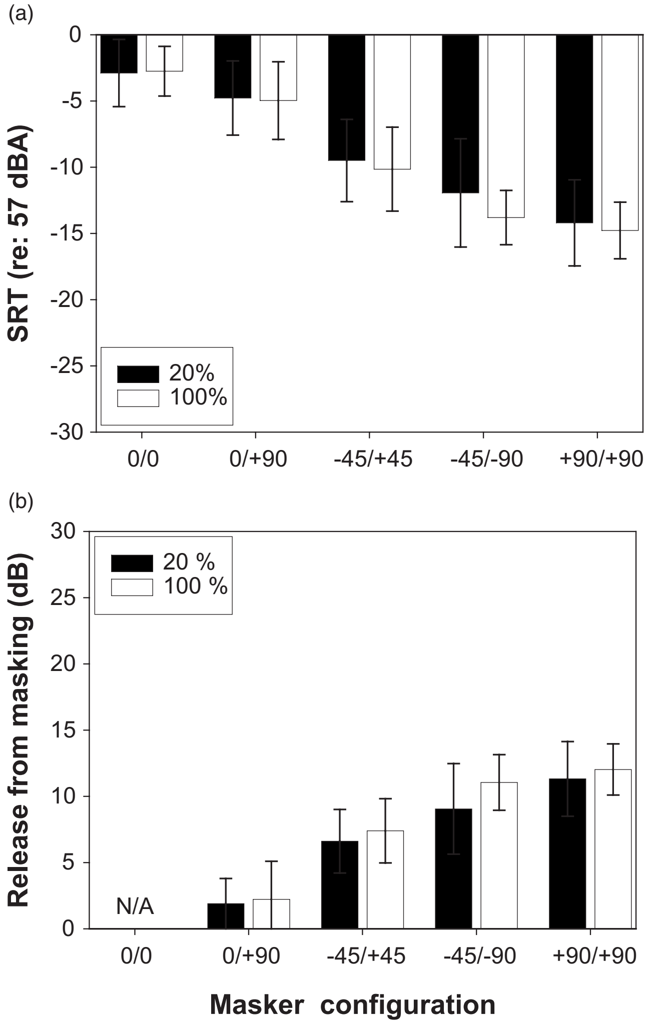

FIG. 4. All five masker configurations in experiment 3 were tested at 20% and 100% predictability. The bar plots above show group mean (±SD) (a) SRTs and (b) SRM by masker configuration and predictability.

4. Data analysis

Every threshold reported for each subject is the average of SRTs from completed adaptive tracks for a given combination of masker configuration and predictability. SRTs were calculated as in the previous experiment. Two-way ANOVAs were performed on SRT, SRM, and the slopes of the psychometric functions, with masker configuration and predictability as factors. *Post hoc* comparisons were made among the masker configurations via Scheffe analysis.

B. Results

Group means (±SD) across masker configuration and predictability are shown in Fig. 4 as SRTs [Fig. 4(a)] and SRM [Fig. 4(b)]. There are visible differences in the plots between the data at 20% predictability and 100% predictability. For example, all SRM values in Fig. 4(b) are somewhat lower at 20% predictability than at 100% predictability. Moreover, the standard deviations in Figs. 4(a) and 4(b) are considerably larger at 20% predictability than at 100% predictability for the two masker configurations in which both maskers were in the same hemifield, particularly for the −45/−90 configuration.

Statistical analysis showed a significant main effect of masker configuration on both SRT [$F(4,36)=125.76$; $p<0.0001$] and SRM [$F(3,27)=76.23$; $p<0.0001$], but there was no significant effect of predictability and there was no interaction. A *post hoc* Scheffe analysis in which the SRT data were collapsed across predictability showed that all masker configurations were significantly different from each other with two exceptions: (1) 0/0 was not significantly different from 0/+90, and (2) the same-hemifield masker configurations, −45/−90 and +90/+90, were not significantly different from each other. There was no effect of masker configuration or predictability on slope, and there was no interaction.

C. Discussion

In experiment 3, as in the previous experiments, there was no effect of predictability on SRTs, SRM, or the slopes of the psychometric functions. Although SRM at 100% predictability was, on average, 1 dB higher than at 20% predictability [see Fig. 4(b)] this difference did not approach significance. Furthermore, a masking effect on the order of 1 dB would probably be too small to be of interest or of functional relevance.

