Effect of spatial uncertainty of masker on masked detection for nonspeech stimuli (L)

Wei Li Fan,^{a)} Timothy M. Streeter,^{b)} and Nathaniel I. Durlach^{c)}
Hearing Research Center, Boston University, Boston, Massachusetts 02215

(Received 1 June 2007; revised 25 April 2008; accepted 30 April 2008)

Research on informational masking for nonspeech stimuli has focused on the effects of spectral uncertainty in the masker. In this letter, results are presented from some preliminary probe experiments in which the spectrum of the masker is held fixed but the spatial properties of the masker are randomized. In addition, in some tests, the overall level of the stimulus is randomized. These experiments differ from previous experiments that have measured the effect of spatial uncertainty on masking in that the only attributes (aside from level) that distinguish the target from the masker are the spatial attributes; in all of the tests, the target and masker were statistically identical, statistically independent, narrowband noise signals. In general, the results indicate that detection performance is degraded by spatial uncertainty in the masker but that compared both to the effects of spectral uncertainty and to the effects of overall-level uncertainty, the effects of spatial uncertainty are relatively small. © 2008 Acoustical Society of America. [DOI: 10.1121/1.2932257]

PACS number(s): 43.66.Dc [RAL]

Pages: 36–39

I. INTRODUCTION AND BACKGROUND

The idea that spatial uncertainty in the masker might degrade detection of a target signal arises naturally from previous results obtained in two areas: informational masking and binaural unmasking. In the first area, it has been shown that the threshold for a tonal target in a multitone masker with most of its energy outside a protected zone around the target tone can be raised 20–40 dB by randomly varying the masker spectrum (see references in [Durlach et al., 2005](#)). Also, it has been shown in studies of speech intelligibility that uncertainty about the spatial properties of the masker degrades the intelligibility of the speech target ([Kidd et al., 2007](#)). Although theoretical efforts have been made to explain these and other informational masking results (e.g., see [Watson, 1987, 2005](#); [Lutfi, 1993](#); [Oh and Lutfi, 1998](#); [Durlach, 2006](#); [Durlach et al., 2003, 2005](#); [Shinn-Cunningham, 2005](#); [Kidd et al., 2007](#)), no satisfactory theory is yet available. Furthermore, there are currently no data available for simple nonspeech stimuli on the effects of spatial uncertainty of the masker for cases in which target and masker have the same spectrum.

In the second area, it has been shown that binaural detection thresholds are decreased when the interaural relations of the masker differ from those of the target. In particular, for a diotic target signal, performance is better when the masker is dichotic than when it is diotic. Moreover, this improvement occurs both when the interaural relations are determined by the locations of loudspeakers in field listening or by the values of interaural time delays (ITDs) or interaural level differences (ILDs) in earphone listening. Relevant data and theory in this area can be found in [Durlach and Colburn](#)

(1978), [Colburn and Durlach \(1978\)](#), and [Stern and Trahiotis \(1995\)](#). Of particular interest here is the idea that the binaural processing effectively creates binaural spatial channels (like the peripheral processing creates frequency channels) with different channels corresponding to different source angles or, more generally, different interaural configurations (i.e., different pairs of ITDs and ILDs). If one further assumes that rather than focusing on the target the listener attempts to reduce the effects of masking by nulling out the masker (see discussion in [Durlach et al., 2003](#); [Durlach, 1972](#); and [de Cheveigne and McAdams, 1995](#)), then it is reasonable to expect that masked detection performance will be especially vulnerable to randomization of the spatial properties of the masker.

Past work of particular relevance to our experiments is that of [Bernstein and Trahiotis \(1997\)](#) in which they measured the effect of roving interaural differences (imposed on both masker and masker plus target) on the detection of a 500 Hz tone in the presence of masking noise. Although the effects of this roving were found to be small (and to be consistent with a version of the equalization and cancellation model), these experiments differed from those reported here in that (among other things) the target and masker had distinguishable spectra.

In this letter, we report results of three preliminary, relatively independent, probe experiments concerned with the effects on target detection of uncertainty in the spatial properties of the masker for cases in which the target and masker are simple nonspeech stimuli that differ only in spatial properties. Although these experiments were limited both in the types of tests performed and the manner in which procedures and parameter values varied across experiments, we believe that they are well worth reporting.

