

Varying Feedback to Evaluate Detection Strategies: The Detection of a Tone Added to Noise

VIRGINIA M. RICHARDS

Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Received: 6 July 2001; Accepted: 8 October 2001; Online publication: 9 January 2002

ABSTRACT

The effectiveness of rewarding strategies associated with different models of detection was evaluated for a tone-in-noise detection task. In three experiments, a tone was added to a narrowband noise masker in a two-interval, forced-choice procedure with feedback. In the first experiment the observer's task was to detect which interval had the added tone. In the second experiment the task was to indicate which interval was more intense. In the third experiment the task was to indicate which interval had the sound with the less modulated envelope. In terms of correct signal detection, thresholds were nearly the same for all three experiments. However, for three of the observers the psychometric functions indicate that observers altered their responses in accordance with the changes in feedback contingencies. This indicates that neither changes in level nor changes in envelope pattern alone can account for the detection of a tone added to noise for these observers. The fourth observer was sufficiently insensitive that the pattern of feedback associated with the level and envelope models did not differ substantially, and for her feedback had little impact on behavior. The results indicate that changes in feedback provide a means of testing alternative psychophysical models. Moreover, the results suggest that it is difficult to model the detection of a tone added to noise in terms of a single cue or a simple combination of cues. Observers are free to choose from at least two cues, and can do so.

Keywords: masking, strategies

INTRODUCTION

Even in simple psychophysical detection and discrimination tasks, observers' decisions might depend on changes in more than one stimulus dimension. This is the situation for the experimental task considered here, the detection of a tone added to a narrow band of noise. Adding a tone to a narrow band of noise leads to, on average, a more intense stimulus (Green and Swets 1966), a stimulus with a less modulated envelope (Moore 1975; Hartmann and Pumplin 1988, Richards, 1992), and a stimulus with a fine structure that is more regular (tends toward periodicity; Moore 1975; Richards 1992). Whether considering the detection of a tone added to noise or some other psychophysical experiment, an important step is to determine whether observers use one or more of the available cues. Though more difficult, it is also important to evaluate the likelihood that the observers' strategies are stationary rather than varying depending on experience, context, or the particular stimulus tokens presented on any one trial or group of trials.

To address these issues experimenters generally adopt one of two approaches. One approach is to vary the stimuli in a controlled way. The experimenter might construct the stimuli so that one or another cue is either constant or random across the stimulus set. Such manipulations discourage observers from incorporating that cue into their decisions regarding, for example, the presence of an added tone. A second approach involves *post hoc* analyses of the stimuli used in the experiment and the resulting data. Traditionally, *post hoc* analyses have relied on probabilistic models of the stimuli and observers' responses. With the increased stimulus control and data analysis afforded by digital signal processing and the relatively large

Correspondence to: Dr. Virginia M. Richards • Department of Psychology • 3815 Walnut Street • University of Pennsylvania • Philadelphia, PA 19104. Telephone: (215) 898-4587; fax: (215) 898-7301; email: richards@cattell.psych.upenn.edu

data-saving opportunities afforded by PCs, experimenters can now evaluate models in terms of observers' responses to particular stimuli (Ahumada and Lovell 1971). A third approach, pioneered by Nachmias (2001), is a more interventionist, online approach. Simply put, one can evaluate the accuracy of a model by providing observers with feedback appropriate for that model. This, of course, assumes that observers incorporate feedback into their decisions.

To appreciate the feedback-based approach applied to a tone-in-noise task, imagine that the detectability of the added signal depends on a change in stimulus intensity. In a typical two-interval, forced-choice (2IFC) procedure, the feedback provided to the observer indicates which of the two intervals actually has the added tone (correct answer feedback). In a second experiment the feedback is changed to indicate which of the two intervals has the more intense stimulus. The feedback in this case is associated with a certain model (or cue) and not necessarily the presence or absence of a tone. Often, but not always, the level-based feedback will be concordant with the tone-present feedback. If observers do rely exclusively on changes in level to detect the added tone, threshold signal levels in these two experiments would be the same. The level-based model for the detection of a tone added to noise would be rejected if reliable deviations in observers' responses were obtained depending on feedback contingencies. More generally, if Model A is a potential model for Task I, an effective means of testing the model is to provide feedback to the observer using the Model A prediction and to compare the results with the results obtained in Task I.

