The information-divergence hypothesis of informational masking

Robert A. Lutfi,^{a)} Lynn Gilbertson, Inseok Heo, An-Chieh Chang, and Jacob Stamas

Auditory Behavioral Research Lab, Department of Communication Sciences and Disorders,
University of Wisconsin, Madison, Wisconsin 53706

(Received 8 March 2013; revised 15 July 2013; accepted 19 July 2013)

In recent years there has been growing interest in masking that cannot be attributed to interactions in the cochlea—so-called informational masking (IM). *Similarity* in the acoustic properties of target and masker and *uncertainty* regarding the masker are the two major factors identified with IM. These factors involve quite different manipulations of signals and are believed to entail fundamentally different processes resulting in IM. Here, however, evidence is presented that these factors affect IM through their mutual influence on a single factor—the *information divergence* of target and masker given by Simpson–Fitter’s d_a [Lutfi *et al.* (2012). *J. Acoust. Soc. Am.* **132**, EL109–113]. Four experiments are described involving multitone pattern discrimination, multi-talker word recognition, sound-source identification, and sound localization. In each case standard manipulations of masker uncertainty and target-masker similarity (including the covariation of target-masker frequencies) are found to have the same effect on performance provided they produce the same change in d_a . The function relating d' performance to d_a , moreover, appears to be linear with constant slope across listeners. The overriding dependence of IM on d_a is taken to reflect a general principle of perception that exploits differences in the statistical structure of signals to separate figure from ground.

© 2013 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4817875]

PACS number(s): 43.66.Dc, 43.66.Ba, 43.66.Fe, 43.66.Qp [EB]

Pages: 2160–2170

I. INTRODUCTION

Over the past two decades there has been an explosion of studies investigating what is popularly termed *informational masking* (IM) (see Lutfi, 1993; Kidd *et al.*, 2008, for reviews). Although definitions vary, the term is used in practice to identify masking that has no obvious explanation in known processes occurring at the auditory periphery (i.e., at the level of the cochlea or auditory nerve). Figure 1 makes the analogy to vision. In the right panel of this figure, one clearly sees a pattern of a repeating orange square (target). In the left panel the pattern continues to be present, but because of the different context in which it occurs (masker), it is not nearly as apparent. Notably, in neither panel does the context overlap with or occlude the orange squares in any way; hence, the effect can reasonably be said to originate at some central level of the visual system beyond the retina. In the vision literature this is referred to as *pattern masking* (Turvey, 1973); it is the direct counterpart to IM in audition. Replace the squares of different colors with tones of different frequencies or the babble of different talkers and the result is fundamentally the same. In the broader context, IM has been linked to the phenomena of auditory object formation (Kubovy and Valkenburg, 2001), sound source segregation (Micheyl and Oxenham, 2010), auditory scene analysis (Bregman, 1990), and the cocktail party listening effect (Cherry, 1953).

Research on IM has a long history dating back to the seminal studies of Pollack (1975) and Watson *et al.* (1975);

Pollack having originated the term. The early research undertook to identify and document the effects of relevant acoustic parameters and their interactions and that effort largely continues today. What has emerged is a distinction between two general categories of effects believed to have fundamentally different causes. The first is associated with *uncertainty* regarding the masker, created when one or more acoustic properties of the maskers vary unpredictably (at random) from trial to trial (Note by comparison the context of random colored squares in Fig. 1). The most recent work on this factor has focused on the effects of uncertainty associated with talker voice and location on multi-talker word recognition (Arbogast *et al.*, 2005; Brungart, 2001; Brungart and Simpson, 2004; Brungart *et al.*, 2009; Freyman *et al.*, 1999; Kidd *et al.*, 2005; Kidd *et al.*, 2008). However, by far, the most data exist for the discrimination of multi-tone patterns. Here the detrimental effects of uncertainty associated with the frequencies, levels, durations, and timing of tones have all been well documented, as have their interactions with the number of tones in the pattern, their relative level and duration, the number and position of tones identified as targets and the type of discrimination task (see Lutfi, 1993; Kidd *et al.*, 2008, for reviews). Some success, moreover, has been made in predicting the results of the tonal pattern studies using a measure of *relative entropy* from information theory to quantify masker uncertainty (Lutfi, 1993; Lutfi and Doherty, 1994; Oh and Lutfi, 1998, 1999).

The second class of effects associated with IM relates to the acoustic *similarity* between target and masker (Compare to the same versus different geometries of target and context elements in Fig. 1). This category includes cases wherein target and masker are temporally coherent (Micheyl *et al.*,

^{a)}Author to whom correspondence should be addressed. Electronic mail: ralutfi@wisc.edu

FIG. 1. Demonstration of visual pattern masking as the counterpart to auditory informational masking.

2010), are harmonically related (Micheyl *et al.*, 2010; Oh and Lutfi, 2000), are in close spatial proximity to one another (Arbogast *et al.*, 2005; Martin *et al.*, 2012), are narrowly separated in frequency (Watson *et al.*, 1975; Bregman, 1990, p. 18) or, as in the case of different talkers, have qualitatively similar voices (Brungart, 2001; Kidd *et al.*, 1994; Kidd *et al.*, 2008). In the broader context, target-masker similarity has also been identified with conditions for which target and masker covary in frequency over time (Durlach *et al.*, 2003b; Kidd *et al.*, 1994; Kidd *et al.*, 2002). In this latter case, the effects of target-masker similarity have been interpreted in terms of the Gestalt principle of “common fate” (Bregman, 1990; Durlach *et al.*, 2003b; Kidd *et al.*, 1994).

A serious appraisal of the current state of research on IM must conclude that it is at a critical juncture. Four decades of work have documented many results but have yielded little in the way of an overall conceptual framework, working principle or model for interpreting these results and guiding future research. Many, in fact, now question whether the term IM continues to have any useful meaning, given that it has become a label for just about any form of masking not readily explained by known peripheral processes (Watson, 2005; Durlach, 2006; Durlach *et al.*, 2003a). In this regard, Watson (2005) has described IM as a “suitcase term”; a label for “a number of loosely related, but arguably different phenomena [that] includes too many cause-effect relationships to belong to a single theoretical framework.” Watson goes on to suggest that “similarity- and uncertainty-based shifts in threshold are the result of quite different processing inefficiencies in the nervous system. It is unlikely that a single theory of IM will provide a satisfactory account of both forms of loss of information.”

