Atten Percept Psychophys (2011) 73:2639–2648
DOI 10.3758/s13414-011-0213-8

Using the ideal observer to predict performance in perceptual tasks: An example from the auditory temporal masking domain

Justin A. MacDonald

Published online: 18 October 2011

Abstract This article provides a demonstration of an analytical technique that can be used to investigate the causes of perceptual phenomena. The technique is based on the concept of the ideal observer, an optimal signal classifier that makes decisions that maximize the probability of a correct response. To demonstrate the technique, an analysis was conducted to investigate the role of the auditory periphery in the production of temporal masking effects. The ideal observer classified output from four models of the periphery. Since the ideal observer is the best of all possible observers, if it demonstrates masking effects, then all other observers must as well. If it does not demonstrate masking effects, then nothing about the periphery requires masking to occur, and therefore masking would occur somewhere else. The ideal observer exhibited several forward masking effects but did not exhibit backward masking, implying that the periphery has a causal role in forward but not backward masking. A general discussion of the strengths of the technique and supplementary equations are also included.

Keywords Audition · Psychoacoustics · Decision making

Imagine that you are a vision researcher trying to determine whether the information about the visual environment available at the retina is sufficient to explain human performance in tasks requiring depth perception (see Liu & Kersten, 1998, for the full details of this example). You could choose to measure the responses of the cells at the retina during the presentation of a stimulus, but it is not clear how measurements of the output of a physiological system will lead to firm conclusions about the causes of the behavioral phenomena related to depth perception. This article provides a demonstration of a tool that can used to bridge the gap: a tool to generate behavioral predictions from physiological data.

In this article, the transformation from physiological measure to behavioral result is accomplished by attaching a discrimination device to the output of the physiological structure under study. To continue the depth perception example, perhaps views of different three-dimensional objects are provided to the subject, and the resulting response at the optic nerve to each object is recorded. Using some decision rule, the discrimination device determines which 3-D object is likely to have produced each of the measured outputs. If the same task is given to the human observer, then the two sets of results are directly comparable. However, this situation is still not ideal: The results of the discrimination device depend on the properties of both the retina and the decision rule used. Any match or discrepancy between the two sets of results could be a consequence of either (or both) factors. In an effort to avoid this confound, and hopefully to generate useful conclusions from the physiological measures, this article focuses on the use of the *ideal observer* (IO; see Green & Swets, 1988) as the discrimination device. The IO forms the basis of signal detection theory and other statistical models of decision-making under uncertainty. The signal-known-exactly version of the IO (see Tanner, Birdsall, & Clarke, 1960) classifies an incoming stimulus by comparing it to perfect (noise-free) templates of all experimental stimuli. Performance is brought off of the ceiling by assuming some noise in the incoming stimulus. In a discrimination task, the IO considers all possible factors that would help to increase the probability of determining correctly the identity of the incoming stimulus, and then outputs the decision that is most likely to be correct.

J. A. MacDonald (✉)
New Mexico State University,
MSC 3452/P.O. Box 30001, Las Cruces, NM 88003, USA
e-mail: jmacd@nmsu.edu

Given a set of stimuli to discriminate between, the IO's strategy results in the best possible performance (Green & Swets, 1988).

Although the IO can sometimes be used effectively as a predictive model of higher-level decision-making processes (e.g., Dau, Puschel, & Kohlrausch, 1996; Gresham & Collins, 1998; Jepsen, Ewert, & Dau, 2008), it remains a simplistic and almost certainly incorrect representation of the human decision-maker. It has a major advantage over other decision-making models, however, in that its optimality allows for identification of an upper bound on discrimination performance. Accordingly, in this article the IO is not treated as a model of human decision-making, but instead as a tool to measure the information content of stimuli after they pass through the physiological structure under study. The methods detailed here can operate on any digitized measurements of perceptual phenomena: visual, auditory, tactile, or otherwise. To demonstrate this technique effectively, I chose to apply it within the domain of auditory perception, specifically forward and backward masking. Although masking phenomena are not the main focus of this paper, a brief description of them is warranted for the sake of clarity.

Auditory masking occurs when a listener's ability to perceive a signal (the "target") decreases due to the presence of another sound (the "masker"). Forward and backward (or more generally, temporal) masking occurs when the presentations of the signal and masker do not coincide in time. "Forward" masking occurs when the presentation of the masker ends before the signal onset, whereas "backward" masking occurs when the masker is presented after signal offset. Both forward and backward masking typically increase (i.e., the target becomes harder to detect) as the time between the target and masker (the interstimulus interval, or ISI) decreases (Fastl, 1976, 1977, 1979; Luescher & Zwislocki, 1949; Miller, 1947; Penner, 1974).

A long and lively debate raged for decades (and still rages, to some extent) about the causes of forward and backward auditory masking phenomena. I chose this domain to demonstrate the IO-based approach because the debate about the role of the auditory periphery in the production of temporal masking effects has largely been resolved. For the purposes of this article, the auditory periphery includes all auditory structures in the outer, middle, and inner ear, from the external ear through to the auditory nerve. While the periphery is thought to be at least partially responsible for forward masking (Duifhuis, 1973; Harris & Dallos, 1979; Nelson, Smith, & Young, 2009; Oxenham & Moore, 1995; Oxenham & Plack, 2000), backward masking is largely thought to have a more central cause (Recio & Rhode, 2000; Samoilova, 1959; Shul'gina & Murav'ev, 2000). Thus, I can use the IO to test the role of the periphery while knowing the correct result beforehand. I have chosen several well-known and replicated effects for this purpose, including the effects of ISI, masker duration, and masker intensity on target detection thresholds.