II. EXPERIMENTS

All signals were presented through earphones; targets and maskers were statistically identical, statistically indepen-

^{a)}Electronic mail: wei.l.fan@gmail.com

^{b)}Electronic mail: timstr@cns.bu.edu

^{c)}Author to whom correspondence should be addressed. Tel.: 617 253 2534. FAX: 617 258 7003. Electronic mail: durlach@mit.edu

dent random noise signals with identical spectra; the spectrum of these signals was flat in the region 300–800 Hz and fell off rapidly outside this region; the signal durations were 250 ms in experiment 1 and 500 ms in experiments 2 and 3, with target and masker gated on and off together; and the experimental paradigm was a one-interval 2AFC (respond “target present” or “target absent”) detection paradigm with trial-by-trial feedback and equal a priori probabilities for target present and target absent. The target was always diotic and the masker (with only one exception in experiment 3) always lay outside a spatially protected zone around the target. In some cases, the spatial properties of the masker were held fixed, and in other cases they varied randomly from trial to trial. The spatially fixed case is denoted SF, and the spatially random case SR. Finally, in some of the tests the overall stimulus level (masker or masker plus target) was roved randomly to reduce loudness cues. When the overall level is fixed, it is denoted LF, when random, LR. Thus, for example, LRSF denotes “level random, space fixed.” Because the roving level applied to the whole stimulus and not just the masker, it never affected the target-to-masker ratio (TMR). In all experiments, subjects were between the ages of 18 and 30, had normal hearing, and participated in training sessions (1 h long) prior to and identical to the test sessions.

In experiment 1, the maskers consisted of triplets (three simultaneous noise samples) with each component of the triplet having zero ILD and an ITD selected randomly without replacement from the set $\{\pm 600, \pm 400, \pm 300, \pm 200 \mu\text{s}\}$. The protected zone around the diotic target (ITD=0 μs) was thus -200 to $200 \mu\text{s}$. The masker was selected randomly from a fixed set of ten such triplets. The level was always roved randomly over the range 50–70 dB SPL and the TMR was always computed using the energy in the whole masker triplet. In experiment 2, the ITDs of the maskers were replaced by virtual azimuth angles using HRTFs based on recordings made with the KEMAR mannequin in anechoic space (Gardner and Martin, 1995). The ten triplets used were drawn from the azimuth set $\{\pm 70^\circ, \pm 45^\circ, \pm 30^\circ, \pm 20^\circ, \pm 15^\circ\}$ and the protected zone was -15° to $+15^\circ$. Again, the masking triplets were chosen randomly and the masking level includes all three components of the triplet. The base level was 54 dB SPL and, when the level was roved, the range was 54–64 dB SPL. In experiment 3, the ten triplets of angles used for the maskers were replaced by six unitary angles. These angles were drawn from the set $\{\pm 75^\circ, \pm 40^\circ, \pm 20^\circ\}$, thus providing a protected zone of -20° to $+20^\circ$. In addition, LF tests were performed in which the masker, like the target, was diotic (so that the only detection cue was stimulus energy). The base level in this experiment was 58 dB SPL, and in those tests where the overall level was roved, the range was 58–68 dB SPL. In the SF tests, all angles in the above set of angles were used. Also, the TMR for a given masker was defined as the average TMR across the two ears. In all experiments, for each TMR, condition, and subject, the stimulus-response matrix obtained was processed to obtain estimates of sensitivity d' and response bias β in the usual

fashion (e.g., Durlach *et al.*, 2005). Further comments on experimental procedures and data-analysis procedures are presented below.

III. EXPERIMENTAL RESULTS

The d' vs TMR curves obtained in experiment 1, as well as threshold and slope estimates based on linear fits, are shown in the top row of Fig. 1. In graphs for individual subjects, and for both LRSF and LRSR, each point is based on 1000 trials (100 per triplet). For the LRSF case, the points were obtained by computing d' for each fixed case and then averaging over these values of d' (the error bars giving the standard deviations across this set of d' 's). The ordering of the tests was random, except that the LRSF tests were conducted before the LRSR tests and, for each of these cases, the value of TMR was held constant during each 100-trial run, with different values of TMR being tested in descending order. In the graph showing the results averaged over subjects, the error bars show the standard error in the mean.

The d' curves, together with the threshold and slope estimates, obtained in experiment 2 are shown in the middle row of Fig. 1. In the graphs for individual subjects, the points for the spatially fixed cases (LFSF and LRSF) are based on 2000 trials (200 per case) and for the spatially random cases (LFSR and LRSR) on 400 trials. For the spatially fixed cases, the values of d' were again obtained by averaging across the d' values obtained with the individual fixed constituents (the error bars again showing the standard deviations across these values of d'). The various conditions in this experiment were tested in the order LFSF, LFSR, LRSF, LRSR and the various TMRs in descending order.

The d' results obtained in experiment 3 are shown in the bottom row of Fig. 1. The number of trials per point and the ordering of the tests were similar to those in experiment 2 with the following exceptions: (a) the number of trials per point for the fixed cases was 1200 rather than 2000 (corresponding to the reduction in the number of maskers from 10 to 6) and (b) the case in which the masker, like the target, was diotic was tested at the end of the experiment (with 400 trials per TMR value).

