Note on informational masking (L)

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

(Received 9 August 2002; revised 5 February 2003; accepted 24 February 2003)

Informational masking (IM) has a long history and is currently receiving considerable attention. Nevertheless, there is no clear and generally accepted picture of how IM should be defined, and once defined, explained. In this letter, consideration is given to the problems of defining IM and specifying research that is needed to better understand and model IM. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1570435]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Lj [MRL]

I. INTRODUCTION

This letter is intended to stimulate and broaden discussions about informational masking (e.g., Watson, 1987; Leek *et al.*, 1991; Neff *et al.*, 1993; Neff, 1995; Neff and Dethlefs, 1995; Oh and Lutfi, 1998, 2000; Lutfi *et al.*, 2003; Kidd *et al.*, 1998, 2002; Wright and Saberi, 1999; Richards *et al.*, 2002; Brungart *et al.*, 2001; Freyman *et al.*, 2001; Arbogast *et al.*, 2002). Although this term has been used in many ways, it is common to equate informational masking (IM) with nonenergetic masking, where energetic masking (EM) is defined as masking that results from competition between target and masker at the periphery of the auditory system, i.e., overlapping excitation patterns in the cochlea or auditory nerve (AN). Thus, EM is often equated with peripheral and IM with central masking. Also, because a primary function of peripheral processing is frequency analysis, most research in this area has focused on the frequency dimension. In a related definitional thread, IM has sometimes been defined as the elevation in threshold caused by stimulus uncertainty. Independent of the precise definition, and despite the association of IM with central attentional factors, IM is clearly distinguishable from general inattention to the overall experimental task (e.g., by differences in stimulus-response correlations).

Within the domain of nonenergetic masking, this note is confined to the detection of tonal targets in the presence of simultaneous multitone maskers. Nevertheless, it is hoped that it will prove useful when considering sequential masking, discrimination and recognition performance, or situations involving speech stimuli. [Some results of pioneering research on the effects of uncertainty that focuses on sequential masking for nonspeech stimuli are summarized in Watson (1987) and Espinoza-Varas and Watson (1989).]

IM is well illustrated by the case in which the target is a fixed-frequency tone and the masker is a ten-tone complex with the component frequencies selected randomly on each presentation, subject to the constraint that they all lie outside a protected region around the target tone (to minimize EM). In such a case, and for many listeners, the target threshold can exceed the average threshold obtained with the fixed exemplars of the random masker, or with a Gaussian noise masker having the same power, by as much as 40 dB. Apparently, under such conditions, these ‘‘holistic’’ listeners are severely distracted by the masker and find it difficult to perform well even though there is little masker energy in the frequency region of the target. Furthermore, the performance of these listeners cannot be easily improved by instructions, target cueing, or modest practice (Neff *et al.*, 1993). In contrast, the performance of these listeners can be greatly improved by altering the stimuli so that target-masker similarity is decreased, thereby reducing the tendency to confuse or group the target and masker (e.g., Kidd *et al.*, 1994; Neff, 1995; Oh and Lutfi, 2000; Durlach *et al.*, 2003). Also of importance, at the other end of the continuum, there exist ‘‘analytic’’ listeners who are highly resistant to spectral uncertainty (e.g., Neff and Dethlefs, 1995).

II. DEFINITIONAL ISSUES

Many factors drive the need for improved definitions, including (a) the lack of clarity in the notion of ‘‘overlap’’ or ‘‘competition’’ among peripheral channels; (b) the possibility of elevated masked thresholds being caused by uncertainty in dimensions other than frequency and/or by conditions other than stimulus uncertainty; and (c) the relativistic aspects of the distinction between peripheral and central (the meaning depending on one’s physiological vintage point).

To help address such issues, for any location L in the ascending auditory pathway, we define PM(L)=peripheral masking at L=masked threshold for the ideal detector operating on the input at L; and CM(L)=central masking at L=masked threshold of human observer minus the quantity PM(L). In general, determination of PM(L) and CM(L) for a given experimental condition requires (a) constructing a statistical model of neural activity at L for that condition; (b) computing the performance of the ideal detector operating on this model activity, and (c) comparing the performance obtained with this ideal detector to that of human listeners under the same condition. EM(L) can then be identified with PM(L) and IM(L) can be regarded as a component of CM(L) that includes uncertainty effects.