For the demonstration analyses, stimuli filtered through the auditory periphery will serve as templates for a discrimination task to be completed by the IO. If the IO demonstrates temporal masking effects at this point, it would follow that all other observers must also exhibit these effects, since the IO performs best out of all possible observers. In essence, if the periphery forces even the best decision-maker to exhibit masking effects, then a strong case can be made for the periphery being at least partially responsible for psychophysical masking. However, if the IO does not demonstrate masking effects, then nothing about the information content of the stimuli themselves necessitates masking effects. In this case, it is reasonable to conclude that the periphery is not responsible for temporal masking. In this way, the IO can be used as a measure of the filtering properties of the structure: If the IO doesn't exhibit masking effects when operating on the inputs to the structure, but does when operating on the outputs, the structure essentially forces all observers, including the ideal, to exhibit masking effects. If no masking effects are found by the IO at the output of the structure under study, the periphery can be ruled out as a potential cause. Instead, something more central must cause masking effects: Either another structure filters out information before it reaches the observer, thereby causing masking, or the human observer makes suboptimal use of the available information, or some combination of both.

Ideally, IO analyses would be conducted on actual measurements of the auditory periphery. For the purposes of this demonstration, however, four models served as substitutes to be tested using the IO: the auditory image model (Patterson, Allerhand, & Giguere, 1995), the dual resonance nonlinear (DRNL) filter model (Meddis, O'Mard, & Lopez-Poveda, 2001), the Carney (1993) model, and the Zilany model (Zilany & Bruce, 2006). These models were chosen to represent a range of approaches to approximating the function of the periphery, although the differences between them are not particularly important for the purposes of this study. All models were implemented in software that takes a discrete waveform as input and generates another discrete waveform as output that represents the probability of spike activity along the auditory nerve over time. These output waveforms served as the templates used by the IO. Using models has the benefit of allowing for unlimited stimulus presentations, which in turn allows for the measurement of extremely complete psychometric functions.

Assessing the relative performance of each periphery model using the IO is not entirely straightforward. In essence, overprediction can rule out models, while underprediction cannot. In this example analysis, if the periphery models underpredict masking effects, they are consistent with the observed behavioral data. Masking effects measured using behavioral methods likely result from a combination of peripheral and central effects, so any periphery model that predicts a portion of that total is consistent with the data. Only when a periphery model predicts more than that total can we start discriminating between models. Of course, if each periphery model/IO concatenation is treated as a model of the human system, one could compare quantitative fits across models to determine which combination best accounts for the human data. These conclusions apply to the periphery model + IO combinations, though, and not necessarily to the periphery models themselves.

Ideal-observer analyses

Analysis I: Forward masking functions

The first analysis was conducted to determine the ability of the periphery models to predict the relation between ISI and threshold in a forward masking paradigm. The general consensus in the literature is that the periphery is at least partially responsible for forward masking effects. Harris and Dallos (1979) conducted a comprehensive series of measurements on the output of auditory nerve (AN) fibers in response to forward masking stimuli. AN response to the target decreased with the ISI, suggesting a role of the periphery in forward masking. However, Nelson, Smith, and Young (2009) reported evidence suggesting that additional masking occurs at the level of the inferior colliculus (IC). Single-neuron recordings made in the central nucleus of the IC demonstrated stronger masking than was observed at the AN, suggesting a contribution to forward masking from subcortical structures.

The first analysis determined the performance of the IO in a two-interval forced choice (2IFC) detection task modeled after Experiment 1 from Jesteadt, Bacon, and Lehman (1982). Jesteadt et al. used pure-tone targets and maskers to measure forward masking functions at ISIs ranging from 5 to 40 ms. Consistent with the empirical literature, each of the periphery models should exhibit forward masking effects, although perhaps not to the extent of those measured in behavioral studies.

Ideal-observer analysis method

Stimuli All of the stimuli used in this article were Microsoft WAV files sampled at 44.1 kHz and 16 bits. Because WAV files are recorded in arbitrary units, it was necessary to rescale them into micropascals before they were presented to the models of the periphery. For the first simulation, the masker was a 300-ms, 1-kHz pure tone that was presented at 60 dB SPL. The target stimuli were 1-kHz pure tones and were 20 ms in duration. Both the target and masker (T+M) stimuli had 10-ms onset and offset ramps. T+M stimuli in the forward masking conditions had the following structure: A masker was followed by an ISI of 5, 10, 20, or 40 ms, followed by the target. Except for a 20-ms period of silence replacing the target, M stimuli had the same structure as the T+M stimuli.

Auditory periphery model settings The AIM, DRNL, and Carney models were implemented using the Development System for Auditory Modeling (DSAM) software package (Patterson, 2001). The Zilany model was implemented using a MATLAB executable available from the authors on the Internet (Zilany & Bruce, 2007). The implementations of each of the models were held constant throughout all of the analyses reported in this study, with the exception that the characteristic frequency of the inner hair cell was altered to match the frequency of the target stimulus. All of the models take a signal in micropascals (or pascals, in the case of the Zilany model) as input and produce as output a series of samples corresponding to the spike probability associated with a single AN fiber (the fiber associated with the 1-kHz frequency, in this analysis) per unit time. Simulating actual spike events rather than probabilities was considered as a possibility, but doing so would only add noise to the output and would not alter the results in any positive way. Outputting probabilities has the additional advantage of rendering the models completely deterministic, thereby eliminating the need for multiple simulation runs when pure-tone stimuli are used as input. The AIM, DRNL, Carney, and Zilany models were implemented using their default parameters, as listed in their respective published descriptions (Carney, 1993; Meddis et al., 2001; Patterson et al., 1995; Zilany & Bruce, 2006).