Watson’s assessment seems to represent the current consensus on IM, but it may be premature. The present paper provides evidence for a different view. It proposes that similarity and uncertainty effects associated with IM can be understood to reflect a general principle of perception that recognizes differences in the statistical structure of signals as vital information for separating figure from ground (Attneave, 1954; Barlow, 1961; Bregman, 1990). Applying this principle to IM, we speculate that IM will depend singularly on the *information divergence* between target and masker—a measure of statistical difference between two information sources used in information theory (Kullback and Liebler, 1951). The information-divergence hypothesis, as

we shall call it, draws together IM phenomena by forcing a fundamentally different way of thinking about IM. It redefines what constitutes a signal for the listener, from a time-waveform that varies along one or more different acoustic dimensions, to an ensemble of waveforms having a unique set of statistical properties. The statistics, not the acoustics, is where one finds the common link among these phenomena. In what follows, a series of experiments is undertaken involving a variety of different psychophysical tasks (multi-tone pattern discrimination, word recognition, sound source identification, and sound localization). The basic phenomena associated with IM (uncertainty, similarity, and frequency-covariation effects) are replicated in these tasks. Support for the information-divergence hypothesis is then given by the finding of a common functional relation across experiments between listener performance and a single measure of information divergence.

II. UNCERTAINTY, SIMILARITY, AND INFORMATION DIVERGENCE

In this section we describe the methodological framework and divergence metric used in experiments to test the information-divergence hypothesis. The approach represents an application of a larger theoretical framework recently developed to deal with vagaries of the terms “uncertainty” and “similarity” as they have been used in the literature (Lutfi *et al.*, 2012). Before proceeding, it will be helpful to consider how the hypothesis manages predictions for two very different masking scenarios. We compare the studies of Kidd *et al.* (2008) and Brungart (2001). Both studies required listeners to recognize words in a target phrase masked by a masker phrase spoken by another talker. In the Kidd *et al.* (2008) study, a detrimental effect on performance was observed when the spatial location of the masker phrase was selected at random from trial to trial (spatial uncertainty effect). In the Brungart study, a detrimental effect on performance was observed when the target and masker phrases were spoken by talkers of the same gender (voice similarity effect). The two studies are representative of several others that have shown spatial location and voice quality of talkers to be major factors involved in IM and the “cocktail party listening” effect.

Now, let us suppose that the perceived differences in voice in the Brungart study roughly equate to differences in fundamental frequency, F0 (not an entirely unreasonable assumption). Conceivably, one could then measure probability density functions (PDFs) of F0 that specify the statistical differences in voice for each talker (mostly average differences associated with prosody, intonation, and gender). Similarly, one could determine PDFs in the Kidd *et al.* study that specify the statistical variation in the spatial locations of each talker. Let $p(x)$ and $q(x)$ denote these PDFs for the target and masker in each case, x being either F0 or spatial location. The information-divergence hypothesis maintains that, through repeated observations of x , the listener “picks up” information about p and q and uses this information to distinguish between the target and masker phrases. The prediction is that listener performance will depend, not on the particular acoustic parameter associated with x , but on the magnitude of

the difference between p and q . The acoustics are of little consequence in this prediction except when they exceed the peripheral ear's capacity to resolve differences in p and q ; cases by definition that would not be associated with IM.

In information theory, the measure that quantifies the statistical differences between an arbitrary pair of PDFs, as given by p and q , is information divergence, also known as Kullback–Leibler divergence, discrimination information or relative entropy (Kullback and Liebler, 1951). The measure is defined as the expected value of the log-likelihood ratio of x under p and q ,

$$D_{\text{KL}}(p||q) = \int_x p(x) \ln \left[\frac{p(x)}{q(x)} \right] dx = E_p \ln \left[\frac{p(X)}{q(X)} \right]. \quad (1)$$

In general applications, D_{KL} is often interpreted as the *discriminable* difference between two information sources (target and masker in the present case).¹ This is easily seen for the special case in which p and q have equal-variate, normal densities, $\mathcal{N}_p$ and $\mathcal{N}_q$. Here, D_{KL} reduces to the more familiar index of discriminability,

$$D_{\text{KL}}(\mathcal{N}_p||\mathcal{N}_q) = d'_x = \frac{E[X|p] - E[X|q]}{SD[X]}. \quad (2)$$

The desirable properties of D_{KL} are that it has no physical unit and depends only on the statistical properties of x . It thus provides a measure of discriminability regardless of the physical dimensions along which x might vary. In the present example, this allows meaningful comparisons to be made between such different influences as uncertainty regarding talker location and similarity in talker voice. Returning to the information-divergence hypothesis, the prediction is that, all other influences being equal, the factor causing the greatest change in D_{KL} will be the one to have the greatest influence on performance and, conversely, that factors producing the same change in D_{KL} will have identical influence on performance.

The above example illustrates in principle how predictions are to be made in the present study. In practice, there are only two differences. First, in the present experiments the densities p and q are known exactly rather than measured, they are determined procedurally to be normal with prespecified means, variances, and covariances. This greatly simplifies the predictions for each experiment by allowing divergence to be expressed in terms of these known quantities. Second, for our initial application, we chose a metric of divergence closely related to D_{KL} , but more commonly used in the psychoacoustic literature. The metric is Simpson–Fitter's d_a ,

$$D_{\text{KL}}(\mathcal{N}_p||\mathcal{N}_q) \approx d_a = \frac{\mu_T - \mu_M}{\sqrt{(\sigma_T^2 + \sigma_M^2)/2}}, \quad (3)$$

where μ_T and μ_M are the parameter means of target and masker, σ_T^2 and σ_M^2 are their variances and $|\mu_M - \mu_T| \geq |\sigma_T - \sigma_M|$.² Our reason for choosing d_a initially is that it is an ideal statistic for divergence computed on simple differences in x ; d_a in this case being proportional to the normal transform of the area under the receiver operating characteristic (Simpson and Fitter, 1973). Later we consider the case where divergence may

include higher moments of x requiring the more general expression for D_{KL} given by Eq. (1).