Note that the word ‘‘masking’’ is used here to refer to a variety of different processes associated with threshold elevation [an issue discussed by Tanner (1958)]. Note also that the distinction between PM(L) and CM(L) is consistent with the approach introduced by Siebert and Colburn (e.g., Siebert, 1968; Colburn, 1973) for the special case L=auditory nerve (AN). For a variety of tasks, they determined how

[a]Electronic mail: durlach@mit.edu

 0001-4966/2003/113(6)/2984/4/$19.00

much information was lost in the transformation from acoustical stimulus to AN firing patterns and how much in the transformation from these firing patterns to the human psychophysical responses. With this approach, ''energetic masking'' (at L) is identical to ''peripheral masking'' (at L) and the notion of ''overlapping excitation patterns'' (at L) is replaced by the characterization ''masking that cannot be overcome even by the ideal detector'' (at L). Note finally that the extent to which CM(L)=IM(L) is left open.

Although this definitional structure is responsive to the problems cited above, it is not problem-free. For example, it does not specify which central limitations that lead to threshold elevation should be included under CM (an issue related to Tanner's problem). More relevant to our current concerns, it fails to specify which components of CM should be included in IM. In particular, should the effects of similarity as well as of uncertainty be included? In addition, the definitions are difficult to apply (even though they are in principle operational definitions). Application not only requires considering physiological as well as psychophysical data, but it becomes increasingly difficult as one proceeds up the auditory system because of increased system complexity. Furthermore, as evident even in the simple standard definition of masking (elevation in threshold due to the presence of the masker), there is considerable operational imprecision. For example, how much and what kinds of training should precede the threshold measurements? This issue becomes particularly crucial in the area of interest because of possible differential learning effects associated with PM and CM. Despite such problems, it is hoped that the above thoughts can prove useful in organizing various masking phenomena. In the remainder of this note, unless explicitly stated otherwise, it is assumed that L=AN.

III. NEEDED RESEARCH

A. Uncertainty and similarity

Two main factors controlling the magnitude of nonenergetic masking (i.e., CM) that have been considered (apart from listener identity) are stimulus uncertainty and target-masker similarity. However, further research is required to develop an adequate understanding of how these factors generate masking and how they interact.

Strictly speaking, stimulus uncertainty is neither necessary nor sufficient to produce nonenergetic masking. On the other hand, if the randomization occurs over a range that is small compared to the listener's resolution, the randomization will not elevate the listener's threshold. On the other hand, if a sequence of stimuli is sufficiently complex and rapidly varying to appear random to the listener even though it is technically deterministic, then the listener's threshold can be greatly elevated. Clearly, as far as uncertainty is concerned, what matters is the deviation between what the listener hears on a given trial and what the listener expects to hear on that trial. Further evidence that stimulus uncertainty does not necessarily produce large amounts of masking is evident in the results (a) for ''analytic'' listeners who are resistant to the effects of uncertainty (Neff and Dethlefs, 1995; Oxenham *et al.*, 2003) and (b) from experiments in which the effects of uncertainty are reduced by decreasing target-masker similarity (Kidd *et al.*, 1994; Neff, 1995; Oh and Lutfi, 2000; Durlach *et al.*, 2003). Additional indications that nonenergetic masking can occur without uncertainty are available in data on cross-frequency effects in binaural hearing (Bernstein and Trahiotis, 1993; Culling *et al.*, 2003). Illustrative data on the effects of different types or degrees of uncertainty can be found in Watson *et al.* (1976) and Lutfi (1992).

In the same vein, although target-masker similarity is an important factor in causing nonenergetic masking, there are unresolved problems here, too. One such problem resides in the notion that it is the target-masker similarity itself that counts. Because the task in these detection experiments is not to discriminate between the masker M and the target T but rather between M and M+T, what is really important here is not the similarity between M and T but between M and M+T. Although these two similarity factors are related, they are not interchangeable; there are many situations in which the addition of T to M leads to a change in the sound of the overall stimulus that is not well described by the notion that one ''hears out'' T. A second such problem concerns how best to quantify the similarity factor. Although the similarity–dissimilarity dimension is closely related to the grouping–segregating distinction considered in auditory scene analysis (Bregman, 1990), it is not obvious how best to quantify similarity.

Finally, once the definitions of uncertainty and similarity factors are clarified, there will still be the need to determine how best to combine them in an overall model of nonenergetic masking.