Analysis implementation The templates used by the IO consisted of the responses of the periphery models to the T+M and M stimuli. The IO was allowed prior knowledge of the ISI on the upcoming trial. Informal testing indicated that knowledge of the ISI significantly reduces the computational requirements without changing the results of the simulation. Knowledge of the ISI allowed the observer to focus only on the portions of the incoming stimulus that might contain a response due to the target. For example, if the ISI was 20 ms in a forward masking task, the IO ignored all neural activity until 20 ms after masker offset, as this portion of the interval would not include any information that could be used to increase the proportion of correct responses. The IO also knew the duration of the response of the auditory nerve to the presentation of a 20-ms signal, which in the case of the periphery models under consideration is not more than 45 ms. At maximum stimulus intensity, neural activity returns to baseline approximately 20 ms after stimulus offset in all model implementations. To decrease the number of calculations necessary to complete the analysis, therefore, both of the templates used in this analysis were made up of only those portions of the intervals that might contain the target (45 ms in each interval, so that each incoming stimulus and template was 90 ms in duration). The IO had prior knowledge of the target intensity, and therefore was given the pair of templates corresponding to the intensity of the target in the upcoming trial.

The performance of the IO for each ISI/target intensity combination was calculated analytically using Eq. A17 (see the Appendix). This equation allows for the calculation of the probability of a correct response in an arbitrary two-alternative forced choice (2AFC) discrimination task, and can easily be adapted to a 2IFC paradigm; a 2IFC task can be treated as a special case of the standard 2AFC discrimination task. On a given trial, the concatenated interval pair is treated as a single stimulus, one that is labeled "Interval 1" if the T+M stimulus is in the first interval or "Interval 2" if the T+M stimulus is in the second interval. To determine the detection threshold associated with a particular ISI, the performance of the IO was calculated for each target intensity, ranging from 0 to 120 dB SPL. The 70.7% threshold at each ISI was estimated from the resulting psychometric function using linear interpolation between the closest points. The standard deviation parameter (σ) varied across models. The value for each model was chosen so that the threshold in quiet exhibited by each of the models matched the 19-dB SPL threshold in quiet observed with the human subjects in Jesteadt et al. (1982). Values for the σ parameter were set at 0.0046, 0.0795, 0.0789, and 0.2083 for the AIM, DRNL, Carney, and Zilany models, respectively.

A custom MATLAB script was written to administer the simulations. All stimuli were generated in the script and then passed to the DSAM program (in the case of the AIM, DRNL, and Carney models) or the MATLAB implementation of the Zilany model (Zilany & Bruce, 2007). The script then retrieved the results from the periphery model and calculated thresholds for each ISI/periphery model combination.

Analysis results and discussion

The results of the forward masking analyses for each model of the periphery, as well as the results from Jesteadt et al. (1982), are shown in Fig. 1. All four models of the auditory periphery predicted the expected qualitative pattern of masking results: Masked thresholds decreased as ISI increased, presumably because the AN adaptation at signal onset decreased as the ISI was increased. The AIM, DRNL, and Carney models underpredicted masking effects at all ISIs, in some cases by as much as 20 dB. It is important to note, however, that underpredicted masking effects are consistent with the behavioral data, and overpredicted masking effects are not. In fact, there is some evidence that physiological measures of AN responses exhibit smaller masking effects than do those measured behaviorally (Harris & Dallos, 1979; Relkin & Turner, 1988). This suggests that forward masking has both peripheral and central causes, and the results of the AIM, DRNL, and Carney models are consistent with this. The slight overprediction of the Zilany model in the 5-ms ISI condition is more diagnostic, however. When making optimal use of the information available at the output of the Zilany model, the IO performed worse than the human listeners in the Jesteadt et al. study. Since no observer can perform better than optimal, it follows that human observers had more information available to them during the decision-making process. Therefore, given the serial nature of the auditory periphery, the Zilany model reduces the information content of the auditory stream to a greater extent than the human auditory periphery does.[1] In what manner this occurs is well beyond the scope of this article.

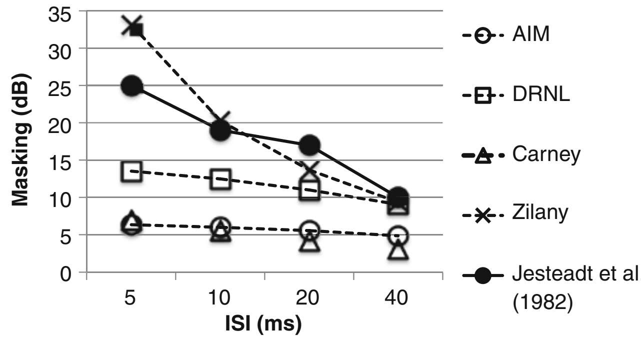

Fig. 1 Forward masking functions predicted by the periphery models. Although masking effects were observed for all four models, three of them underpredicted the amount of masking by as much as 20 dB. The Zilany model overpredicted masking effects by 8 dB in the 5-ms ISI condition

Analysis II: Backward masking functions

A follow-up analysis was conducted to investigate the role of the periphery in the production of backward masking effects. Although there is considerably less consensus in the literature about the role of the periphery in backward masking, the preponderance of evidence suggests that backward masking

[1] The Zilany model also consistently produced nonmonotonic psychometric functions across a range of forward masking stimuli, meaning that there was not always a positive relationship between target intensity and AN response. This typically occurred at signal levels between 80 and 100 dB SPL, and only when a masker preceded the target. Fortunately, this nonmonotonicity occurred away from the 70.7% threshold levels for all simulations reported in this article.

has a central cause. Pastore and MacLatchy (1975) demonstrated that backward masking effects could be reduced if a contralateral stimulus was presented within a few milliseconds of target onset, suggesting that attentional factors play a role in backward masking effects. In addition, backward masking effects can be reduced or entirely eliminated with practice (Miyazaki & Sasaki, 1984; Samoilova, 1959; Strait, Kraus, Parbery-Clark, & Ashley, 2010). Therefore, the periphery models are not expected to exhibit backward masking effects in this analysis. A simulated backward masking experiment patterned after Elliott (1962) provided a test of this prediction.

