I ma t a ma : C t ac t t c t tm c a t b d c a a tma m a t[a)]

Nathaniel I. Durlach,[b)] Christine R. Mason, Barbara G. Shinn-Cunningham,
Tanya L. Arbogast, H. Steven Colburn, and Gerald Kidd, Jr.
Hearing Research Center, Boston University, Boston, Massachusetts 02215

sary or a suf®cient condition to produce nonenergetic masking, there is no doubt that stimulus uncertainty can produce large amounts of such masking under a wide variety of conditions. Further comments on some of these conceptual issues are available in Durlach *et al.* (2003a).

In this paper, we report the results of a series of detection experiments (involving tonal targets and random multiquency region around the signal tone (called the ''protected region''). Because the amount of energetic masking decreases with the size of the protected region (cf. Neff *et al.*, 1993), many of the experiments designed to focus on informational masking use protected regions that are equal to or greater than the ''critical band'' around the given target frequency. Also, because in many cases both energetic and informational masking are expected to occur at least to some degree, attention is given to how much masking is energetic and how much is informational, and to how the two types of masking interact (e.g., Lutfi, 1990). The number of tonal components in the masker, and the frequency range of those components, as well as the extent to which the components are randomized in amplitude and frequency, varies with the experiment. Also, the relative effect of randomizing the spectrum of the masker between intervals and between trials in two-interval paradigms has been examined (cf. Neff and Green, 1987; Neff, 1995; Neff and Dethlefs, 1995; Wright and Saberi, 1999; Richards *et al.*, 2002).

The effect of randomizing the spectrum of the masker can be exceedingly large. However, the results of such experiments are strongly listener dependent. Whereas some listeners, occasionally referred to as ''holistic'' or ''synthetic'' listeners, evidence very large effects of the uncertainty in the multitone masker, other listeners, often referred to as ''analytic'' listeners, show hardly any effect at all (Espinoza-Varas and Watson, 1989; Neff and Dethlefs, 1995; Lutfi *et al.*, 2003). Moreover, it appears that the variation in the size of this effect arises primarily from variation in the masked threshold for the uncertain-masker case rather than for the certain-masker case (which is often broadband noise). Questions of current interest in this area include: To what extent does a listener's ability to resist informational masking vary with the experimental task? What other differences among listeners correlate with this ability? How much can this ability be enhanced by training? According to a recent study by Oxenham *et al.* (2003), there is a significant positive correlation between resistance to informational masking and musical training.

Despite the large amount of data on informational masking that has become available over the past few years, there have been only a few attempts to model informational masking. Currently there is no model that satisfactorily accounts for all of the empirical results, even when limited to the body of work on detecting a target tone in a simultaneous random multitone masker discussed above. The most extensive effort to date is the CoRE (component relative entropy) model proposed by Lutfi (1993). Oh and Lutfi (1998) have shown that the CoRE model, which uses the weighted outputs (mean levels and variances) of a set of peripheral filters in addition to a variable bandwidth ''attentional'' filter, can predict the variation in threshold with number of masker tones (as originally found by Neff and Green, 1987) with considerable accuracy. In other cases, however, such as the detection threshold for an inharmonic tone embedded in a randomized harmonic multitone masker, the model is less successful (Oh and Lutfi, 2000). In distinct but related efforts, both Wright and Saberi (1999) and Richards *et al.* (2002) have interpreted informational masking data in terms of channel-weighting analyses.

Apart from the modeling work noted above, which is focused primarily on uncertainty in the stimulus combined with channel weights, the main theoretical notions that have been proposed to help understand informational masking phenomena concern the perceptual grouping or segregation of target and masker (Leek *et al.*, 1991; Kidd *et al.*, 1994; Neff, 1995; Oh and Lutfi, 2000). At a crude intuitive level, informational masking occurs because the listener finds it difficult to focus attention on the target in the presence of a distracting or confusing masker. Although uncertainty is clearly relevant to this phenomenon, so is the extent to which the target ''sounds like'' the masker and is grouped with the masker. In the words of Leek *et al.* (1991, pp. 205–206), ''Informational masking is broadly defined as a degradation of auditory detection or discrimination of a signal embedded in a context of other similar sounds'' and ''A target that is sufficiently different from the surrounding tones along some acoustic dimension will be heard with increased precision.'' Thus, in addition to uncertainty, similarity, which is well known to be a factor in the extent to which auditory objects may be grouped into a single auditory image or segregated into separate images (e.g., Bregman, 1990), has also been distinct images is difficult to quantify, there is some hope that eventually one or more metrics of target-masker similarity can prove useful in predicting the amount of informational masking that occurs. It should also be noted that target-masker similarity appears to be important in a wide range of complex auditory detection and recognition tasks. For example, there is substantial evidence that target-masker similarity plays a major role in speech reception tasks: informational masking tends to increase as the masker goes from noise to speech to same-sex talker to same talker (e.g., Freyman *et al.*, 1999, 2001; Brungart, 2001; Brungart *et al.*, 2001; Arbogast *et al.*, 2002).[2] Furthermore, a recent study by Kidd *et al.* (2002) provides support for the proposition that target-masker similarity affects informational masking for nonspeech pattern recognition. Finally, it should be noted that similarity is a well-known factor in the degree to which stimuli interfere with or mask each other in sequential as well as simultaneous masking [for extensive work on temporal patterns and sequential masking, see the work by Watson and his colleagues as exemplified in Watson *et al.* (1976), Watson and Kelly (1981), Watson (1987), and Espinoza-Varas and Watson (1989)] and in sensory channels other than audition [see, for example, Turvey (1973) for a consideration of pattern masking in vision].