B. Nonfrequency domains, sites central to the auditory nerve

Consistent with the above definitions, it is appropriate to consider not only the case in which the peripheral channels are the frequency channels at the level of the AN, but also cases in which the location L of interest is higher up in the system and the relevant ''peripheral channels'' concern domains other than frequency. For example, it might be enlightening to consider the case in which the frequency domain is replaced by the spatial domain and to explore the extent to which masked thresholds are elevated when uncertainty is introduced into the spatial characteristics of the stimulus rather than the frequency characteristics. A possible experiment in this area would determine the increase in the masked threshold of a target noise source located at a fixed azimuth θ_T caused by randomizing the azimuth θ_M of an independent masking noise, where the random draw of θ_M is constrained by a protected angular region about θ_T to minimize the spatial-domain EM that would occur when target and masker overlap in some relatively peripheral azimuthal channel. Although some results suggest that uncertainty has only small effects in the spatial domain (Bernstein and Trahiotis, 1997), further research is needed to adequately explore this area. Obviously, one could consider similar experiments in other domains as well (e.g., amplitude modulation). In order to truly understand nonenergetic masking, it is important to determine the extent to which the various phenomena observed in the frequency domain occur in other domains. [For relevant results of this type in sequential masking, see Watson and Kelly (1981)]. In the same spirit, it would be useful to compare results obtained in audition to those in vision (e.g., Turvey, 1973; Nakayama and Joseph, 1998; Cusack and Carlyon, 2000).

C. Psychometric functions, receiver operating characteristics, sequential effects

Most data on nonenergetic masking consists of thresholds measured using adaptive procedures. Relatively few data are available on psychometric functions, receiver operating characteristics (ROCs), or trial-to-trial sequential effects. Furthermore, initial examination of these elements (either by means of probe experiments or crude intuitive modeling) suggests that these elements may differ substantially in the different types of masking considered. For example, preliminary data from our lab (see also Wright and Saberi, 1999) indicate that psychometric functions from experiments like that outlined at the beginning of this note show important differences (in slope as well as lateral position) for the following three cases: (a) tests with a randomized masker, (b) tests with fixed exemplars of a randomized masker, and (c) tests with a randomized masker in which the results are sorted after the test to construct a psychometric function for each exemplar. It is expected that equally important differences will appear with ROCs and sequential effects. The results of such studies can provide important constraints on theoretical models of masking.

D. Individual differences

In the domain of nonenergetic masking studied to date (the frequency domain), intersubject differences are enormous. Whereas some listeners have their thresholds elevated by as much as 40 dB when masker uncertainty is introduced into a situation where target-masker similarity is high, other listeners appear insensitive to such uncertainty. Among the questions that arise here are (a) What are the sources of these intersubject differences? (b) How constant are they across conditions and domains? and (c) To what extent can they be reduced by training?

In the attempt to understand individual differences (e.g., Lutfi *et al.*, 2003), more data are needed in which the same subjects are tested in a wide variety of conditions concerned with the frequency domain as well as with other domains. In most previous studies, the subjects have varied between experiments. Also, in addition to continuing the search for new ways to characterize performance differences among subjects, more attention should be given to comparing methods already developed (involving critical bands, attentional bands, weighting constants, efficiency factors, etc.).

One special thrust to pursue concerns the Listener-Max-vs Listener-Min distinction (Durlach *et al.*, 2002; de Cheveigne and McAdams, 1995): Whereas Max is envisioned as an archetypal analytic listener who attempts to maximize the T/M ratio by maximizing T, Min is envisioned as an archetypal holistic listener who attempts to maximize the T/M ratio by minimizing M. In the frequency domain, Max constructs an acceptance filter focused on T without regard for M, whereas Min constructs a multiple notch-rejection filter matched to M without regard for T. In the spatial domain, Max points an acoustical searchlight at T, whereas Min points a set of nulls at the locations of the masker components. An interesting prediction that arises in connection with this distinction is the reversal in who does best when uncertainty shifts from M to T: when M is uncertain, Max should be best; when T is uncertain, Min should do best. Although previous data indicate that uncertainty in T is less disruptive than in M, no data are available to test this subject-reversal prediction. Research is also needed to explore the extent to which nonadditivity of masking (even when uncertainty is minimized) can be explained by special costs associated with the need for Min to create simultaneous multiple nulls.