Now Eq. (3) is for the case wherein each presentation of a stimulus provides, in effect, a single independent observation from p and q (i.e., the values of x for target and masker are fixed within the presentation). For the case in which each presentation yields multiple observations from p and q we assume that the observations are optimally combined. In past studies involving such multiple observations, the number of observations from p is often equal to that of q , corresponding to a fixed number N of target-masker pairs [cf. the Kidd *et al.* (2008) study described above]. In such cases the optimal combination rule is

$$d_{aN} = \sqrt{1 + (N - 1)(1 - r^2)} \cdot d_a, \quad (4)$$

where r is the Pearson product-moment correlation between observations. Here the term under the radical can be thought of as the number of *independent* observations of the difference between p and q ; one independent observation for $r = 1$ (covariation case), and N for $r = 0$.

Taken together, Eqs. (3) and (4) serve to titrate the effects of target-masker uncertainty, similarity and covariation into the influence of a single critical variable, d_{aN} . Each quantity that enters into the computation of d_{aN} is associated with one of these three factors consistent with the way they have been conceptualized in the literature; stimulus uncertainty with an increase in the variance terms σ_T^2 and σ_M^2 , target-masker similarity with a decrease in the difference of means, $\mu_T - \mu_M$ and target-masker covariation with a value of $r > 0$. The question for the remainder of this paper is whether performance will bear the same functional relation to d_a when these quantities are chosen to yield the same values of d_a across different conditions and different listening tasks.

III. TESTS OF THE INFORMATION-DIVERGENCE HYPOTHESIS

A. Multitone pattern discrimination

The first case considered involves the discrimination of multitone patterns differing in frequency. The experiment is similar to the vast majority of multitone pattern discrimination studies, which has focused on frequency differences among patterns. Target and masker patterns were each a sequence of $N = 5$ tone bursts. The tone bursts were 50 ms in duration and gated on and off with 10-ms, cosine-squared ramps. So as to minimize interactions in the cochlea, target and masker bursts alternated in time without silent intervals between them. The frequencies of the target tones were selected independently and at random on each presentation from normal distributions with mean $\mu_T = 1000$ Hz and standard deviation $\sigma_T = 25$ Hz. In the two-interval, forced-choice procedure, the listener's task was to detect an increment in the target tone frequencies averaging 25 Hz across trials. Correct feedback was given after each trial.

Three masker conditions were investigated as shown in Fig. 2. For the masker uncertainty and similarity conditions, the frequencies of the masker tones were selected independently

FIG. 2. Stimulus configurations for multitone pattern discrimination experiment. Target and masker are sequences of $N=5$ alternating pure tones, where the individual tones vary in frequency from trial to trial. The statistical distributions of frequencies are given by the continuous curves drawn to the left of each panel (black for target, gray for masker). Reading left to right: Target-masker similarity condition (small $\mu_T - \mu_M$), frequency covariance condition ($r=1$), masker uncertainty condition (large σ_M). Values of $\mu_T - \mu_M$, r , and σ_M are chosen in each condition to produce the same value of d_{aN} .

and at random on each presentation. The variance in masker frequencies was chosen to be comparably large (uncertainty condition) or the mean difference between the masker and target frequencies was chosen to be comparably small (similarity condition). These two conditions are representative, generally, of the many IM studies for which the focus has been on the frequency uncertainty associated with masker tones (e.g., Neff and Green, 1987; Watson, *et al.*, 1976; Lutfi, 1993; Kidd *et al.*, 2008). The similarity condition is also representative of studies where small frequency separations are found to cause perceptual fusion of melodic patterns (Bregman, 1990, p.18), an effect sometimes implicated in IM. For the third masker condition (covariation condition), the frequency of the first masker tone in the sequence was selected at random while the frequencies of the remaining $N-1$ masker tones varied to maintain the same frequency separation from the target as the first masker tone. This condition is comparable to those of other IM studies where target and masker tones are constructed to share the same pitch contour (e.g., Durlach *et al.*, 2003). In all three masker conditions the values of μ_M , σ_M , and r associated with the masker frequencies were chosen to yield the same value of d_a . For the masker uncertainty condition these values respectively were 2236 Hz, 75 Hz, and 0; for the similarity condition they were 1553 Hz, 25 Hz, and 0, and for the covariation condition they were 2236 Hz, 25 Hz, and 1.

All sounds were played at a 44 100-Hz sampling rate with 16-bit resolution using a mark of the unicorn audio interface. From the interface the sounds were buffered through a Rolls RA62c headphone amplifier and then delivered diotically over Beyerdynamic DT 990 headphones to listeners seated individually in a double-walled, industrial

acoustics sound-attenuated chamber sound-attenuated chamber. A loudness balancing procedure was used to calibrate the overall sound power to be approximately 70 dB sound pressure level (SPL) at the eardrum (Lutfi *et al.*, 2008). Listeners were 3 male and 13 female students at the University of Wisconsin–Madison ranging in age from 19–23 yr. They were paid at an hourly rate for their participation. The results of a standard hearing evaluation showed all listeners to have normal hearing sensitivity of 15 dB hearing level (HL) or better from 250 Hz to 8 kHz (ANSI S3.6-2004). All listeners also performed perfectly on a pre-data collection block of 15 trials with target presented alone.

Figure 3 shows the pairwise comparisons of the percent correct reduction in performance from the pre-data collection block for the three masker conditions. Each point represents the performance reduction of a single listener averaged over 450 trials. The data indicate somewhat less IM than predicted in the covariance condition compared to the similarity and uncertainty conditions. This is likely due to the inability of listeners in the similarity and uncertainty conditions to benefit from more than a few independent observations of the difference between target and masker on each trial. In comparable conditions involving frequency discrimination of target-tone sequences without maskers, listeners have been found to benefit from only 2–3 independent observations (Lutfi, 1990). Notwithstanding, in light of the large performance differences across listeners, there is remarkable agreement in IM across the three masker conditions for each listener. Consistent with the information-divergence hypothesis, the results clearly show a greater dependence of IM on the value of d_a than on any of the specific manipulations

FIG. 3. Pairwise comparisons of the reduction in performance from the no-masker condition for the three masker conditions of the multitone pattern discrimination task. Each symbol represents the reduction in performance for a single listener.

associated with masker uncertainty, similarity, and frequency covariation.