The primary advantage of the feedback method, compared with methods in which the stimuli are controlled or models are evaluated in a *post hoc* fashion, is that changes in feedback allow a direct evaluation of a model. Compared with methods in which stimuli are varied, the feedback method is advantageous because the same stimulus set is tested across different conditions. Compared with *post hoc* fits of a model to psychophysical data, an advantage of the feedback-based method is that there is no need to estimate an "internal" noise. This is because the observers' own data, which include internal noise, are used to evaluate the model of interest.

A tone-in-noise task was chosen to examine the usefulness of the feedback-based approach because it is known that observers are sensitive to the changes in level and the changes in envelope pattern associated with the addition of a threshold-level tone (Richards and Nekrich 1993). Provided the feedback-based approach generates sufficiently sensitive data, it was reasonable to expect that the current results would

also demonstrate sensitivity to changes in both level and envelope patterns.

Three experiments are reported below. In Experiment I, feedback indicating the presence of the signal tone is used. This experiment is referred to as the tone-in-noise, or tone, experiment. In Experiment II, feedback indicates the stimulus with the higher level. This is referred to as the level experiment. In Experiment III, the envelope experiment, feedback indicates the stimulus with the less modulated envelope. Regardless of the feedback, the same adaptive tracking procedure was used in which the level of the added tone was altered. In all three experiments psychometric functions relating the proportion of correct *signal detections* and signal level formed the data set. That is, after completing each experiment, the data were evaluated in terms of the proportion of times the observer correctly indicated the interval containing the added signal. By comparing psychometric functions between the first (tone-in-noise) and second (level) or first and third (envelope) experiments, one can evaluate the efficacy of the level and envelope models for the detection of a tone added to noise.

A variety of results might occur. For example, if the psychometric functions for Experiments I and II were the same, the result would support the level model for the detection of a tone added to noise; or the psychometric functions for Experiments I and III might be the same, indicating support of the envelope model; or psychometric functions in the three experiments might be about the same, indicating a lack of support for one model over another. Anticipating such a situation, the maskers were chosen so as to emphasize level and envelope cues. As a result, observers' responses could be partitioned conditional on whether the stimuli in the two intervals had large versus small changes in either level or envelope patterns. As described in detail below, compared with psychometric functions based on all trials, "extracted" psychometric functions provide a more sensitive indication of the impact of feedback contingencies. Whatever the ultimate results of the experiments, it is important to appreciate that if observers alter their responses depending on feedback contingencies, it follows that more than one strategy is available to the observer in traditional tone-in-noise experiments.

The narrowband noise maskers tested are sufficiently narrow, 60 and 120 Hz, that the "mean" envelope modulation rates (approximately half the bandwidth) were within the range over which observers can readily detect the introduction of modulation (Viemeister 1979). Under these circumstances, both level-based and envelope-based models of detection are fitting. The contribution of fine structure cues is inferred by evaluating data collected using stimuli centered at 1000 and 5500 Hz. Phase locking is

