Impact of uncertainty on sound perception

Xiang Cao

Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA

Address for correspondence:
Dr. Xiang Cao, Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA. E-mail: xiangcao@seas.upenn.edu

DOI: 10.4103/0976-9668.71664

Abstract
In the auditory study, masking caused by sound uncertainty is a hot topic because research on this topic has a potential to be applied to improve human perception in a real world. This article introduces the origination of the study on this so-called informational masking and lists some key results obtained. The informational masking is widely accepted to result from the central auditory system since the classical auditory peripheral model fails to account for the data. This article reviewed the currently most satisfactory model on informational masking and its ability and disability in accounting for current experimental data. In the end, potential sources of the informational masking are discussed as an indication for the future research direction. The review is mostly based on articles published in JASA and JARO, the two most prestigious journals in the auditory study.

Key words: CoRE model, informational masking, masking

INTRODUCTION

Perceiving the sounds from a particular sound source becomes much more difficult when sounds from other independent sources are presented at the same time. These “masking” sounds compete with the “target” sound. During the 70s of the last century, Green and Swets[1] proposed a critical-band energy-detector model to predict the competition between the masking and target sounds within the auditory system. The model consists of a “critical band” filter,[2] followed by a physiologically inspired rectifier, an integrator (based on psychophysical estimates of temporal integration), and a mechanism that based decisions on stimulus energy. The model has successfully predicted the amount of masking produced by noise with different width of bands.

However, this model fails to predict the amount of masking when maskers comprise N randomly selected tonal components where N was varied over a wide range.[1] The task was to detect a pure-tone target of a fixed and known frequency (0.5, 1, and 4 kHz) located within a distribution of potential masker components, for masker samples drawn at random on every presentation. These multitone maskers were essentially sparse samples of components from Gaussian noise with the parameter of interest. Large amounts of masking are found for small numbers of components that, according to the classical energy-detector theory, should produce very little masking. Based on intuitions drawn from the critical-band energy-detector model, because it would rarely happen that the masker components would fall near enough the target to mask it, little amount of masking is expected. Surprisingly, large amounts of masking (more than 50 dB for 1 and 4 kHz) were found for maskers comprising as few as 10 components and significant amounts of masking were observed for only 2 components.

This extra masking caused by the masker frequency uncertainty is usually referred to as “informational masking,” a term first introduced in an abstract by Irwin Pollack in 1975.[4] In subsequent years, the simultaneous multitone masking procedure was used by many investigators to study the masking in addition to that predicted by the critical-band energy-detector model.[5-12] It was found that the maximum amount of masking did not occur for the masker consisting of the most components (i.e., true Gaussian noise) but instead for a masker comprising about 10–20 frequency components.[7] When the masker sample was randomized only between trials, but was the same on both intervals within a trial, the amount of masking was substantially reduced. When the sound pressure level of the maskers (total energy) was held constant, the amount of masking initially increased,
reached a plateau, and then decreased. The interpretation is that when there are very few components, the masking that is observed is dominated by informational masking while, at the opposite extreme where the masker is true Gaussian noise, the masking observed is almost entirely energetic, dominated by the energy-detector model. Varying the number of masker components thus changes the ratio of energetic to informational masking. The plateau region from roughly 10 to 100 components indicates component densities where the successive masker components are sufficient to create a significant uncertainty while also producing significant amounts of energetic masking.

The only model to date that has been used to explain the results from a wide range of studies by taking into account both energetic and informational masking is the component-relative entropy (CoRE) model, first proposed by Lutti. In the original article describing the CoRE model, Lutti demonstrated that it could predict the results from a variety of multitone masking experiments, such as those reported by Neff and Green. However, it also accurately predicted the findings from other types of studies such as the profile analysis experiment described by Kidd et al., in which sensitivity to differences in a spectral shape was measured for randomly perturbed reference spectra. The model relies on the statistical summation over trials of the outputs of a set of auditory filters spanning the audible frequency range. The amount of masking that is predicted is related to both the target-to-masker ratio in the band containing the target and to the variability of the outputs of the attended to non-target bands. For instance, when detecting a pure-tone target in Gaussian noise, the variability in the outputs of the nontarget bands, computed across trials, would be relatively low (depending on bandwidth, duration, etc.). The threshold for the target would primarily be determined, therefore, by the target-to-masker energy in the target’s band with very little contribution to the overall amount of masking from nontarget bands. In contrast, for a random-frequency multitone masker comprising a few components and with a “protected region” surrounding the target in which no masker component frequency falls in, the variability of the outputs of the nontarget bands across trials may be quite large while the target-to-masker ratio at the masked threshold in the target’s band may be very high. This situation consists of a small amount of energetic masking with a relatively large amount of informational masking. The predictions of the CoRE model are relatively accurate to predict these results.

Although the CoRE model provides an excellent account of the data, as well as a number of other informational masking conditions, there are some findings that it cannot explain. For example, Oh and Lutti report that the CoRE model does not provide a satisfactory explanation for the large decrease in masking found by causing a target tone to be mistuned slightly in a multitone masker having masker components drawn at random from a set of harmonically related tones. Also, Kidd et al. noted that the CoRE model does not capture the trend in masking apparent for the multiple-bursts different masker (discussed below) as the number of masker bursts and interburst intervals are varied. Furthermore, the CoRE model does not take the target–masker similarity into account. Nonetheless, the CoRE model represents an important conceptual tool that can help explain many of the findings from a large subset of informational masking studies.