B. Word recognition

The second task considered was word recognition. The specific conditions were selected to be roughly similar to those of the Kidd *et al.* (2008) and Brungart (2001) studies described earlier (also see Brungart and Simpson, 2002; Brungart *et al.*, 2001). The target and masker were each a sequence of $N=3$ words selected at random on each trial from the Maryland consonant-nucleus-consonant (CNC) word lists, 500 words altogether from a native English-speaking male (Peterson and Lehiste, 1962). Using MATLAB version 7.0.1, the recordings were edited to eliminate the indicator phrase “Ready,” leaving only approximately 50 ms of silence preceding and following each recorded word. In different conditions the temporal overlap of target and masker words was either 0, 0.1, or 0.2 s. Within each block of trials the fundamental frequency (F0) of words was varied at random from one presentation to the next. F0 was varied using the synchronized overlap-add, fixed-synthesis algorithm of Henja and Musicus (1991) to maintain the natural duration of the words. Across all conditions the mean and standard deviation of the F0s for the target words were fixed at $\mu_T = 120$ Hz and $\sigma_T = 15$ Hz, respectively. The listener’s task was to recognize as many of the three target words as possible on each trial. At the end of each trial, three columns of words corresponding to the three targets in order of their position in the sequence appeared on the listener’s computer monitor. Each column contained the correct target word corresponding to the correct temporal position for that trial, as well as 9 foils randomly selected from the Maryland list. The listener highlighted their choice of the correct target word in each column by pointing and clicking on a single word in each column. Each choice was identified as correct or incorrect only after the listener had made a selection in all three columns.

Three masker conditions were examined, as in the first experiment. For the masker uncertainty condition the variability of masker F0s, given by σ_M , was made comparably large, much like the random-voice condition of Kidd *et al.* (2008). For the target-masker similarity condition the mean difference between the target and masker F0s, given by $\mu_T - \mu_M$, was made comparably small, much like the same talker condition in the study by Brungart (2001). Finally, in the frequency-covariation condition the F0s of target and masker words covaried the same way as did the frequencies of target and masker tones in the first experiment. The covariation condition permitted comparison to results of the first experiment, but otherwise the authors know of no studies examining comparable conditions using speech. As before, the values of μ_M , σ_M , and r were chosen to yield the same value of d_a for each condition. For the masker uncertainty condition they were, respectively, 170 Hz, 33.5 Hz, and 0 for the target-masker similarity condition they were 149 Hz, 15 Hz, and 0, and for target-masker covariation condition they were 170 Hz, 15 Hz, and 1. These values were selected to be roughly representative of the means and standard

deviation of F0s for adult males, adult females, and children (Fitch and Holbrook, 1970; Peterson and Barney, 1952; Keating and Buhr, 1978). The apparatus and method for generating sounds and presenting them to listeners were identical to the first experiment described, as were the procedure for screening listeners. Listeners were 1 male and 8 female students at the University of Wisconsin–Madison, none of whom had participated in any of the other experiments. Their ages ranged from 20–27 yr. Baseline performance in a no-masker condition was measured for each listener prior to collecting data for the three masker conditions.

Figure 4 shows the pairwise comparisons of the reduction in percent correct performance from baseline resulting from the addition of the masker in the three masker conditions. Each point represents the performance reduction for a single listener averaged over 90 trials each for the masker and no-masker conditions. Once again, the data show large individual differences in performance across listeners. For some listeners increasing the temporal overlap of words has a deleterious effect, while for others it has little or no effect. For all listeners, however, performance is remarkably similar across the three masker conditions, consistent with the outcome of the first experiment. The results lend further support to the information-divergence hypothesis by showing the same dependence of IM on d_a for fundamentally different stimuli and psychophysical task.

C. Sound source identification

The foregoing two experiments provide preliminary support for the information-divergence hypothesis in showing the same performance for the same value of d_a across different masker conditions. The next experiment was undertaken to test the stronger prediction that the *function* relating performance to d_a will be the same across different masker conditions. A sound source identification task was chosen similar to that described in the studies of Lutfi and Liu (2007, 2011) and Lutfi *et al.* (2008). Sound source identification studies attempt to bridge the gap between pure-tone discrimination and word recognition tasks by using the impact sounds of simple resonant sources to maintain some degree of real-world naturalness without giving up a large measure of stimulus control.

In the present experiment, the target and masker were synthesized impact sounds of a loosely suspended circular plate (details of the synthesis can be found in Lutfi and Liu, 2011). The impact sounds consisted of a sum of exponentially damped sinusoids with a decay modulus of 50 ms and frequencies $f_0 \times [1.00 \ 2.80 \ 5.15 \ 5.98 \ 9.75 \ 14.09]$. The values of f_0 for target and masker were selected independently and at random on each presentation, as would correspond to changes in the size of the plate. The exact values were such that $f_0 = \mu + z\sigma$, where μ and σ are the mean and standard deviation of target or masker f_0 s and z is a randomly sampled normal deviate. For the target, $\mu = \mu_T$ was 500 Hz and $\sigma = \sigma_T$ was 25 Hz. For the masker $\mu = \mu_M$ and $\sigma = \sigma_M$ were varied in two ways. In the first, the effect of target-masker similarity was measured by fixing σ_M at 25 Hz and varying μ_M from 550 to 900 Hz. In the second, the effect of

FIG. 4. Pairwise comparisons of the reduction in performance from the no masker condition for the three masker conditions of the word recognition task. Each symbol represents the reduction in performance for a single listener. Data are given for three degrees of temporal overlap of target and masker words (panel rows).

masker uncertainty was measured by fixing μ_M at 900 Hz and varying σ_M from 25 to 282 Hz. The values of the varied parameter were chosen to yield the same values of d_a . The total duration of each impact sound was 200 ms, time enough for the sounds to decay to silence. An instance of the masker immediately preceded and followed the target on each trial, same f_0 for each masker. In the two-interval, forced-choice procedure, the listener's task was to judge which of the two intervals contained the target sound corresponding to the larger-sized (lower f_0) plate. Correct feedback was given after each response. The apparatus and method for generating

sounds and presenting them to listeners were identical to the first experiment described. Listeners were 1 male and 4 female students at the University of Wisconsin–Madison ranging in age from 20–24 yr. Data collection from two other listeners was terminated after they showed near perfect performance in initially run masker conditions. Listeners were paid at an hourly rate for their participation. All had normal hearing with auditory thresholds from 250 to 8000 Hz equal or less than 20 dB HL, and all performed perfectly or near perfectly on a pre-data collection block of 15 trials with target presented alone.