In all experiments, informational masking in the spatial domain occurs to the extent that the SR curves are on the right of (or lower than) the SF curves (compare LFSR to LFSF and LRSR to LRSF). Such results are clearly apparent in experiments 2 and 3 (where the roving-level range was less than or equal to 10 dB) but not in experiment 1 (where the roving-level range was always 20 dB). In terms of estimated thresholds, the shifts between the SR cases and the SF cases (examined over both the LR and LF conditions in both experiments 2 and 3) are of the order of 2–4 dB. The consistency of these results, combined with the small standard error in the estimates of threshold across subjects (less than 1 dB for all test conditions in both experiments 2 and 3), is a strong indication of the degrading effect of spatial uncertainty.

The tendency for roving level to play a dominant role in these experiments is evidenced not only by the results of experiment 1 (where any effect of spatial uncertainty has

FIG. 1. Experimental results on sensitivity d' . Rows 1, 2, and 3 give results for experiments 1, 2, and 3, respectively. LFSF means level fixed, spatial structure fixed; LRSF means level random, spatial structure fixed; LRSR means level random, spatial structure random; and diotic means that the masker signal (like the target signal) is the same in both ears. Results for LFSF and LRSF show averages over results for the various fixed maskers with any $d' > 3$ being given the value 3 and with error bars indicating standard deviations across fixed spatial conditions. Graphs in the rightmost box in each row give averages over subjects, with error bars indicating standard errors over subjects. In some cases, the error bars are hidden behind the symbols used for marking the average data points. Estimates of threshold values (in dB and defined by the condition $d' = 1$), as well as estimates of the slopes, for all the average-subject d' curves, are shown (along with the standard error of these estimates) at the right edge of each row.

been overwhelmed by level uncertainty) but also by the results of experiments 2 and 3. Compared with the threshold shifts of 2–4 dB associated with spatial uncertainty, threshold shifts associated with level uncertainty are 7–10 dB.

The importance of roving level is also evident in comparisons of thresholds (defined by $d' = 1$) across experiments. Whereas the thresholds in experiment 1 (where there was a rove of 20 dB) are roughly 0 dB, the thresholds in experiments 2 and 3 for cases in which there was no level rove are roughly –10 dB. (That this difference in thresholds is due primarily to the roving-level factor and not other differences between the experiments is indicated by the similarities in the thresholds when one compares the results of experiment 1 to those of experiments 2 and 3 when the level is roved in experiments 2 and 3). In addition, roving level appears to affect the slope of the d' curves: when the rove was increased from 0 to 10 dB in experiments 2 and 3, the slope always decreased (for both the SF cases and the SR cases).

Differences among the d' values across individual maskers for the LFSF and LRSF cases (see the error bars associated with the filled symbols for individual subjects) generally appear modest. Differences across subjects shown by the error bars in the average-subject graphs vary from very small (see filled circles for experiment 2) to very large (e.g., see top open circle in experiment 2 and filled circle at TMR = –5 dB in experiment 3).

The results for the diotic tests in experiment 3 are roughly consistent with what one would expect from previous results on intensity JNDs. According to the data in Fig. 1, the threshold for the diotic case is roughly –8 dB, which corresponds to a Weber fraction of approximately 0.5 dB. Note also that the difference in performance for this case and the LFSF case in experiment 3 indicates that the detection cue for the LFSF case cannot be merely the total energy level in the whole stimulus.

Overall, the results on β showed little response bias. Pooling the data across subjects, test conditions, and values of TMR, we found the mean and standard error for each experiment to be –0.22 and 0.15 (experiment 1), 0.03 and 0.09 (experiment 2), and –0.02 and 0.07 (experiment 3). Further analysis of the data for experiment 1 in which trials for the top half and bottom half of the 20 dB roving-level range were kept separate showed that the negative mean bias for this experiment (–0.22) was due entirely to the existence of a negative bias in the top half of the range: whereas the mean bias for the bottom half was 0.07, the mean bias for the top half was –0.57. Inasmuch as a negative value of β corresponds to a bias to say target present, this result is consistent with the notion that the overall stimulus level served as at least one component of the detection cue used in experiment 1. (An equivalent analysis of the d' data in experiment 1 showed a slight tendency for d' to be larger at the higher levels).

IV. DISCUSSION

In general, the results indicate that sensitivity d' tends to be reduced by spatial randomization of maskers with little energy at the target location, but that the magnitude of this "spatial informational masking" is small compared both to the effect of randomizing the level of the stimulus or, as documented in the literature (e.g., Durlach *et al.*, 2005), to the effect of "spectral informational masking."