One aspect of informational masking that is clear for most studies and a diverse set of experimental procedures is that the listener attends to frequency regions that provide no useful information for solving the task. Neff et al. sought to determine whether listeners who were very susceptible to informational masking (“high-threshold” listeners) exhibited wider listening bandwidths than less susceptible (“low-threshold”) listeners. In order to obtain listening bandwidth estimates, they adapted the techniques normally used to measure “auditory filter” characteristics and applied them to the multitone masking experiment. In the more common procedure, a set of threshold estimates is obtained for pure-tone targets masked by notched-filtered noise as the bandwidth of the notch is varied. Based on these threshold estimates, a best-fitting set of filter parameters (making an assumption about the type of filter) are then computed. Neff et al. performed a similar analysis using data obtained from the multitone masking experiment where the variation in the width of the “notch” was accomplished by changing the size of the “protected region” around the target frequency. They found that the estimated “attentional” filter bandwidths and processing efficiency (related to the target-to-masker ratio in the filter at the masked threshold) were both lower and poorer, respectively, in the high-threshold group than in the low-threshold group with large differences found between subjects. Furthermore, the more susceptible high-threshold group generally exhibited large amounts of informational masking even for extremely broad protected regions.

Informational masking often results from stimulus uncertainty. Attention, grouping and segregation, memory, and general processing capacity, are all factors that are related to producing informational masking or in causing release from informational masking. For example,
Durlach et al.[18] speculated that the representation of a target at one physiological level might be sufficient, i.e., not energetically masked, yet at a higher level that would not be true. Another possible cause of informational masking would be whether the outputs of the neural elements representing the target at a given physiological site are combined (or “grouped”) with other elements representing irrelevant stimuli. This could happen for a variety of reasons, such as the target and masker being “mixed” along a particular stimulus dimension and presented synchronously. One possible illustration of informational masking due to a failure of segregation was reported by Kidd et al.[19] Informational masking, consistent with the definition proposed by Durlach et al.,[18] could also be caused by an incorrect selection of the available neural elements – either to enhance or to suppress – at a given physiological site. And finally (although there are many other doubtless possibilities not considered here), limitations on the short-term storage and retrieval of sounds in memory, or interruptions in the processing of stored sounds, can produce informational masking. In general, the way by which stimulus uncertainty produces informational masking keeps open to us.

SUMMARY

It is much harder to perceive a target sound when the competing sounds are random than when the competing sounds are known in advance. This is categorized as informational masking, which occurs not at the auditory periphery but at the central auditory system. The CoRE model has been proposed to account for a large part of the data. However, it fails to account for informational masking with harmonic stimuli or streaming. The origin of informational masking is unclear yet, but it is believed to relate to segregation or memory of auditory stimuli.

ACKNOWLEDGMENT

This work was supported by grant RO1 DC02012 from NIH/NIDCD.

REFERENCES

1. Green DM, Swets JA. Signal Detection Theory and Psychophysics.: Florida, Krieger Publishing Company; 1974.
2. Fletcher H. Auditory patterns. Rev Mod Phys 1940;12:47-65.
3. Neff DL, Green DM. Masking produced by spectral uncertainty with multicomponent maskers. Percept Psychophys 1987:41:409-15.
4. Pollack I. Auditory informational masking. J Acoust Soc Am 1975;57:55.
5. Neff DL, Dethlefs TM, Jesteadt W. Informational masking for multicomponent masks with spectral gaps. J Acoust Soc Am 1993;94:3112-26.
6. Kidd G Jr, Mason CR, Deliwala PS, Woods WS, Colburn HS. Reducing informational masking by sound segregation. J Acoust Soc Am 1994;95:3475-80.
7. Oh EL, Lutfi RA. Nonmonotonicity of informational masking. J Acoust Soc Am 1998;104:3489-99.
8. Wright BA, Saberi K. Strategies used to detect auditory signals in small sets of random maskers. J Acoust Soc Am 1999;105:1765-75.
9. Richards VM, Tang Z, Kidd G Jr. Informational masking with small set sizes. J Acoust Soc Am 2002;111:1359-66.
10. Richards VM, Huang R, Kidd G Jr. Masker-first advantage for cues in informational masking. J Acoust Soc Am 2004;116:2278-88.
11. Durlach NI, Mason CR, Shinn-Cunningham BG, Arbogast TL, Colburn HS, Kidd G Jr. Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. J Acoust Soc Am 2003;114:368-79.
12. Durlach NI, Mason CR, Gallun FJ, Shinn-Cunningham B, Colburn HS, Kidd G Jr. Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers. J Acoust Soc Am 2005;118:2482-97.
13. Lutfi RA. A model of auditory pattern-analysis based on component-relative-entropy. J Acoust Soc Am 1993;94:748-58.
14. Kidd G Jr, Mason CR, Green DM. Auditory profile analysis of irregular sound spectra. J Acoust Soc Am 1986;79:1045-53.
15. Oh EL, Lutfi RA. Effect of masker harmonicity on informational masking. J Acoust Soc Am 2000;108:706-9.
16. Kidd G Jr, Mason CR, Richards VM. Multiple bursts, multiple looks, and stream coherence in the release from informational masking. J Acoust Soc Am 2003;114:2835-45.
17. Patterson RD, Nimmo-Smith I, Weber DL, Milroy R. The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. J Acoust Soc Am 1982;72:1788-803.
18. Durlach NI, Mason CR, Kidd G, Arbogast TL, Colburn HS, Shinn-Cunningham BG. Note on informational masking (L). J Acoust Soc Am 2003;113:2984-7.
19. Kidd G Jr, Mason CR, Brughera A, Chiu CY. Discriminating harmonicity. J Acoust Soc Am 2003;114:967-77.

Source of Support: RO1 DC02012 from NIH/NIDCD. Conflict of Interest: None declared.