FIG. 5. Performance as d' is plotted as a function of d_a for the 5 listeners (panel columns) participating in the sound source identification task. Error bars give the standard deviation of the d' values obtained in six, 50-trial blocks. The lower row of panels gives the data for the condition in which $\mu_T - \mu_M$ is varied; the upper row of panels give the data for the condition in which σ_M is varied. The curves drawn through the data give the results of a linear regression for which the slope of the best-fitting curve was forced to be the same across all listeners and all conditions. The intercepts were allowed to vary across listeners but were forced to be the same across conditions within each listener. The values of slope (a) intercept (b), and variance accounted for (r^2) are given at the top of each panel.

Figure 5 gives d' performance as a function of d_a for the five listeners participating in the experiment (panel columns). The two masker conditions are given by the panel rows (variable σ_M , top panel, variable $\mu_T - \mu_M$, bottom panel). Each data point represents the average of 300 trials. Error bars give the standard deviation of the d' values obtained in the corresponding six, 50-trial blocks. Repeated data points are replications of the first conditions run. The curves drawn through the data give the results of a linear regression for which the slope of the best-fitting curve was forced to be the same across all listeners and all conditions. The intercept was allowed to vary across individuals but was forced to be the same within listeners across the two conditions. The fits across listeners are generally quite good, the one exception being that for listener TDT for the variable σ_M condition. Even in this case, the slope would not be expected to deviate much from the fitted value were it allowed to vary independently of the other listeners. The results of the regression suggest that the rate at which performance improves with increases in d_a is quite similar across listeners, this despite significant differences in the overall performance of listeners. The results once again show a greater dependence of performance on the value of d_a than on the specific manipulations designed to increase masker uncertainty or target-masker similarity.

D. Sound localization

The final task in the series was sound localization. The target and masker were brief Gaussian noise bursts whose perceived location on the azimuthal plane was varied using KEMAR HRTFs. The synthesized locations for target and masker were normally distributed as azimuthal angle, with mean and standard deviation denoted μ and σ as before. For the target, μ_T was fixed at 0° and σ_T was fixed at 10° . Two masker conditions were investigated as shown in Fig. 6. For the target-masker proximity condition (left panel) σ_M was fixed at 20° and μ_M took on values of 10° , 20° , or 30° . For

the masker uncertainty condition (right panel) μ_M was fixed at 30° and σ_M took on values of 10° , 15° , 20° , or 25° . Because the masker angles were constrained between $\pm 90^\circ$, the values of d_a differed for the two conditions. Noise bursts were played in sequence as masker-target-masker triads, the two masker bursts having the same location within a presentation. All noise bursts were 100 ms in duration and were gated on and off with 10-ms cosine-squared ramps, the inter-burst interval was 100 ms. The apparatus for playing sounds and presenting them to listeners was identical to that of the other experiments in this series. In the two-interval, forced-choice procedure, the listeners task was to judge whether the target moved from right to left or left to right across the two observation intervals. Correct feedback was given after each response. Listeners were 4 male and 4 female students at the University of Wisconsin–Madison ranging in age from 19–22 yr. Two other listeners were not included in the experiment as their performance failed to improve much from chance with additional training. All listeners had the same screening as described in the first experiment and all had participated in one or more of the previous experiments.

FIG. 6. Stimulus configurations for sound localization experiment. Target and masker are broadband Gaussian noise bursts whose spatial locations vary at random from one presentation to the next given by normal distributions (black-continuous for target, gray-dashed for masker). Left panel: target and masker locations are made more 'similar' by decreasing the mean spatial separation between target and masker, $\mu_T - \mu_M$. Right panel: the masker location is made more uncertain by increasing the standard deviation of the masker location, σ_M .

FIG. 7. Performance as d' is plotted as a function of d_a for the 8 listeners (panels) participating in the sound localization task. Error bars give the standard deviation of the d' values obtained in the corresponding six, 50-trial blocks. To constrain azimuthal angles between -90° and 90° , different values of d_a were required for the two masker conditions. For the variable σ_M (fixed μ_M) condition the values of d_a were 1.6, 1.9, 2.3, and 3.0. For the variable μ_M (fixed σ_M) condition the values. The curves drawn through the data give the results of a linear regression for which the slope of the best-fitting curve was forced to be the same across all listeners. The values of slope (a), intercept (b), and variance accounted for r^2 are given at the top of each panel.

Figure 7 gives d' performance as a function of d_a for the eight listeners participating in the experiment (panels). Each point represents the average of 300 trials. Error bars give the standard deviation of the d' values obtained in the corresponding six, 50-trial blocks. Not all listeners were able to participate in the condition for which $\mu_M = 30^\circ$ and $\sigma_M = 25^\circ$; hence, only five data points are shown for these listeners. Once again, the curves drawn through the data give the results of a linear regression for which the slope of the best-fitting curve was forced to be the same across all listeners, only the intercept was allowed to vary across individuals. The best-fitting value of the slope differs from that of the sound source identification experiment. This, however, is to be expected given the vast differences between the two experiments in task, stimuli and manner of presentation. We return to the issue of cross-study comparisons in the discussion. At present, the more significant point is that the fitted curves continue to suggest good agreement across listeners in the rate at which performance improves with d_a , consistent with the information-divergence hypothesis.