Ideal-observer analysis method

Stimuli As in Elliott (1962), the masker was a 50-ms Gaussian noise with 10-ms onset and offset ramps. The target stimuli were 1-kHz pure tones and were 10 ms in duration, with 1-ms onset and offset ramps. The masker was presented at 90 dB SPL, and the target stimuli ranged in intensity from −10 to 30 dB SPL. T+M stimuli had the following structure: The target was followed by an ISI of 0, 1, 3, 5, 10, 15, 25, or 50 ms, followed by a masker. Except for a 10-ms period of silence replacing the target, M stimuli had the same structure as the T+M stimuli.

Auditory periphery model settings All of the analysis parameters were identical to those used in Analysis I, with the exception of the free parameter σ. Here, σ was set for each model so that the threshold of the IO was at 0 dB in quiet. This value was chosen because thresholds in quiet were not reported in Elliott (1962), and 0 dB SPL is by definition the threshold for a normal listener at 1 kHz. Values for the σ parameter were set at 0.0003, 0.0050, 0.0060, and 0.0131 for the AIM, DRNL, Carney, and Zilany models, respectively.

Analysis implementation As in Analysis I, the templates used in this analysis were made up of only those portions of the intervals that might contain the target (45 ms in each interval, so that each incoming stimulus template was 90 ms in duration). The psychometric function associated with each ISI was calculated as in Analysis I.[2]

[2] It should be mentioned here that the inclusion of a noise stimulus in this and later simulations allowed for trial-to-trial variability in the sample of noise that served as the masker. Given that each noise sample is normalized to be exactly equal to the required intensity, that the models themselves are completely deterministic, and that the IO uses as templates only those portions of the stimulus that might contain the target (i.e., no masker), the amount of variability introduced into the measured threshold was trivial. Informal testing indicated that the thresholds measured were reliable to the second decimal place from simulation to simulation. Since a greater degree of precision was not important for the purposes of this study, multiple runs were not conducted.

Analysis results and discussion

The results of the backward masking simulations, as well of those from Elliott (1962), are depicted in Fig. 2. As expected, the models of the periphery did not predict backward masking effects. In fact, two of the models trended toward predicting slightly *reduced* thresholds at low ISIs. These findings are clearly at odds with the backward masking psychophysical literature, which typically reports masking effects with ISIs of up to 25–30 ms (Elliott, 1962; Penner, 1974). This suggests that either all four models of the periphery are missing the component that causes backward masking in the periphery, or (congruent with the literature) the periphery doesn't have any role in the masked thresholds obtained through psychophysical means.

The results of Analysis I indicated that each of the four models of the auditory periphery altered the forward masking stimuli so that even the best-performing observer exhibited elevated thresholds. It is clear from the results of the Analysis II, however, that the models did not give even the slightest indication of backward masking effects. These results are in accord with the literature on the topic (Harris & Dallos, 1979; Strait et al., 2010), indicating the validity of both the approach and the models used. Given the successful demonstration of forward masking effects using the periphery and IO, subsequent analyses focused on predicting the effects of varying masker duration and intensity in a forward masking paradigm.

Analysis III: Masker duration

The reports of masker duration effects in the forward masking literature have been relatively congruent. Zwicker (1984) used maskers 5 and 200 ms in duration and reported that the longer maskers were more effective across a range of ISIs. Kidd and Feth (1982) reported a similar increase in

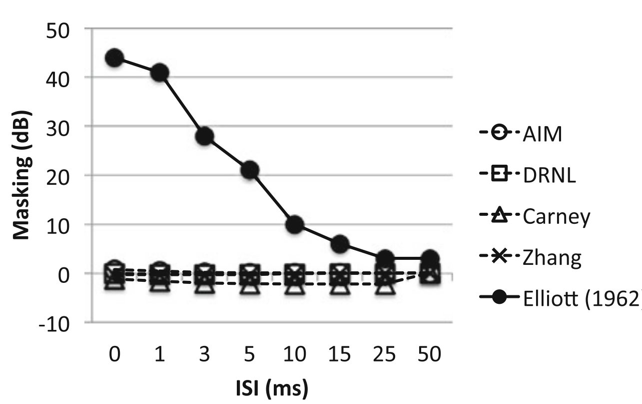

Fig. 2 Backward masking functions predicted by the periphery models. The models predicted either essentially unchanged or slightly reduced thresholds, contrary to behavioral measures. Note that the results from Elliott (1962) were obtained using an adaptive version of the method of limits, and are therefore not directly comparable

forward masking when the duration of the masker was increased to up to 500 ms. Penner (1974) found strong effects of masker duration using maskers that ranged from 1 to 625 ms. Oxenham and Plack (2000) reported an effect of masker duration using 40-dB spectrum level Gaussian noise maskers that ranged from 5 to 200 ms in duration. Accordingly, an IO analysis was conducted to determine the ability of the periphery models to predict this effect of masker duration. The analysis method was patterned after the forward masking experiment detailed by Zwicker, so that simulation results could be compared to human data. In Zwicker's experiment, detection thresholds for a pure tone preceded by a uniform noise masker were measured at several masker durations and ISIs. Zwicker reported a strong effect of masker duration on the target detection threshold. An IO analysis was expected to reproduce these results.