The purpose of the present study was to examine informational masking, and release from informational masking, for conditions in which target-masker similarity was varied sfTl8se i...1(re5w6T2.1(i..isee,)-u5/F12,)-u512(to)-edretfa(to)n condition is shorter than the target in the S condition (i.e., the target is turned on after the onset of the masker). This experiment is similar to one performed by Neff (1995).

2. The reversed-frequency-sweep experiment (Sweep)

As shown in the bottom two panels of Fig. 1, the masker tones are all upward frequency glides. In the S condition, the target is a glide with the same extent and direction as the masker components. In the D condition, the target glide is in the opposite direction from the masker components.

3. The separate-spatial-channels experiment (Spatial)

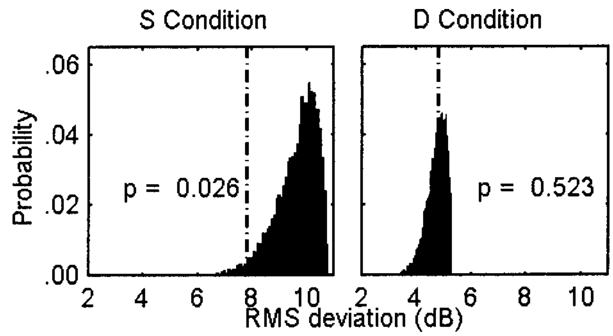
Figure 2 illustrates the third experiment. The S condition consists of a diotic multitone masker with a diotic tonal target. In the D condition, the masker component increased by a factor of 1.49 over the 300-ms duration of the masker. In order to maintain the 5000-Hz upper limit on the frequencies present in the masker, the highest possible starting frequency of any masker component was 3356 Hz (5000 Hz/1.49). In the S case, the target was an upward glide from 820 to 1220 Hz. In the D case, it was a downward glide covering the same frequency range. In all cases, each component had a duration of 300 ms including 20 ms cosine squared onsets and offsets.

In the Spatial experiment (Fig. 2), the masker was the same as that used in the Duration experiment, except that it was presented diotically rather than monotically. The target experiments there is considerably more masking for the S conditions than the D conditions. Given the specific target-masker parameters used and these specific five listeners, the most masking (average of approasking cases, the difference $S-D$ is negligible, in other cases it is nearly 40 dB. The apparent dependence of this difference on both the experiment and the listener clearly reflects the statistical significance of the three-way interaction of these factors mentioned previously.

Fourth, it is evident in Fig. 5 that the variation among the error bars is extremely large (

always presented prior to each trial. Similarly, in the Kidd *et al.* multiple-burst experiments, unlike our multiple-burst experiments, the S condition was transformed into the D condition by altering the masker rather than the target. In addition, in neither study was the set of listeners held fixed across the experiments (thus preventing comparisons among studies of listener consistency across experiments). Nevertheless, to the extent that comparisons can be made across studies, the results appear relatively consistent. For example, using both a single-burst paradigm and a four-burst para-

A further issue that has not been addressed by the above experiments concerns the extent to which the observed releases from masking caused by the reductions of target-masker similarity in the various experiments would have occurred even if there had been no uncertainty in the masker. It has been implied implicitly by our use of the phrase ''combating uncertainty'' that if there were no uncertainty, there would be no nonenergetic masking for the decrease in target-masker similarity to combat. However, it is possible that even if the masker uncertainty had been totally eliminated, the decrease in target-masker similarity in going from condition S to condition D would have caused significant release from masking. In order to adequately explore this issue, it would have been necessary to measure thresholds for the S and D conditions in each experiment for all frozen exemplars of the random masker. To the extent that the results of this additional (massive) set of experiments showed a clear release from masking in going from condition S to condition D (and this release from masking were of sufficient magnitude to rule out explanations in terms of possible changes in energetic masking that might have occurred in some of the frozen cases in going from S to D), one would be forced either to define informational masking so as to include ef-



where L denotes the listener and can assume any of the values L1,L2,...,L5; E denotes the experiment and can assume any of the values Duration, Sweep, Spatial, MBS, or MBD; M(L,E) denotes the amount of masking (in dB) for listener L and experiment E (as shown in Fig. 5); $\overline{M(L,E)}^{L}$ denotes the average of M(L,E) over L (the group mean profile shown in Fig. 4); $\overline{M(L,E)}^{E}$ denotes the average of M(L,E) over E (as reported in Table I); and $\overline{M(L,E)}^{L,E}$ denotes the average of M(L,E) over both L and E (the grand mean of all the data as reported in the last column of Table I).