IV. DISCUSSION

This paper began with the premise that IM can be understood as a manifestation of a single general principle of perception that extracts figure (target) from ground (masker) based on differences in their statistical structure. The information-divergence hypothesis was proposed as a mathematical expression of this principle as applies to IM. The

results of the present study provide preliminary support for this hypothesis. They demonstrate a singular dependence of IM on the statistical divergence between target and masker as measured by Simpson–Fitter’s d_a . Standard manipulations of masker uncertainty and target–masker similarity (including the covariance of target and masker frequencies) are found to have the same effect on performance provided they produce the same change in d_a . This dependence is replicated for common psychophysical tasks involving multitone pattern discrimination, multi-talker word recognition, sound source identification, and sound localization. The function relating d' performance to d_a , moreover, appears to be linear with constant slope across listeners; only the intercepts differing across listeners. Generally, the results argue against the popular suggestion in the literature that uncertainty and similarity-based effects associated with IM involve fundamentally different underlying mechanisms (cf. Durlach, 2006; Watson, 2005).

The idea that vital information for perception exists in the statistical structure of signals is, of course, hardly new. Historically, it has served as the foundation for both the information-theoretic view of perception as redundancy reduction (Attneave, 1954; Barlow, 1961) and the ecological view of perception as the extraction of structure in the world from lawful invariants (Gibson, 1966). It is even implicit in a number of Gestalt principles of perceptual grouping that depend on the statistical structure of signals, as, for example, the principle of common fate applied to time-covarying signals (Dau *et al.*, 2009; Durlach *et al.*, 2003b). What the

present work has done, in practice, is to fold these principles into a single metric describing a statistical structure that can be used to predict the effect of the major factors influencing IM and their interaction. In this regard, the approach mirrors an earlier effort to model uncertainty-based effects on IM, before similarity-based effects on IM had been widely published (Lutfi, 1992, 1993; Oh and Lutfi, 1998). Much like the present application, the component-relative-entropy (CoRE) model derives predictions for IM from the information divergence (relative entropy) of target and masker. The divergence, however, is computed on presumed *internal* representations of target and masker rather than their acoustic representations. This allows the model to predict a number of results that would not otherwise have been predicted based on the acoustic representation alone. Most notably among these is the nonmonotonic relation found between IM and the number of tones comprising a multitone masker (Oh and Lutfi, 1998), and the generally smaller amount of IM that is observed in listeners with sensorineural hearing loss (Alexander and Lutfi, 2004).

There also appear to be parallels between the present results and those of studies involving the integration of information from multiple auditory sources (IM being the complement of integration in that it requires the *differentiation* of information from multiple sources). One approach taken in these studies entails a procedure called sample discrimination (Lutfi, 1989, 1990). On each trial the listener is presented with an N -tone complex. In different conditions the N frequencies, durations or levels of the tones are selected independently and at random from one of two, equivariate, normal distributions differing in mean. The listener must decide on each trial whether tones were selected from the distribution with the high or low mean. The divergence of the parameter distributions in this case is given by $D_{KL} = d_{aN}$ and, much like the present IM data, performance shows the same dependence on divergence regardless of the particular parameter the listener is asked to judge (Lutfi, 1989, 1990).

Still further parallels are found in vision, where D_{KL} has served recently as the basis of several models of object saliency in scenes and video (Gao *et al.*, 2008; Mahadevan and Vasconcelos, 2010; Klein and Frintrop, 2011; Itti and Baldi, 2009). This work appears as the direct counterpart to the present work in that it uses D_{KL} to measure the saliency of objects in various contexts, in much the same way it is used in the present application to gauge the “saliency” of targets embedded in a masker. The study by Itti and Baldi (2009) is particularly noteworthy in this regard. Itti and Baldi (2009) find that the shift of visual gaze to different features of a scene is well predicted by the information divergence between the actual features of the scene and the observer’s expectations regarding those features. In audition, of course, there is nothing comparable to visual gaze. However, the success of the CoRE model in predicting many uncertainty-based effects of IM owes to the basic premise that listener attention is directed toward the most novel or unexpected elements in an auditory scene. Selectivity to novel acoustic events has also been found in the primary auditory cortex of the cat (Ulanovsky *et al.*, 2003) and in the caudal lateral

mesopallium the zebra finch (Gill *et al.*, 2008), the later study using a metric of novelty closely related to D_{KL} .

The parallels to vision, in particular, underscore the potential for the information divergence hypothesis to be tested under a variety of conditions approaching that of real-world listening. In a cocktail party setting, for example, lip-reading may facilitate an individual’s ability to segregate one talker’s speech from another (Sumbly and Pollack, 1954; Devergie *et al.*, 2011), as well as visual cues signaling when and where to listen in a room (Kong and Shinn-Cunningham, 2009). In such cases, the divergence hypothesis might be tested by manipulating, as covariance, the degree to which visual and acoustic cues for target and interfering speech separately agree. Evidence also suggests that segregation can be aided by linguistic differences between target and interfering speech (e.g., Festen and Plomp, 1990; Brouwer *et al.*, 2012). Here, a test of the hypothesis might be constructed by manipulating the probabilities (p and q) associated with different orders of word entropy for target and masker phrases (cf. Shannon, 1951); although in this case there would be no prediction for the possible effect of sequential word probabilities within target and masker phrases. Finally, in any realistic cocktail party setting a listener ultimately intends to understand the meaning of the of talker’s speech, the *message* the talker wishes to convey. Understanding meaning is clearly a greater challenge than recognizing individual words (the focus of the present and most past IM experiments); however, the approach to evaluating predictions of the hypothesis would proceed the same way. The hypothesis does not take into consideration what the listener intends to do with the target speech. It considers only how various properties of the target speech differ statistically from those of the speech that serves as interference. In this sense, the hypothesis is truly task independent.