Ideal-observer analysis method

Stimuli As in Zwicker (1984), the signal was a 2-kHz pure tone presented for 5 ms with 1-ms onset and offset ramps. Target stimuli ranged in intensity from 0 to 120 dB SPL. The masker was uniform white noise presented at 80 dB SPL with a duration of either 5 or 200 ms. T+M stimuli comprised a masker followed by a 0-, 5-, 15-, 45-, or 95-ms period of silence, followed by a 5-ms target. M stimuli were generated in exactly the same way, with the exception that 5 ms of silence replaced the target. As in the previous analysis, the intervals were presented in pairs, and the IO was required to identify the pair that contained the T+M stimulus in the first interval.

Auditory periphery model settings All analysis parameters were identical to those used in the previous analyses, with two exceptions. First, the characteristic frequency associated with each of the periphery models was set to 2 kHz to match the frequency of the target stimulus. Second, σ was set for each model so that the threshold of the IO matched that reported for the human listeners in quiet (21 dB SPL) in Zwicker (1984). The values for the σ parameter were set at 0.0024, 0.0460, 0.0294, and 0.2190 for the AIM, DRNL, Carney, and Zilany models, respectively.

Analysis implementation The implementation was identical to that in the previous analyses.

Analysis results and discussion

The results of this analysis, as well as the results from Zwicker (1984), are shown in Fig. 3. All four models predicted an increase in masked threshold as a function of masker duration, although the magnitude of that prediction varied considerably

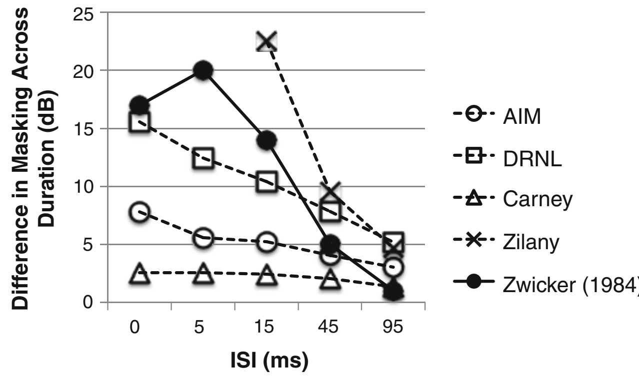

Fig. 3 Model predictions regarding masker duration. The figure does not include the points associated with the 0- and 5-ms ISIs for the Zilany model, because the difference in masking thresholds across masker durations was well over 100 dB in both of those conditions

across models. The AIM and Carney models consistently underpredicted that increase, at times by 15 dB or more. The DRNL model slightly overpredicted the effect of masker duration at longer ISIs. The Zilany model failed in this analysis: The amounts of masking in the long-duration 0- and 5-ms conditions were well over 100 dB. The duration of the masker appears to have an inappropriately large inhibitory effect on the intensity of the target signal that follows, at least when using these particular stimuli. Again, if humans do better than the IO, then humans have more information than the IO, suggesting a fault in the model. Although one might argue that the difference in methodology used to obtain the threshold measurements (a Békésy tracking procedure vs. 2IFC) makes them difficult to compare directly, predicting a greater than 100-dB increase in masking across duration conditions is clearly inappropriate.

Analysis III: Masker intensity

Not surprisingly, many researchers have reported the direct relationship between a masker's intensity and efficiency in a forward masking paradigm (e.g., Fastl, 1976; Jesteadt et al., 1982; Penner, 1974; Zwislocki, Pirouda, & Rubin, 1959). AN adaptation after the presentation of a masker has been found to increase in proportion to the intensity of the masker (Jesteadt et al., 1982). If a target is presented during this fatigued period, a relatively small response will be observed at the AN. This response will be more similar to the activity at the AN in response to silence. In the terms of the IO, neural adaptation decreases the magnitude of the differences between the T+M and M stimuli, thereby decreasing performance. Therefore, the IO should easily predict the effects of masker intensity on target detection thresholds.

Another related finding is less obvious, however: The rate of recovery from forward masking as ISI is increased is greater for higher masker intensities. In other words, the slope of the masking function in a forward masking experiment becomes more negative when the level of the masker is increased (Fastl, 1976; Moore & Glasberg, 1983; Oxenham & Moore, 1995; Zwicker, 1984). The final IO analysis was conducted to determine how well the periphery models predict these effects of masking intensity. The methodology of the experiment detailed in Jesteadt et al. (1982) was followed as closely as possible, so that the results of the simulation could be compared to the findings reported in that article. For simplicity, the results of the simulation in Analysis I served as the low-mask-intensity condition in the present simulation. Given the previous successes of the periphery at predicting forward masking effects, the IO was expected to exhibit both an overall increase in masking and a faster recovery from masking as masker intesity was increased.

Ideal-observer analysis method

Stimuli All stimuli used in this analysis were identical to those used in Analysis I, with the exception that a high-intensity mask condition was added in which the masker was presented at 80 dB SPL.

Auditory periphery model settings All stimuli were filtered through the same implementations of the periphery models that had been used previously. Characteristic frequencies for the models were set at 1 kHz, and the values of σ were the same as in Analysis I.

Analysis implementation The implementation was the same that had been used in Analysis I.

Analysis results and discussion

The results of the analysis are shown in Fig. 4. The high-intensity masker was more effective at all ISIs, regardless of the model of the periphery that was used. Each of them also showed a stronger release from masking with the high-intensity masker, indicated by the negative slopes in the figure. All four models underpredicted these effects, indicating that the models are consistent with the behavioral results. The consistent underprediction exhibited by all four models indicates either that the effects of increasing masker intensity have a central component in the human listener, or that the effects are peripheral and the models are deficient. An investigation of this question is beyond the scope of this study.