As a follow-up, these data were examined for other indications of informational masking such as high individual variability. Such individual variability would be consistent with the use of different min and max strategies (Durlach *et al.*, 2003a), as described in the Introduction. For example, do most subjects have similar SRM at both 20% and 100% predictability, or is this highly variable? Also, is there a connection between spatial separation of maskers from the target and the relative amount of SRM at 20% and 100% predictability? The relationship between SRM values at 20% and 100% predictability is shown for all ten subjects in Fig. 5. The data in Fig. 5 indicate that for the majority of subjects, SRM is highly similar at 20% and 100% predictability. Overall, the individual variability in these data is not substantially greater along one axis than the other, and the differences observed between the group mean SRM values at 20% and 100% predictability are due to the results of just two subjects, SHY and SID. In addition, there is no evidence in Fig. 5 of a connection between spatial separation and the relative amount of SRM at 20% and 100% predictability.

The relative sizes of the error bars in the two same-hemifield configurations are intriguing. The fact that the error bars are larger at 20% predictability would be consistent with greater informational masking at 20% predictability than at 100% predictability. It is worth noting that this same pattern is seen in the experiment 2 data in Fig. 3(b), particularly for the simple loudspeaker array. Moreover, it seems plausible that the min listening strategy described above would be most effective when the spatial locations to be suppressed are near each other and in the same hemifield. Note, however, that the 2–2.5 dB standard deviations of the group means in experiments 2 and 3 do not fit the high individual variability seen with informational masking [e.g., see Durlach *et al.* (2003b)]. Thus, the relative sizes of the