Finally, we consider whether the present evidence for the information-divergence hypothesis significantly raises the prospect for developing a comprehensive model of IM. The hypothesis, it should be stressed, is not a model, it merely points to the difference in statistical properties of target and masker (given by p and q) as the critical factor underlying IM. To expand the hypothesis into a formal model, at least three important issues would need to be addressed. The first is to determine what conditions, if any, exist for which higher moments of p and q need to be included in the computation of divergence. In the vast majority of IM studies the properties of target and masker differ not only in their mean values, but also in their variances and possibly other higher moments. Such higher-moment statistics serve to differentiate target from masker in addition to the difference in means; however, the extent to which they are (or can be) used by listeners to segregate target from masker remains unknown. Some preliminary data relevant to this issue is provided by the unequal target-masker variance conditions of the present study. Here the strong dependence of listener performance on d_a (which ignores all higher moment statistics) suggests that listeners failed to use the difference in variance as a means to segregate target from masker. The outcome might be different in the exceptional case where target and masker can only be segregated based

on higher moments, i.e., where $\mu_T = \mu_M$ (cf. [Lutfi et al., 1996](#); [Viemeister et al., 2011](#)). Most early uncertainty-based studies of IM, in fact, come close to such conditions. In these cases the CoRE model, which assumes that judgments continue to be based on simple differences, has had good success in predicting the data (see [Lutfi, 1993](#); [Kidd et al., 2008](#), for reviews). The second issue is how one models the specific relation of performance to divergence. An important question here is whether the linear (affine) relation found in the present study generalizes broadly to other conditions. The task of modeling would be simplified if, indeed, it does and even more so if the slope of the function is found to be the same across different psychophysical tasks. As earlier noted, our conditions do not permit direct comparisons of slopes because of other stimulus differences across tasks; however, future studies, no doubt, should be able to implement the necessary controls that would make such comparisons possible. Last, a significant challenge for models is to provide some account of the large individual differences in performance that are so often observed in IM studies. The present results suggest some progress inasmuch as the rate at which performance improves with d_a appears to be constant across listeners; differences attributed almost entirely to differences in overall performance. It seems, therefore, that at least one significant component of IM, the rate at which performance improves with information divergence, is invariant across listeners.

ACKNOWLEDGMENTS

We would like to thank Dr. Emily Buss and two anonymous reviewers for helpful comments on an earlier version of this manuscript. This research was supported by a NIDCD grant R01 DC001262-20.

¹Note that D_{KL} is an asymmetric measure in that, except in special cases, $D_{KL}(p||q) \neq D_{KL}(q||p)$. A special case is when p and q are equal-variate normal, the precondition of Eq. (2). We have chosen the asymmetric form $D_{KL}(p||q)$ because the listener's task is to identify the target whose statistical properties are given by p .

²The correct expression for the normal, unequal variance case is

$$D_{KL}(p||q) = \frac{(\mu_M - \mu_T)^2 + \sigma_M^2 - \sigma_T^2}{2\sigma_T^2} + \ln\left(\frac{\sigma_T}{\sigma_M}\right).$$

This measure was found not to have as strong of a relation to performance as d_a for the conditions in which $\sigma_T \neq \sigma_M$. This suggests that listeners could not, or did not, take advantage of the difference in variance in this condition to segregate target from masker. The outcome might be expected to be different where $|\sigma_M - \sigma_T|$ is quite large relative to $|\mu_M - \mu_T|$, a case considered later in the discussion.

Alexander, J. M., and Lutfi, R. A. (2004). "Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights," *J. Acoust. Soc. Am.* **116**, 2234–2247.

ANSI S3.6-2004 Specification for audiometers, American National Standards Institute, New York.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (2005). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **117**, 2169–2180.

Attneave, F. (1954). "Some informational aspects of visual perception," *Psychol. Rev.* **61**, 183–193.

Barlow, H. B. (1961). "Possible principles underlying the transformation of sensory messages," *Sensory Commun.* 217–234.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA), pp. 1–773.

Brouwer, S., Van Engen, K. J., Calandruccio, L., and Bradlow, A. R. (2012). "Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content," *J. Acoust. Soc. Am.* **131**, 1449–1462.

Brungart, D. S. (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. (2009). "Multitalker speech perception with ideal time-frequency segregation: Effects of voice characteristics and number of talkers," *J. Acoust. Soc. Am.* **125**, 4006–4022.

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.

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

Dau, T., Ewert, S., and Oxenham, A. J. (2009). "Auditory stream formation affects comodulation masking release retroactively," *J. Acoust. Soc. Am.* **125**, 2182–2188.

Devergie, A., Grimault, N., Gaudrain, E., Healy, E. W., and Berthommier, F. (2011). "The effect of lip-reading on primary stream segregation," *J. Acoust. Soc. Am.* **130**, 283.

Durlach, N. I. (2006). "Auditory masking: Need for an improved conceptual structure," *J. Acoust. Soc. Am.* **120**, 1787–1790.

Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., Shin-Cunningham, B. (2003a). "Note on informational masking," *J. Acoust. Soc. Am.* **113**, 2984–2987.

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

Festen, J. M., and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *J. Acoust. Soc. Am.* **88**, 1725–1736.

Fitch, J., and Holbrook, A. (1970). "Modal vocal fundamental frequency of young adults," *Arch. Otolaryngol.* **92**, 379–382.

Freyman, R. L., Helfer, K. S., Mcall, 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.

Gao, D., Mahadevan, V., and Vasconcelos, N. (2008). "On the plausibility of the discriminant center-surround hypothesis for visual saliency," *J. Vision* **8**(7), 1–18.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems* (Houghton-Mifflin, Boston).

Gill, P., Woolley, S. M., Fremouw, T., and Theunissen, F. E. (2008). "What's that sound? Auditory area CLM encodes stimulus surprise, not intensity or intensity changes," *J. Neurophysiol.* **99**, 2809–2820.

Henja, D., and Musicus, B. R. (1991). "The SOLAFS time-scale modification algorithm," Bolt, Beranek and Newman (BBN) Technical Report.

Itti, L., and Baldi, P. (2009). "Bayesian surprise attracts human attention," *Vision Res.* **49**, 1295–1306.

Keating, P., and Buhr, R. (1978). "Fundamental frequency in the speech of infants and children," *J. Acoust. Soc. Am.* **63**(2), 567–571.

Kidd, G., Jr., Best, V., and Mason, C. R. (2008). "Listening to every other word: Examining the strength of linkage variables in forming streams of speech," *J. Acoust. Soc. Am.* **124**, 3795–3802.