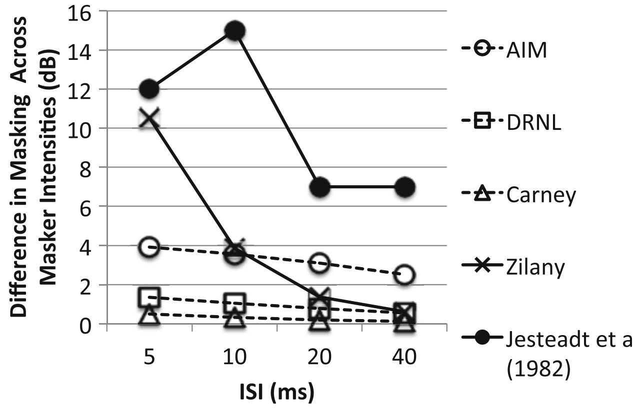

Fig. 4 Predicted differences in threshold as a function of ISI. Each point in the figure represents the difference in threshold between the high- and low-intensity masker conditions. Mirroring the results of Jesteadt et al. (1982), the models predict that this difference will decrease as ISI increases. All of the models underpredicted the effect of increasing masker intensity

Discussion

The results of the example IO analysis conducted in this article are straightforward: Models of the periphery are largely able to account for forward masking effects, although those effects are typically underpredicted relative to human data obtained through psychophysical methods. The models were completely unable to account for backward masking. These results correspond very well with the results from the empirical literature, highlighting the ability of the IO to predict behavioral results from physiological measurements.

A casual reader of this article might be tempted to conclude that since the various periphery model + IO combinations did a reasonable job of predicting forward masking effects, the IO is a good choice for a model of the human decision-maker, at least in a masking context. In fact, several models in the auditory masking domain do just that (e.g., Dau et al., 1996; Gresham & Collins, 1998), and the d' sensitivity measure from signal detection theory is directly based on the IO (Green & Swets, 1988). However, the idea that the human observer deliberately stores templates that correspond to each of 120 possible signal intensities seems implausible, even when the observer has knowledge of the intensity of the upcoming signal. Such a practice would not make sense from the observer's point of view: The observer is only required to distinguish between signals and noises, and not between signals of differing intensities. Realizing the strength of such criticisms, early users of the IO (e.g., Tanner, 1961) restricted the amount of information available to it, to see whether its performance matched that of human observers. Tinkering with the input in this way is equivalent to modeling higher-level processes, however. Changing the templates used by the IO essentially introduces a model of higher-level memory by assuming that the memory of a stimulus is stored in a particular format. This approach suffers from the same drawback that all modeling efforts suffer from: A belief in its assumptions is required before confidence in its conclusions can occur.

One might be tempted to conclude, then, that the IO is unlikely to be the true model of human decision-making, and therefore that any use of it is a waste of time. Treating the IO solely as a model of decision-making would be missing the point of this study, however. To demonstrate this, imagine that we actually conducted the auditory masking IO analysis to determine the true role of the periphery rather than to demonstrate an analysis methodology. In this case, the purpose of the analysis would be to determine whether the periphery itself was responsible for temporal masking effects. One possible way to test this would be to attach all possible higher-level models of decision-making to the output of the periphery and record the results. If all possible models predicted temporal masking effects, then the periphery must be responsible. If just one model could be found that did not predict temporal masking effects, then the periphery could not be responsible, since enough information was available at the output of the periphery so that at least one higher-level model was able to perform without any degradation in performance. The version of the IO used in this article (the signal-known-exactly version; see Tanner et al., 1960) is the decision-making model that maximizes performance. Therefore, if any subset of the higher-level models does not predict increased thresholds at small ISIs, the IO model must be an element of that subset. On the other hand, if the IO predicts increased thresholds at small ISIs then all other possible models must predict thresholds that are no lower than those predicted by the IO model.

Therefore, the strength of the IO arises from its optimality rather than its performance as a model of human decision-making. The IO identifies an upper bound on performance that can be used to draw strong conclusions about the filtering properties of the structure under study, be it the auditory periphery or any other physiological structure subject to measurement.

Appendix

The performance of the IO in a discrimination task for digitized stimuli is here derived analytically. On each trial, either an A or a B stimulus is presented to the IO. Let the digitized A and B stimuli be denoted by $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$, respectively, where n is the number of samples. The stimulus presented on a particular trial (the "incoming stimulus") is assumed to undergo some amount of degradation before it reaches the IO, which will be represented as additive noise (Green & Swets, 1988). Let the incoming stimulus be denoted by $\{x_1, \ldots, x_n\}$, and let e_i be the noise added to the ith sample of the incoming stimulus. The stimulus received by the IO is therefore $\{x_1 = a_1 + e_1, \ldots, x_n = a_n + e_n\}$ on A trials and $\{x_1 = b_1 + e_1, \ldots, x_n = b_n + e_n\}$ on B trials. Assume that the e_is are independent and identically distributed random variables that are sampled from a normal distribution with mean zero and variance σ^2 (see Green & Swets, 1988, for a justification of this assumption). In this case, the ith sample of the incoming stimulus will be normally distributed with mean a_i and variance σ^2 on A trials and mean b_i and variance σ^2 on B trials.

The task of the IO is to classify the incoming stimulus as either an A or a B stimulus. The IO generates a response by comparing the incoming stimulus to each of the templates, $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$. The comparison is conducted sample by sample in order to make use of all available information. According to the definition of the IO, all classifications are to be made using some monotonic transformation of the likelihood ratio. Since the variances of the A and B distributions are equal, and if the A and B presentation probabilities are each set to .5, the likelihood ratio associated with the incoming waveform (see Green & Swets, 1988, p. 163; see also Whalen, 1971, pp. 156–160) is

$$LR = \frac{\prod_{i=1}^{n} f(x_i|A)}{\prod_{i=1}^{n} f(x_i|B)} = \prod_{i=1}^{n} e^{\frac{(x_i - b_i)^2 - (x_i - a_i)^2}{2\sigma}}. \tag{A1}$$