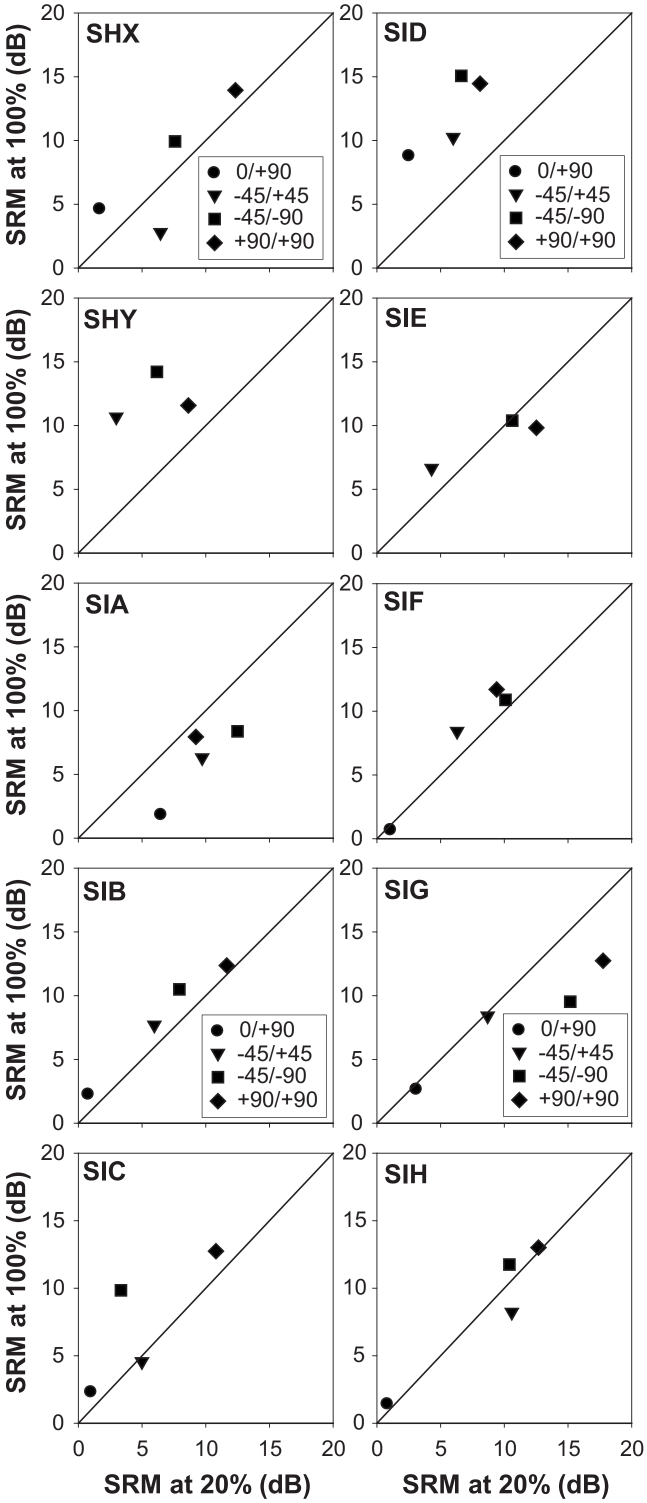

FIG. 5. Experiment 3 SRM data at 100% predictability and at 20% predictability are compared for each of the ten participants. Data points near the 45° line indicate that SRM was similar at these two levels of predictability. Note that the 0/+90 data point is not visible for two subjects whose measured SRM was slightly negative in this masker configuration, SHY and SIE.

error bars for the same-hemifield masker configurations cannot be considered strong evidence for the hypothesis.[5]

When comparing these results to data from other speech-on-speech masking experiments, it may seem surprising that SRTs in this experiment were at negative SNRs, even in the most challenging masker configurations. However, the spondee identification task in this experiment is substantially easier than sentence identification tasks employed in many other studies [e.g., see Freyman *et al.* (2001) and Hawley *et al.* (2004)]. It should be noted that group mean SRTs were negative for all masker configurations in experiment 2 as well (data not shown).

One could question the use of a constrained MLE procedure, which limits the influence of "inattention trials" on the threshold and slope of the estimated psychometric function, when calculating effects of informational masking. In the Appendix we show that the absence of an effect of predictability is not due to the constraints on the MLE procedure. Finally, it was noted earlier that SRM was 15 dB for the +90/+90 configuration in experiment 2, which is about 3 dB higher than SRM for this configuration in experiment 3. The reason for this difference is not known. Overall, however, the findings of the three experiments are quite consistent with each other.

V. GENERAL DISCUSSION

These experiments were motivated by theoretical considerations and by previous findings concerning effects of uncertainty in studies of the cocktail party problem. Masker uncertainty can be a source of informational masking, and this could have important implications for relatively unpredictable cocktail party environments. In previous studies in this field, in which the number and locations of maskers were held fixed throughout each block of trials, possible effects of predictability of the number and/or locations of maskers were not examined. We hypothesized that as predictability of the number and/or locations of speech maskers is increased, there will be (1) improved speech intelligibility and (2) increased SRM.

The data from the first of the experiments presented above did not support the hypothesis. The second and third experiments were each designed, in part, to more fully exploit differences in the predictability of masker locations than their predecessors. However, each further experiment lent additional evidence *against* an effect of predictability. Thus, these experiments offer no evidence for an effect of predictability of the number and locations of speech maskers on speech intelligibility or SRM in a cocktail party environment, at least not when the location of the target is known.

One further issue that needs to be addressed is the extent to which informational masking was present with the target/masker combinations used in these experiments. Despite evidence that energetic masking may play a relatively small role in the overall masking that occurs when speech is masked by interfering speech [e.g., see Brungart *et al.* (2006)], one cannot assume that each instance of masking of speech targets by speech interferers necessarily gives rise to informational masking. In fact, the high performance of subjects in the one-masker configurations of experiment 1 (at a very unfavorable SNR of −19 dB) suggests that there was little, if any, informational masking in that portion of experiment 1. If, however, we focus our attention on the two-masker configurations, which form the bulk of the data in these experiments, two lines of evidence support the view that informational masking was present. First, a comparison of the 0 and 0/0 configurations in experiment 1 reveals that when a second masker was added at the target location (while decreasing the level of each masker to maintain the same overall masker level), performance dropped from 77% to 21%. The key question here is whether the masking that resulted from adding a second speech interferer was greater than the amount that can be accounted for by energetic masking. This can be determined by comparing the data from experiment 1 to the data from experiments 2 and 3. If one compares Fig. 2(a) to Fig. 4(a), performance with a single masker at the target location was very close to 80% correct at a SNR of −19 dB [see the "0" configuration in Fig. 2(a)]; in contrast, the 80% threshold with two maskers at the target occurred at a SNR of −2.8 dB [see the "0/0" configuration in Fig. 4(a)]. This difference of just over 16 dB is far beyond the <3 dB reported by Bronkhorst and Plomp (1992) for the addition of a second masker under diotic stimulus presentation in an experiment in which the maskers were envelope-modulated noise.[6] A precedent for a large effect of adding a second speech masker at 0° may be found in experiments performed in free field [e.g., see Cullington and Zeng (2008)] and in data collected under diotic or monaural stimulus presentation conditions (Brungart, 2001; Brungart *et al.*, 2001). Second, further support for the view that informational masking was present in the two-masker configurations comes from the large amount of SRM in experiments 2 and 3. SRM for the +90/+90 configuration was approximately 15 dB in experiment 2 [Fig. 3(b)] and 12 dB in experiment 3 [Fig. 4(b)]. This is greater than SRM with energetic maskers [e.g., see Zurek (1993) and Bronkhorst (2000)] but is comparable to the ~12 dB of SRM that Hawley *et al.* (2004) found for speech or reversed-speech interferers. In addition, the sizable SRM we report in experiments 2 and 3 is consistent with the finding by Yost *et al.* (1996) of a particularly large benefit of spatial separation when more than two utterances are presented simultaneously. To summarize, (1) the experiments reported here were conducted under test conditions that give rise to informational masking in the two-masker configurations at least, and (2) there was no evidence of additional masking due to uncertainty about the number or location(s) of maskers.