A second thrust concerns the idea that, despite the focus on central processing in discussions of uncertainty effects, the observed intersubject differences may result from differences in peripheral processing (Carney, 2002; Lauter, 2002). For example, consider a random masker and suppose that two listeners, L1 and L2, have identical central processors but different peripheral representations of the ensemble of maskers. L1 and L2 might then reveal differences in susceptibility to uncertainty because (a) the same central processing is used to combat the uncertainty evident in two different peripheral representations of the masker ensemble or (b) two different central processing schemes are used, each of which is selected by the same central processor to optimize performance with the given peripheral representation. In either event, differences in peripheral processing would play a major role in causing the observed differences in susceptibility to uncertainty. Note also that, abstractly, the question of how susceptibility differs for L1 and L2 is essentially the same as the question of how susceptibility changes for either L1 or L2 when the ensemble of masking stimuli is changed. Note further that in order to pursue these issues, it would be useful not only to conduct relevant theoretical analyses (e.g., on the effects of different types of peripheral nonlinearities), but also to expand previous studies of the effects of sensorineural impairments (Doherty and Lutfi, 1999; Micheyl *et al.*, 2000; Kidd *et al.*, 2001). Of particular interest would be a comparison of monaural uncertainty effects between the two ears of subjects with unilateral impairments [similar to the work on grouping by Rose and Moore (1997)].

Finally, it is essential that further research be conducted on how much the very strong effects of uncertainty (or target-masker similarity) observed for some listeners can be reduced by training (and the extent to which generalization to other nonenergetic masking tasks occurs). Although such improvement in performance would appear less likely if the cause of the poor performance is peripheral rather than central, peripheral limitations would not be ruled out by such a finding. In any case, previous results on the difficulty of improving performance by target cueing, coaching, and limited training (e.g., Neff *et al.*, 1993) indicate that the training challenge is substantial.

In general, the fact that when target-masker similarity is high the effects of uncertainty are monstrous for some subjects and negligible for others implies that an understanding of individual differences in the effects of uncertainty, both pre- and posttraining, is essentially equivalent to understanding the effects themselves. Without such understanding, creating an insightful theory of informational masking will not be possible.

In conclusion, it should be noted that a comprehensive theory of informational masking will need to address not only simultaneous masking (the focus of this letter), but also sequential masking. [Extensive research in this area that includes consideration of individual differences and training has been performed by Watson and colleagues (e.g., Watson *et al.*, 1976; Leek and Watson, 1984; Espinoza-Varas and Watson, 1986; Leek *et al.*, 1991; Surprenant and Watson, 2001; Watson and Kidd, 2002).]

ACKNOWLEDGMENTS

We are indebted to Dr. L. Bernstein, Dr. L. Carney, Dr. J. Culling, Dr. J. Lauter, Dr. M. Leek, Dr. R. Lutfi, Dr. V. Richards, and Dr. C. Watson for useful communications about the material in this note. Our work was supported by Grant Nos. DC00100 and DC04545 from NIH/NIDCD and F49620-01-1-0005 from AFOSR.

Arbogast, T. L., Mason, C. R., and Kidd, Jr., G. (**2002**). ''The effect of spatial separation on informational and energetic masking of speech,'' J. Acoust. Soc. Am. **112**, 2086–2098.

Bernstein, L. R., and Trahiotis, C. (**1993**). ''Spectral interference in a binaural detection task: effects of masker bandwidth and temporal fringe,'' J. Acoust. Soc. Am. **94**, 735–747.

Bernstein, L. R., and Trahiotis, C. (**1997**). ''The effects of randomizing values of interaural disparities on binaural detection and discrimination of interaural correlation,'' J. Acoust. Soc. Am. **102**, 1113–1120.

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

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

Carney, L. H. (**2002**). Personal communication.

Colburn, H. S. (**1973**). ''Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination,'' J. Acoust. Soc. Am. **54**, 1458–1470.

Culling, J. F., Hodder, K. I., and Colburn, H. S. (**2003**). ''Interaural correlation discrimination with spectrally remote flanking noise: constraints for models of binaural interaction,'' J. Acoust. Soc. Am. (in review).