The natural logarithm of this quantity is

$$\begin{aligned}\ln(LR) &= \sum_{i=1}^{n} \frac{(x_i - b_i)^2 - (x_i - a_i)^2}{2\sigma} \\ &= \frac{1}{2\sigma}\left[\sum_{i=1}^{n}(x_i - b_i)^2 - \sum_{i=1}^{n}(x_i - a_i)^2\right].\end{aligned} \tag{A2}$$

An unbiased IO will respond A if

$$\ln(LR) = \frac{1}{2\sigma}\left[\sum_{i=1}^{n}(x_i - b_i)^2 - \sum_{i=1}^{n}(x_i - a_i)^2\right] > 0, \tag{A3}$$

and will respond B otherwise. Rearranging the terms in Eq. A3, the unbiased IO will respond A when

$$\sum_{i=1}^{n}(x_i - b_i)^2 > \sum_{i=1}^{n}(x_i - a_i)^2, \tag{A4}$$

and will respond B otherwise. Let $SSE_A = \sum_{i=1}^{n}(x_i - a_i)^2$ and $SSE_B = \sum_{i=1}^{n}(x_i - b_i)^2$. Using this notation, the unbiased decision rule for the IO in this task is

$$\text{response} = \begin{cases} A & \textit{if } SSE_B > SSE_A \\ B & \textit{otherwise} \end{cases}. \tag{A5}$$

In words, the unbiased IO chooses the classification that is associated with the smallest sum of squared errors across all samples.

All that remains is to determine the probability of a correct response under this decision rule. On an A trial,

$$SSE_A = \sum_{i=1}^{n} (x_i - a_i)^2 = \sum_{i=1}^{n} (a_i + e_i - a_i)^2 = \sum_{i=1}^{n} e_i^2 \tag{A6}$$

and

$$SSE_B = \sum_{i=1}^{n} (x_i - b_i)^2 = \sum_{i=1}^{n} (a_i + e_i - b_i)^2. \tag{A7}$$

The probability of a correct response, given that an A stimulus was presented, is therefore

$$p(\text{correct}|A) = p(SSE_A < SSE_B|A) = p\left(\sum_{i=1}^{n} e_i^2 < \sum_{i=1}^{n} (a_i + e_i - b_i)^2\right). \tag{A8}$$

Distributing the exponent on the right-hand side of the inequality and rearranging the terms leads to

$$p(\text{correct}|A) = p\left(-2\sum_{i=1}^{n}(a_i - b_i)e_i < \sum_{i=1}^{n} (a_i - b_i)^2\right). \tag{A9}$$

The elements of $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$ are known constants, so

$$(a_i - b_i)e_i \sim \mathrm{N}\left[0, \sigma^2 (a_i - b_i)^2\right]. \tag{A10}$$

It follows that

$$\sum_{i=1}^{n}(a_i - b_i)e_i \sim \mathrm{N}\left[0, \sigma^2 \sum_{i=1}^{n} (a_i - b_i)^2\right], \tag{A11}$$

and therefore that

$$-2\sum_{i=1}^{n}(a_i - b_i)e_i \sim \mathrm{N}\left[0, 4\sigma^2 \sum_{i=1}^{n} (a_i - b_i)^2\right]. \tag{A12}$$

Therefore,

$$p(\text{correct}|A) = \Phi\left(\frac{\sum_{i=1}^{n} (a_i - b_i)^2}{\sqrt{4\sigma^2 \sum_{i=1}^{n} (a_i - b_i)^2}}\right) = \Phi\left(\frac{\sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}}{2\sigma}\right), \tag{A13}$$

where Φ is the cumulative distribution function associated with the standard normal. As expected, the probability of a correct response on A trials increases as the difference between the A and B templates increases.

A similar procedure allows for the calculation of the probability of a correct response on B trials. In this case,

$$p(\text{correct}|B) = p(SSE_B < SSE_A|B) = p\left(\sum_{i=1}^{n} e_i^2 < \sum_{i=1}^{n} (b_i + e_i - a_i)^2\right). \tag{A14}$$

Distributing the square and rearranging terms allows this equation to be reduced to

$$\begin{aligned} p(\text{correct}|B) &= p\left(-2\sum_{i=1}^{n}(b_i - a_i)e_i < \sum_{i=1}^{n} (b_i - a_i)^2\right) \\ &= p\left(2\sum_{i=1}^{n}(a_i - b_i)e_i < \sum_{i=1}^{n} (a_i - b_i)^2\right). \end{aligned} \tag{A15}$$

Using the same logic as above, this equation further reduces to

$$p(\text{correct}|B) = \Phi\left(\frac{\sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}}{2\sigma}\right). \tag{A16}$$

Therefore, $p(\text{correct}|A) = p(\text{correct}|B)$, and

$$\begin{aligned} p(\text{correct}) &= p(\text{correct}|A)p(A) + p(\text{correct}|B)p(B) \\ &= \Phi\left(\frac{\sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}}{2\sigma}\right)(p(A) + p(B)) \\ &= \left(\frac{\sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}}{2\sigma}\right). \end{aligned} \tag{A17}$$

References

Carney, L. H. (1993). A model for the responses of low-frequency auditory-nerve fibers in cat. *Journal of the Acoustical Society of America, 93,* 401–417.

Dau, T., Puschel, D., & Kohlrausch, A. (1996). A quantitative model of the "effective" signal processing in the auditory system: I. Model structure. *Journal of the Acoustical Society of America, 99,* 3615–3622.

Duifhuis, H. (1973). Consequences of peripheral frequency selectivity for nonsimultaneous masking. *Journal of the Acoustical Society of America, 54,* 1471–1488.

Elliott, L. L. (1962). Backward masking: Monotic and dichotic conditions. *Journal of the Acoustical Society of America, 34,* 1108–1115.