Note that by collecting and rearranging terms, Eq. (A1) can be rewritten simply as

$$M(L,E)=\overline{M(L,E)}^{L}+\overline{M(L,E)}^{E}-\overline{M(L,E)}^{L,E}. \tag{A2}$$

The relationship described by Eqs. (A1) and (A2) assumes that the results for listener L can be estimated by adding $\overline{M(L,E)}^{E}$ (a constant for each value of L) to the group-mean profile $\overline{M(L,E)}^{L}$, normalized by the overall group mean $\overline{M(L,E)}^{L,E}$. Note, furthermore, that equations (A1) and (A2) perfectly describe the data both when performance is the same for all listeners [because then $M(L,E)=\overline{M(L,E)}^{L}$ and $\overline{M(L,E)}^{E}=\overline{M(L,E)}^{L,E}$ for all L and E] and when performance is the same for all experiments [because then $M(L,E)=\overline{M(L,E)}^{E}$ and $\overline{M(L,E)}^{L}=\overline{M(L,E)}^{L,E}$ for all L and E].

In order to evaluate the extent to which Eq. (A2) represents the data for both the S and D conditions, the rms deviation between the predicted values of M(L,E) and the measured values of M(L,E) was computed (separately for S and D conditions). The results of this computation, included in Table III, show that the rms deviation for the S condition is 7.8 dB and the rms deviation for the D condition is 4.8 dB. If instead of using Eq. (A2) to estimate M(L,E), we used simply

$$M(L,E)=\overline{M(L,E)}^{L}, \tag{A3}$$

i.e., we ignored subject differences and just used the group-mean profile to estimate M(L,E), then the rms deviations (also shown in Table III) would have been 10.8 dB for the S condition and 5.3 dB for the D condition. Although in an absolute sense, the rms deviation between data and predictions is larger in the S condition than in the D condition, subject differences account for a larger percentage of the variation in the S condition [$(10.8-7.8)$ dB out of 10.8 dB or 28%] than in the D condition [$(5.3-4.8)$ dB out of 5.3 dB or 8.6%].

An alternative way to compare Eqs. (A2) and (A3) is to calculate the correlations between the predicted and actual results in each case and determine the percentage of variation in the data for which the model accounts. These calculations (see Table III) show that 24% of the variance is accounted for in the S condition and 18% for the D condition when the mean alone is used [Eq. (A3)], but that these values increase to 61% in the S condition and 33% in the D condition when a listener-specific term is included in the predictions [Eq. (A2)]. Thus, incorporating knowledge of listener identity explains 37% more of the variance for the S condition, but only an additional 15% of the variance in the D condition (compared to using only knowledge of the experiment).

While these analyses suggest that knowledge of listener identity improves prediction accuracy, Eq. (A2) has more degrees of freedom than Eq. (A3); thus it is not a "fair" comparison. Even if data points for each experiment are randomly assigned to "pseudo-listeners" [rather than grouping the data by actual listeners in calculating $\overline{M(L,E)}^{E}$], the rms deviation will always decrease using the more-complex model [Eq. (A2)] compared to the experiment-only model [Eq. (A3)]. In order to obtain better insight into this issue, a boot-strapping method was used to determine the extent to which, for the data points we were fitting, the observed improvements in the model predictions is more likely to arise by chance than from the actual listener-specific characteristics. More specifically, in order to assess whether these improvements are better than expected by chance, we determined how often random permutations of the measured data lead to better predictions than the predictions based on grouping the data by listener. In other words, predictions using Eq. (A2) were compared to the results obtained when the correspondence between listeners and the measured values of M(L,E) were randomized. In this analysis, we (a) constructed results for randomized pseudo listeners by randomizing the correspondence between L and M(L,E) (subject only to the constraint that the experiment E was held fixed in the randomization); (b) calculated the rms deviation for each such randomization in the same manner as described previously; (c) performed 10 000 such randomizations and rms-deviation computations; and (d) used these results to estimate the probability density of the rms deviations for these randomized pseudo listeners.

The results of this analysis are shown in Fig. 6. For each of the conditions S and D, the figure shows both the rms deviation obtained with the real listeners (represented by the dashed vertical lines) and the estimated probability density of the rms deviation for the pseudo listeners.[4]

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

Neff, D. L., and Callaghan, B. P. (**1987**). *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 37–46.

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

Neff, D. L., and Dethlefs, T. M. (**1995**). ''Individual differences in simultaneous masking with random-frequency, multicomponent maskers,'' J. Acoust. Soc. Am. **98**, 125–134.

Neff, D. L., Dethlefs, T. M., and Jesteadt, W. (**1993**). ''Informational masking for multicomponent maskers with spectral gaps,'' J. Acoust. Soc. Am. **94**, 3112–3126.

Oh, E. L., and Lutfi, R. A. (**1998**). ''Nonmonotonicity of informational masking,'' J. Acoust. Soc. Am. **104**~