Some of the factors that might be related to our results are the following. First, in experiments 2 and 3, the use of HRTFs necessarily involved the introduction of some spectral differences among the various maskers (and between target and masker). Thus, our explanation of the difference between the results of experiments 2 and 3 and those of experiment 1 solely in terms of the much larger roving-level range in experiment 1 may be incomplete. Second, the number of JNDs in the spatial domain is relatively small compared to that in the spectral domain. Thus, perceptual randomization in the spatial domain might be more difficult to achieve than in the spectral domain. Third, a change in spatial location might be less likely to induce the perception of a change in auditory source ("auditory object") than a change in spectrum. Thus, spatial randomization might be less distracting (attention grabbing) than a change in spectrum. Fourth, susceptibility to informational masking may decrease as the domain of masker randomization becomes more central. It would be interesting to compare spatial to spectral informational masking in a system where spatial processing precedes spectral processing.

Further issues of interest in connection with our results focus on the effects of level randomization. For example, to what extent does the strong effect of level randomization that is seen in our experiments on spatial informational masking also occur in spectral informational masking? Currently available information relevant to this question is rather meager. On the one hand, to the extent that the model considered in Durlach *et al.* (2005) is correct (a model in which the reduction in sensitivity d' caused by spectral randomization of the masker is interpreted in terms of level randomization at the output of a spectral filter), one would predict a strong effect in the spectral domain as well as the spatial domain. On the other hand, there exist at least some data that suggest that the level randomization effect is very weak in the spectral domain (e.g., Mason *et al.*, 1984; Richards and Neff, 2004). In the authors' opinion, further experimental work is needed on the effect of level randomization in both the spatial domain and the spectral domain.

ACKNOWLEDGMENTS

The authors are indebted to Chris Mason for many useful comments on drafts of this material and to the following grants for supporting this work: NIH RO1 DC00100, NIH/NIDCD P30 DC004663, and AFOSR FA9950-05-01-2005.

- Bernstein, L. R., and Trahiotis, C. (1997). "The effects of randomizing values of interaural disparities on binaural detection and on discrimination of interaural correlation," *J. Acoust. Soc. Am.* **102**, 1113–1120.
- Colburn, H. S., and Durlach, N. I. (1978). "Models of binaural interaction," in *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), Vol. IV.
- de Cheveigne, A., and McAdams, S. (1995). "Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement," *J. Acoust. Soc. Am.* **97**, 3736–3748.
- Durlach, N. I. (2006). "Auditory masking: Need for improved conceptual structure," *J. Acoust. Soc. Am.* **120**, 1787–1780.
- Durlach, N. I. (1972). "Binaural signal detection: Equalization and cancellation theory," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York).
- Durlach, N. I., and Colburn, H. S. (1978). "Binaural Phenomena," in *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), Vol. IV.
- Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003). "Note on informational masking," *J. Acoust. Soc. Am.* **113**, 2984–2987.
- Durlach, N. I., Mason, C. R., Gallun, F. J., Shinn-Cunningham, B. G., Colburn, H. S., and Kidd, G., Jr. (2005). "Informational masking for simultaneous nonspeech stimuli: Psychoacoustic functions for fixed and randomly mixed maskers," *J. Acoust. Soc. Am.* **118**, 2482–2497.
- Gardner, W. G., and Martin, K. D. (1995). "HRTF measurements of a KE-MAR," *J. Acoust. Soc. Am.* **97**, 3907–3908.
- Kidd, G. I., Jr., Mason, C. R., Richard, V. M., Gallun, F. J., and Durlach, N. I. (2007). "Informational masking," in *Auditory Perception of Sound Sources*, Springer Handbook of Auditory Research Vol. **29**, edited by W. Yost (Springer, New York).
- Lutfi, R. A. (1993). "A model of auditory pattern analysis based on component-relative-entropy," *J. Acoust. Soc. Am.* **94**, 748–758.
- Mason, C. R., Kidd, G., Jr., Hanna, T. E., and Green, D. M. (1984). "Profile analysis and level variation," *Hear. Res.* **13**, 269–275.
- Oh, E. L., and Lutfi, R. A. (1998). "Nonmonotonicity of informational masking," *J. Acoust. Soc. Am.* **104**, 3488–3499.
- Richards, V. M., and Neff, D. L. (2004). "Cuing effects for informational masking," *J. Acoust. Soc. Am.* **115**, 289–300.
- Shinn-Cunningham, B. G. (2005). "Influences of spatial cues on grouping and understanding sound," *Proceedings of the Forum Acusticum*, 29 August–2 September.
- Stern, R. M., and Trahiotis, C. (1995). "Models of binaural interaction," in *Handbook of Perception and Cognition*, edited by B. C. J. Moore (Academic, New York), Vol. 6.
- Watson, C. S. (1987). "Uncertainty, informational masking, and the capacity of immediate auditory memory," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ).
- Watson, C. S. (2005). "Some comments on informational masking," *Acta Acust.* **91**, 502–512.