It is worth noting that the use of parallel adaptive tracks in experiments 2 and 3 could lead to large variations in SNR from trial to trial. This, in effect, introduced a second source of uncertainty in the blocks with low predictability. Thus, the results of experiments 2 and 3 offer no evidence of an effect of either uncertainty about masker locations or uncertainty about SNR on speech intelligibility. These findings of experiments 2 and 3 are consistent with data reported by Freyman *et al.* (2007), who found that the uncertainty introduced by varying SNR from trial to trial has very little effect on speech intelligibility.

As described in the Introduction, it has been proposed that there are different strategies that could enable listeners to extract a target signal from competing sources (Durlach *et al.*, 2003a). One interpretation of the results of the experiments reported here is that knowing the number or locations of maskers ahead of time is of no benefit, provided that listeners know where to attend for the target signal (Arbogast and Kidd, 2000; Kidd *et al.*, 2005). This interpretation is consistent with the max listening strategy, which is characterized by enhanced sensitivity to sources from the target location. On the other hand, if there is some use of the min strategy by listeners in multitalker environments, then knowing the spatial locations of maskers ahead of time should result in improved performance. The absence of an effect of predictability of the number or locations of maskers suggests that use of the min strategy did not contribute significantly to these results. However, the min listening strategy may still contribute to speech intelligibility under other test conditions. For example, Brungart and Simpson (2007a) found a significant effect on speech intelligibility of varying the predictability of both target and masker locations simultaneously.

An alternative view is that listeners may not need much location information about any one source in order to benefit from spatial differences between target and masker(s) in a speech intelligibility task. Licklider (1948) found that the greatest release from masking in a speech identification task occurred for signal inversions that result in poor lateralization of target (N_0S_π) or masker ($N_\pi S_0$) sources. Culling and Summerfield (1995) measured identification of vowels defined by two noise bands in the presence of a diotic vowel defined by two noise bands; they found that listeners could more readily identify vowels with decorrelated formants than interaurally correlated vowels with a 700 μs ITD. Recent data by Brungart and Simpson (2007b) showed only small improvements in speech intelligibility when ITDs of the maskers, or of the target, were unchanged for several consecutive trials. In experiments using a three dimensional auditory display, which more closely approximates free-field presentation than the other experiments described in this paragraph, Drullman and Bronkhorst (2000) found that spatial separation makes a significant contribution to communication even when localization of the *target* is rather poor.

The findings reported here may have a parallel in recent studies that examined contributions of the predictability of the content of maskers to speech intelligibility in a cocktail party environment. Brungart and Simpson (2004) found that listeners were rather poor at utilizing reduced masker uncertainty to achieve improved speech intelligibility, even when the masking speech was frozen from trial to trial and even when they were given feedback after each trial. The authors speculated that some of their findings might be particular to the speech corpus they used, in which listener responses are drawn from a small set of numbers and colors. However, Freyman *et al.* (2007) reached a similar result with a very different stimulus set in which targets and maskers were nonsense sentences. When masker content was frozen and subjects were specifically instructed to ignore the repeated masker token, the authors found little improvement in speech intelligibility, even though the same masker token was repeated on dozens of consecutive trials. The results of the experiments described in this paragraph, together with the results reported in the current paper, suggest the possibility that contributions of reductions in masker uncertainty to speech intelligibility in a cocktail party environment may be quite limited. Alternatively, there may be ways of manipulating uncertainty that were not explored in these studies and that could yet prove to have significant effects on speech intelligibility in a cocktail party environment. It is worth noting that for masked detection of nonspeech stimuli, an effect of spatial uncertainty of maskers has been reported, but it is smaller than the effect of spectral uncertainty (Fan *et al.*, 2008).