Kidd, G., Jr., Mason, C. R., and Arbogast, T. L. (2002). "Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns," *J. Acoust. Soc. Am.* **111**, 1367–1376.

Kidd, G., Jr., Mason, C. R., and Deliwala, P. S. (1994). "Reducing informational masking by sound segregation," *J. Acoust. Soc. Am.* **95**, 3475–3480.

Kidd, G., Jr., Mason, C. R., and Gallun, F. J. (2005). "Combining energetic and informational masking for speech identification," *J. Acoust. Soc. Am.* **118**, 982–992.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2008). "Informational masking," in *Springer Handbook of Auditory Research: Auditory Perception of Sound Sources*, edited by W. A. Yost and A. N. Popper (Springer-Verlag, New York), pp. 143–190.

- Klein, D. A., and Frintrap, S. (2011). "Center-surround divergence of feature statistics for salient object detection," in *Proceedings of the International Conference on Computer Vision*, Barcelona, Spain.
- Kubovy, M., and Valkenburg, D. V. (2001). "Auditory and visual objects," *Cognition* **80**, 97–126.
- Kullback, S., and Leibler, R. A. (1951). "On information and sufficiency," *Ann. Math. Stat.* **22**, 79–86.
- Kong, L., and Shinn-Cunningham, B. (2009). "How visual cues help us understand speech in a complex environment," *J. Acoust. Soc. Am.* **125**, 2691.
- Lutfi, R. A. (1989). "Informational processing of complex sound: I. Intensity discrimination," *J. Acoust. Soc. Am.* **86**, 934–944.
- Lutfi, R. A. (1990). "Informational processing of complex sound. II. Cross-dimensional analysis," *J. Acoust. Soc. Am.* **87**, 2141–2148.
- 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., Chang, A.-C., Stamas, J., and Gilbertson, L. (2012). "A detection-theoretic framework for modeling informational masking," *J. Acoust. Soc. Am.* **132**, EL109.
- Lutfi, R. A., and Doherty, K. A. (1994). "Effect of component-relative entropy on the discrimination of simultaneous tone complexes," *J. Acoust. Soc. Am.* **96**, 3443–3450.
- Lutfi, R. A., Doherty, K. A., and Oh, E. (1996). "Psychometric functions for the discrimination of spectral variance," *J. Acoust. Soc. Am.* **100**, 2258–2265.
- Lutfi, R. A., and Liu, C. J. (2007). "Individual differences in source identification from synthesized impact sounds," *J. Acoust. Soc. Am.* **122**, 1017–1028.
- Lutfi, R. A., and Liu, C. J. (2011). "Target enhancement and noise cancellation in the identification of a rudimentary sound source in noise," *J. Acoust. Soc. Am.* **129**, EL52–EL56.
- Lutfi, R. A., and Liu, C. J., Stoelinga, C. N. J. (2008). "Level dominance in sound source identification," *J. Acoust. Soc. Am.* **124**, 3784–3792.
- Mahadevan, V., and Vasconcelos, N. (2010). "Spatiotemporal saliency in dynamic scenes," *IEEE Trans. Pattern Anal. Machine Intell.* **32**, 171–177.
- Martin, R. L., Bolia, R. S., Eberle, G., and Brungart, D. S. (2012). "Spatial release from speech-on-speech masking in the median sagittal plane," *J. Acoust. Soc. Am.* **131**, 378–385.
- Michéyl, C., Kreft, H., Shamma, S., and Oxenham, A. J. (2010). "Temporal coherence versus harmonicity in auditory stream formation," *J. Acoust. Soc. Am.* **133**, EL188–EL194.
- Michéyl, C., and Oxenham, A. J. (2010). "Pitch, harmonicity, and concurrent sound segregation: Psychoacoustical and neurophysiological findings," *Hear. Res.* **266**, 36–51.
- Neff, D. L., and Green, D. M. (1987). "Masking produced by spectral uncertainty with multicomponent maskers," *Percept. Psychophys.* **41**, 409–415.
- Oh, E. L., and Lutfi, R. A. (1998). "Nonmonotonicity of informational masking," *J. Acoust. Soc. Am.* **104**, 3489–3499.
- Oh, E. L., and Lutfi, R. A. (1999). "Informational masking by everyday sounds," *J. Acoust. Soc. Am.* **106**, 3521–3528.
- Oh, E. L., and Lutfi, R. A. (2000). "Effect of harmonicity on informational masking," *J. Acoust. Soc. Am.* **108**, 706–709.
- Peterson, G., and Barney, H. (1952). "Control methods used in a study of vowels," *J. Acoust. Soc. Am.* **24**, 175–184.
- Peterson, G., and Lehiste, I. (1962). "Revised CNC list for auditory tests," *J. Speech Hear. Disorders* **27**, 62–70.
- Pollack, I. (1975). "Auditory informational masking," *J. Acoust. Soc. Am.* **57**(Suppl. 1), S5.
- Shannon, C. E. (1951). "Prediction and entropy of printed English," *Bell Syst. Tech. J.* **30**, 50–64.
- Simpson, A. J., and Fitter, M. J. (1973). "What is the best index of detectability?" *Psychol. Bull.* **80**, 481–488.
- Sumby, W. H., and Pollack, I. (1954). "Visual contribution to speech intelligibility in noise," *J. Acoust. Soc. Am.* **26**, 212–215.
- Turvey, M. T. (1973). "On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli," *Psychol. Rev.* **80**, 1–52.
- Ulanovsky, N., Las, L., and Nelken, I. (2003). "Processing of low-probability sounds by cortical neurons," *Nat. Neurosci.* **6**, 391–398.
- Viemeister, N. F., Stellmack, M. A., and Byrne, A. J. (2011). "Discrimination of stimulus variance," *J. Acoust. Soc. Am.* **129**, 2588.
- Watson, C. S. (2005). "Some comments on informational masking," *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.
- Watson, C. S., Wroton, H. W., Kelly, W. J., and Benbassat, C. A. (1975). "Factors in the discrimination of tonal patterns I: Component frequency, temporal position, and silent intervals," *J. Acoust. Soc. Am.* **60**, 1175–1185.