Fastl, H. (1976). Temporal masking effects: I. Broad band noise masker. *Acustica, 35,* 287–302.

Fastl, H. (1977). Temporal masking effects: II. Critical band noise masker. *Acustica, 36,* 317–331.

Fastl, H. (1979). Temporal masking effects: III. Pure tone masker. *Acustica, 43,* 282–294.

Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics*. Los Altos Hills, CA: Peninsula.

Gresham, L. C., & Collins, L. M. (1998). Analysis of the performance of a model-based optimal auditory signal processor. *Journal of the Acoustical Society of America, 103,* 2520–2529.

Harris, D. M., & Dallos, P. (1979). Forward masking of auditory nerve fiber responses. *Journal of Neurophysiology, 42,* 1083–1107.

Jepsen, M. L., Ewert, S. D., & Dau, T. (2008). A computational model of human auditory signal processing and perception. *Journal of the Acoustical Society of America, 124,* 422–438.

Jesteadt, W., Bacon, S. P., & Lehman, J. R. (1982). Forward masking as a function of frequency, masker level, and signal delay. *Journal of the Acoustical Society of America, 71,* 950–962.

Kidd, G., & Feth, L. L. (1982). Effects of masker duration in pure tone forward masking. *Journal of the Acoustical Society of America, 72,* 1384–1386.

Liu, Z., & Kersten, D. (1998). 2D observers for human 3D object recognition? *Vision Research, 38,* 2507–2519. doi:10.1016/S0042-6989(98)00063-7

Luescher, E., & Zwislocki, J. (1949). Adaptation of the ear to sound stimuli. *Journal of the Acoustical Society of America, 21,* 135–139.

Meddis, R., O'Mard, L. P., & Lopez-Poveda, E. A. (2001). A computational algorithm for computing nonlinear auditory frequency selectivity. *Journal of the Acoustical Society of America, 109,* 2852–2861.

Miller, R. L. (1947). Masking effect of periodically pulsed tones as a function of time and frequency. *Journal of the Acoustical Society of America, 19,* 798–807.

Miyazaki, K., & Sasaki, T. (1984). Pure-tone masking patterns in nonsimultaneous masking conditions. *Japanese Psychological Research, 26,* 110–119.

Moore, B. C. J., & Glasberg, B. R. (1983). Growth of forward masking for sinusoidal and noise maskers as a function of signal delay: Implications for suppression in noise. *Journal of the Acoustical Society of America, 73,* 1249–1259.

Nelson, P. C., Smith, Z. M., & Young, E. D. (2009). Wide-dynamic-range forward suppression in marmoset inferior colliculus neurons is generated centrally and accounts for perceptual masking. *Journal of Neuroscience, 29,* 2553–2562. doi:10.1523/JNEUROSCI.5359-08.2009

Oxenham, A. J., & Moore, B. C. J. (1995). Additivity of masking in normal hearing and hearing-impaired subjects. *Journal of the Acoustical Society of America, 98,* 1921–1935.

Oxenham, A. J., & Plack, C. J. (2000). Effects of masker frequency and duration in forward masking: Further evidence for the influence of peripheral nonlinearity. *Hearing Research, 150,* 258–266.

Pastore, R. E., & MacLatchy, C. S. (1975). Some nonmasking auditory postsignal effects. *Perception & Psychophysics, 17,* 455–459. doi:10.3758/BF03203293

Patterson, R. D. (2001). DSAM: Development system for auditory modeling. Available at http://dsam.org.uk (date last viewed 7/12/2011).

Patterson, R. D., Allerhand, M. H., & Giguere, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *Journal of the Acoustical Society of America, 98,* 1890–1894.

Penner, M. J. (1974). Effect of masker duration and masker level on forward and backward masking. *Journal of the Acoustical Society of America, 56,* 179–182.

Recio, A., & Rhode, W. S. (2000). Basilar membrane responses to broadband stimuli. *Journal of the Acoustical Society of America, 108,* 2281–2298.

Relkin, E. M., & Turner, C. W. (1988). A reexamination of forward masking in the auditory nerve. *Journal of the Acoustical Society of America, 84,* 584–591.

Samoilova, I. K. (1959). Masking of short tone signals as a function of the time interval between masked and masking sounds. *Biofizika, 4,* 44–52.

Shul'gina, G. L., & Murav'ev, A. A. (2000). Analysis of the effect of backward masking by means of a complex model of a net of nuron-like elements. *Human Physiology, 26,* 34–40.

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research, 261,* 22–29.

Tanner, W. P., Jr. (1961). Application of the theory of signal detectability to amplitude discrimination. *Journal of the Acoustical Society of America, 33,* 1233–1244. doi:10.1121/1.1908962

Tanner, W. P., Jr., Birdsall, T. G., & Clarke, F. R. (1960). *The concept of the ideal observer in psychophysics* (Tech. Rep. 98). Ann Arbor, MI: University of Michigan, Electronic Defense Group.

Whalen, A. D. (1971). *Detection of signals in noise*. New York: Academic Press.

Zilany, M. S. A., & Bruce, I. C. (2006). Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. *Journal of the Acoustical Society of America, 120,* 1446–1466. doi:10.1121/1.2225512

Zilany, M. S. A., & Bruce, I. C. (2007). Cat auditory nerve model. Available at http://senselab.med.yale.edu/modeldb/showmodel.asp?model=87751 (date last viewed 7/20/2011).

Zwicker, E. (1984). Dependence of post-masking on masker duration and its relation to temporal effects in loudness. *Journal of the Acoustical Society of America, 75,* 219–223.

Zwislocki, J., Pirouda, E., & Rubin, H. (1959). On some poststimulatory effects at the threshold of audibility. *Journal of the Acoustical Society of America, 31,* 9–14.