In the experiments reported here, the same talker produced both targets and maskers. The purpose of this was to facilitate uncertainty-based masking effects, if any, through high target-masker similarity [e.g., see Durlach *et al.* (2003b)]. Nonetheless, in future experiments it would be good for these measures to be repeated with targets and maskers spoken by different talkers of the same gender or different-gender talkers. In addition, it would be useful to obtain these measures in other populations such as children, older adults, and hearing-impaired listeners. One application of this approach is to improve clinical measures of SRM when patients are fitted with bilateral amplification systems or cochlear implants; these measures typically leave out much of the variability of naturalistic listening environments. In the experiments reported in this paper, we varied one type of uncertainty in a simulated cocktail party environment, while other features of the listening environment were held fixed. Although we found no evidence of additional informational masking due to uncertainty about the number and locations of maskers, other aspects of masker uncertainty may yet prove to be a source of informational masking and may need to be incorporated into clinical testing.

ACKNOWLEDGMENTS

The authors are grateful to Jonathan Riley, Jordan Bonnett, and Brent Saffran for their assistance with data collection. We would also like to thank two anonymous reviewers and the associate editor for their thoughtful comments on previous versions of this paper. This work was supported by Grant No. R01 DC030083 from NIH-NIDCD.

APPENDIX: CONSTRAINED MLE

In the constrained MLE method that was used to calculate thresholds in experiments 2 and 3, the value of the lambda parameter, which determines the asymptote of the psychometric function, was confined to a narrow range. Use of this approach in informational masking experiments could raise concerns because informational masking has significant effects on both the slope and the asymptote of the psychometric function (Lutfi *et al.*, 2003). The study of Lutfi *et al.* (2003) found some lambda values on the order of 0.25 under conditions that cause informational masking. On the other hand, in the constrained MLE procedure used here lambda is not allowed to exceed 0.05. Although this method for estimating psychometric functions has been used successfully to measure effects of uncertainty on speech intelligibility in previous research (Johnstone, 2006), it is possible that constraining the lambda parameter to a narrow range could at least partially offset any informational masking effects and prevent significant differences among the tested masker configurations. In order to control for this, MLE estimation of the psychometric function for each adaptive track in experiments 2 and 3 was repeated under conditions in which the constraint on the lambda parameter was relaxed. Namely, in the reanalysis, lambda was allowed to take on values in the much larger range of [0,0.5]. Then possible effects of predictability on the beta (slope) and lambda (asymptote) parameters of the psychometric function were analyzed statistically [see also Lutfi *et al.* (2003)]. Across adaptive tracks, the constraint influenced the estimation of psychometric functions rather infrequently. In fact, when the limitation on lambda values was relaxed, the fitting parameters were unchanged for 97% of adaptive tracks in experiments 2 and 3. Furthermore, ANOVAs were performed on the reanalyzed data to test for possible effects of predictability on the beta and lambda parameters in experiments 2 and 3. In experiment 2 there was no main effect of predictability on beta or lambda and there were no significant interactions involving predictability. In experiment 3 there was no main effect of predictability on either beta or lambda; however, there was a significant interaction between masker configuration and predictability ($p < 0.05$) affecting the beta parameter. In order to explore whether the character of the interaction is consistent with greater informational masking under conditions of uncertainty, the beta values from experiment 3 have been plotted across masker configuration and predictability in Fig. 6. If there were greater informational masking at low predictability, then one would expect the following: (1) in the 0/0 configuration, beta values would be larger (shallower slopes) at 20% predictability than at 100% predictability, and (2) differences in beta values as a function of predictability would decrease as maskers are spatially separated from the target. In fact, the data do not meet the second criterion. Rather, the −45/−90 configuration, which has one of the