Cusack, R., and Carlyon, R. P. (**2000**). ''Auditory pop-out: Perceptual asymmetries in sequences of sounds,'' Br. J. Audiol. **34**, 112.

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.

Doherty, K. A., and Lutfi, R. A. (**1999**). ''Level discrimination of single tones in a multitone complex by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **105**, 1831–1840.

Durlach, N. I., Kidd, Jr., G., Mason, C. R., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. (**2002**). ''Informational masking: Toward improved understanding,'' J. Acoust. Soc. Am. **111**, 2337.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, Jr., G. (**2003**). ''Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity'' (in press).

Espinoza-Varas, B., and Watson, C. S. (**1986**). ''Temporal discrimination for single components of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **80**, 1685–1694.

Espinoza-Varas, B., and Watson, C. S. (**1989**). ''Perception of complex auditory patterns by humans,'' *The Comparative Psychology of Audition: Perceiving Complex Sounds*, edited by R. J. Dooling and S. H. Hulse (Erlbaum Associates, Hillsdale, NJ).

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

Kidd, Jr., G., 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, Jr., G., Arbogast, T. L., Mason, C. R., and Walsh, M. (**2001**). ''Informational masking in listeners with sensorineural hearing loss,'' J. Assoc. Res. Oto. **3**, 107–119.

Kidd, Jr., G., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). ''Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **104**, 422–431.

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

Lauter, J. (**2002**). Personal communication.

Leek, M., Brown, M. E., and Dorman, M. F. (**1991**). ''Informational masking and auditory attention,'' Percept. Psychophys. **50**, 205–214.

Leek, M., and Watson, C. S. (**1984**). ''Learning to detect auditory pattern components,'' J. Acoust. Soc. Am. **76**, 1037–1044.

Lutfi, R. A. (**1992**). ''Informational processing of complex sounds. III. Interference,'' J. Acoust. Soc. Am. **91**, 3391–3400.

Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (**2003**). ''One factor underlies individual differences in auditory informational masking within and across age groups,'' Percept. Psychophys. **65**, 396-406.

Micheyl, C., Arthaud, P., Reinhart, C., and Collet, L. (**2000**). ''Informational masking in normal-hearing and hearing-impaired listeners,'' Acta Otol. **120**, 242–246.

Nakayama, K., and Joseph, J. S. (**1998**). ''Attention, pattern recognition, and pop-out in visual search,'' in *The Attentive Brain*, edited by R. Parasuraman (MIT, Cambridge, MA).

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 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**, 3489–3499.

Oh, E. L., and Lutfi, R. A. (**2000**). ''Effect of masker harmonicity on informational masking,'' J. Acoust. Soc. Am. **108**, 706–709.

Oxenham, A. J., Fligor, B., Mason, C. R., and Kidd, Jr., G. (**2003**). ''Informational masking and musical ability'' (in review).

Richards, V. M., Tang, Z., and Kidd, Jr., G. (**2002**). ''Informational masking with small set sizes,'' J. Acoust. Soc. Am. **111**, 1359–1366.

Rose, M. M., and Moore, B. C. J. (**1997**). ''Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **102**, 1768–1778.

Siebert, W. M. (**1968**). ''Stimulus transformations in the peripheral auditory system,'' in *Recognizing Patterns*, edited by P. Kolers and M. Eden (MIT, Cambridge, MA).

Surprenant, A. M., and Watson, C. W. (**2001**). ''Individual differences in the processing of speech and nonspeech sounds in normal-hearing listeners,'' J. Acoust. Soc. Am. **110**, 2085–2095.

Tanner, W. P. (**1958**). ''What is masking?'' J. Acoust. Soc. Am. **30**, 919–921.

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.

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), pp. 267–277.

Watson, C. S., and Kelly, W. J. (**1981**). ''The role of stimulus uncertainty in the discriminability of auditory patterns,'' in *Auditory and Visual Pattern Recognition*, edited by D. J. Getty and J. H. Howard, Jr. (Erlbaum, Hillsdale, NJ), pp. 37–59.

Watson, C. S., and Kidd, G. (**2002**). ''On the lack of association between basic auditory abilities, speech processing, and other cognitive skills,'' Semin. Hear. **23**, 83–93.

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.

Wright, B. A., and Saberi, K. (**1999**). ''Strategies used to detect auditory signals in small sets of random maskers,'' J. Acoust. Soc. Am. **105**, 1765–1775.