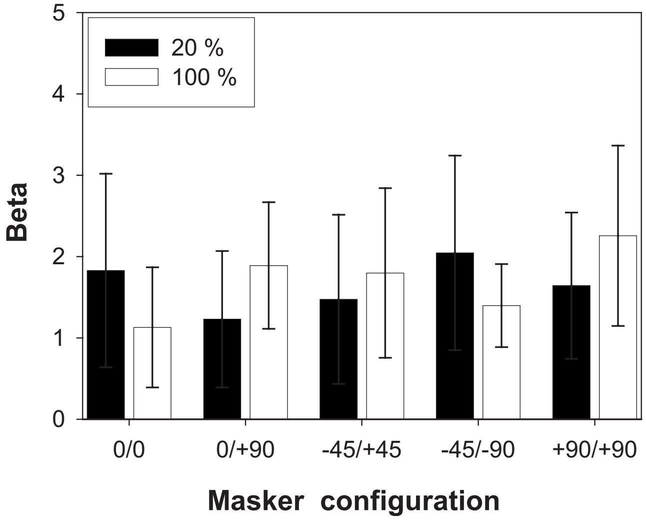

FIG. 6. Group mean (±SD) values of the beta parameter from a revised MLE fit of the experiment 3 data in which the constraint on the lambda parameter was relaxed. An ANOVA on the values of the beta parameter revealed a significant interaction between the masker configuration and predictability in experiment 3. The character of the interaction as revealed by this plot is not consistent with greater informational masking at 20% predictability than at 100% predictability (see the Appendix).

largest target-masker separations, is quite similar to the 0/0 configuration, while the trend in the data for the +90/+90 configuration opposes that of the 0/0 and −45/−90 configurations. Thus, the finding of no additional masking due to uncertainty about the number or locations of maskers was not caused by the constraint on the lambda parameter in the MLE method used for estimating psychometric functions.

[1] Further studies have used variations on the technique described by Freyman *et al.* (1999, 2001) to show that even subtle spatial differences can lead to significant SRM for competing speech (Brungart *et al.*, 2005; Rakerd *et al.*, 2006; Balakrishnan and Freyman, 2008).

[2] Based on some of the results described in this Introduction, particularly the more recent findings, one could argue that when predictability of the masker configuration is high, there will be less informational masking to release and thus *less* SRM. As will be seen in the results and discussion, the direction of this hypothesis did not significantly impact the interpretation of the results of these experiments.

[3] The situation was slightly more complicated for 60% blocks because 40/0.6 is not an integer. In 60% blocks, the 20 familiarization trials were followed by 70 experimental trials, of which 42 trials contained the tested masker configuration. However, in order to keep the number of analyzed trials at 40 for all experiments, in 60% blocks the last two trials of the tested masker configuration were excluded from the data analysis.

[4] Masker durations in this paper are reported to one decimal place. For example, the standard deviation of the duration of the maskers used in experiment 3 was less than 0.05 s and is reported as 0.0.

[5] Note that in experiment 2 each subject's SRT was usually an average of thresholds from multiple adaptive tracks at 80% predictability, whereas the SRT at 20% predictability was typically from one or two adaptive tracks. Thus, it may seem that sampling issues could explain the larger error bars at 20% predictability in Fig. 3. However, such an account leaves unexplained the larger error bars in the same-hemifield masker configurations at 20% predictability in Fig. 4, where there were no such sampling issues. Namely, in experiment 3 the number of adaptive tracks per masker configuration was never higher at 100% predictability than at 20% predictability for any subject or any masker configuration.

[6] Because mean performance for the "0" configuration in experiment 1 was actually 77%, there is some error in equating the SNR of −19 dB with the 80% SRT for this configuration. Using the calculated slopes of each of the psychometric functions from experiment 3, we estimated that for the 0/0 configuration the upper limit of the difference between the 77% correct point and the 80% SRT is <1 dB. Thus, even if one compensates for this difference, the data suggest a large increase in masking when a second speech masker was added.

Arbogast, T. L., and Kidd, G. (**2000**). "Evidence for spatial tuning in informational masking using the probe-signal method," J. Acoust. Soc. Am. **108**, 1803–1810.

Balakrishnan, U., and Freyman, R. L. (**2008**). "Speech detection in spatial and nonspatial speech maskers," J. Acoust. Soc. Am. **123**, 2680–2691.

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

Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

Bronkhorst, A. W., 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. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.

Brungart, D. S., and Simpson, B. D. (**2004**). "Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty," J. Acoust. Soc. Am. **115**, 301–310.

Brungart, D. S., and Simpson, B. D. (**2007a**). "Cocktail party listening in a dynamic multitalker environment," Percept. Psychophys. **69**, 79–91.

Brungart, D. S., and Simpson, B. D. (**2007b**). "Selective spatial attention in a dynamic cocktail party task: Evidence for a strategy based on masker minimization," Presented at the Acoustical Society of America Conference, Salt Lake City, June.

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

Brungart, D. S., Simpson, B. D., and Freyman, R. L. (**2005**). "Precedence-based speech segregation in a virtual auditory environment," J. Acoust. Soc. Am. **118**, 3241–3251.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.

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

Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2004**). "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **116**, 1057–1065.

Culling, J. F., and Summerfield, Q. (**1995**). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am. **98**, 785–797.

Cullington, H. E., and Zeng, F.-G. (**2008**). "Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects," J. Acoust. Soc. Am. **123**, 450–461.

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

Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003a**). "Note on informational masking," J. Acoust. Soc. Am. **113**, 2984–7.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G., Jr. (**2003b**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.

Ericson, M. A., Brungart, D. S., and Simpson, B. D. (**2004**). "Factors that influence intelligibility in multitalker speech displays," Int. J. Aviation Psych. **14**, 313–334.

Fan, W. L., Streeter, T. M., and Durlach, N. I. (**2008**). "Effect of spatial uncertainty of masker on masked detection for nonspeech stimuli," J. Acoust. Soc. Am. **124**, 36–39.

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

Freyman, R. L., Helfer, K. S., and Balakrishnan, U. (**2007**). "Variability and uncertainty in masking by competing speech," J. Acoust. Soc. Am. **121**, 1040–1046.

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

Garadat, S. N., and Litovsky, R. Y. (**2007**). "Speech intelligibility in free field: Spatial unmasking in preschool children," J. Acoust. Soc. Am. **121**, 1047–1055.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of competitor," J. Acoust. Soc. Am. **115**, 833–843.

Johnstone, P. M. (**2006**). "Informational masking and spatial asymmetry in a 'cocktail party' environment: Results with children and adults," Ph.D. dissertation, University of Wisconsin-Madison.

Johnstone, P. M., and Litovsky, R. Y. (**2006**). "Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults," J. Acoust. Soc. Am. **120**, 2177–2189.

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

Kidd, G., Jr., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **104**, 422–431.

Leek, M. R. (**2001**). "Adaptive procedures in psychophysical research," Percept. Psychophys. **63**, 1279–1292.

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

Licklider, J. C. R. (**1948**). "The influence of interaural phase relations upon the masking of speech by white noise," J. Acoust. Soc. Am. **20**, 150–159.

Litovsky, R. Y. (**2005**). "Speech intelligibility and spatial release from masking in young children," J. Acoust. Soc. Am. **117**, 3091–3099.

Lutfi, R. A. (**1993**). "A model of auditory pattern analysis based on component-relative-entropy," J. Acoust. Soc. Am. **94**, 748–758.

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

Oh, E. L., and Lutfi, R. A. (**1998**). "Nonmonotonicity of informational masking," J. Acoust. Soc. Am. **104**, 3489–3499.

Oxenham, A. J., Fligor, B. J., Mason, C. R., and Kidd, G., Jr. (**2003**). "Informational masking and musical training," J. Acoust. Soc. Am. **114**, 1543–1549.

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

Rakerd, B., Aaronson, N. L., and Hartmann, W. M. (**2006**). "Release from speech-on-speech masking by adding a delayed masker at a different location," J. Acoust. Soc. Am. **119**, 1597–1605.

Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.

Shinn-Cunningham, B. G., and Ihlefeld, A. (**2004**). "Selective and divided attention: Extracting information from simultaneous sound sources," Proceedings of the International Conference on Auditory Display.

Watson, C. S. (**2005**). "Some comments on informational masking," Acust. Acta Acust. **91**, 502–512.

Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). "Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty," J. Acoust. Soc. Am. **60**, 1176–1186.

Wichmann, F. A., and Hill, J. (**2001a**). "The psychometric function: I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**, 1290–1313.

Wichmann, F. A., and Hill, J. (**2001b**). "The psychometric function: II. Bootstrap-based confidence intervals and sampling," Percept. Psychophys. **63**, 1314–1329.

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

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