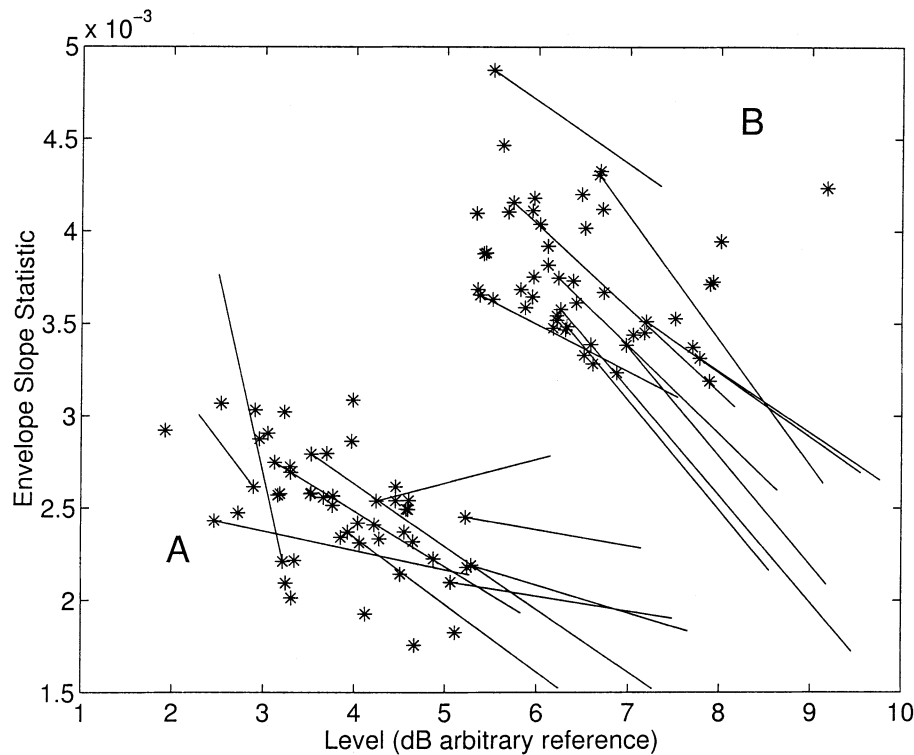


FIG. 1. One hundred maskers are represented on a level-by-slope plot using asterisks. The maskers are divided into two categories, labeled A and B. The rays emanating from some of the masker coordinates terminate at the coordinates associated with signal-plus-masker stimuli for a threshold-level signal.

assumed to be weak or absent at the higher center frequency.

In the next section the masker sets are described. Then the use of feedback as a means of testing alternative models of detection is demonstrated by way of computer simulation results. Finally, the experimental methods and results are presented.

STIMULI AND SIMULATION DEMONSTRATIONS

Maskers

Four types of maskers were tested. The maskers had center frequencies of 1000 and 5500 Hz and bandwidths of 60 and 120 Hz. For each center frequency/bandwidth condition, 100 maskers were tested. The methods used to generate the maskers were the same for each center frequency/bandwidth combination.

The asterisks in Figure 1 show one representation of the 100 60-Hz-wide maskers centered at 1000 Hz. The maskers are plotted in two dimensions; the abscissa is level in dB (using an arbitrary reference) and the ordinate is a measure of envelope modulation, the normalized envelope slope statistic of Richards (1992). The latter statistic involves the extraction of the stimulus envelope, normalization of the envelope to a mean of 1, and subsequent averaging of the absolute value of the "slope" of the envelope as estimated

using differences between adjacent envelope values. Unlike in the original model, low-pass filtering of the stimulus envelope is not included. The lines drawn in Figure 1 show the effect of adding a tone to the maskers, and will be considered below.

The 100 maskers depicted in Figure 1 were chosen from 300 bandpass Gaussian noises. For the original cohort of maskers, the level-by-slope plot is similar to that shown in Figure 1 except that the larger set of maskers form a circular cloud of points. The 100 maskers shown in Figure 1 were extracted from the original set of 300 by limiting the maskers so as to occupy the lower-left and upper-right quadrants and by choosing maskers distant from the center of the original "cloud." The maskers in the lower-left-hand quadrant have small amounts of envelope modulation and sound pressure levels that are relatively low. These maskers comprise masker set "A." For the masker in the upper-right quadrant, the levels are relatively high and the envelopes are more modulated. These maskers will be referred to as masker set "B."

The lines in Figure 1 indicate the effect of adding a signal with a level approximately at threshold (Experiment I, below) to several of the maskers. The lines might be thought of as indicating the trajectory of change in the level-by-slope domain associated with the addition of a 1000-Hz tone. As described above, on average the impact of adding a tone to a narrowband noise masker is to reduce the stimulus' envelope modulation and increase its level. In Figure 1

such a change is indicated by a trajectory downward and to the right. For the sample shown in Figure 1, such a trajectory is especially compelling for the B maskers.

Note that when a tone is added to an A masker, the tone-plus-noise stimulus may have a lower level than a B masker presented alone. If the detection of a tone added to noise depends on intensity differences across intervals, when a tone is added to an A masker and the nonsignal interval contains a B masker, the observer would be expected to generate incorrect decisions more often than, for example, when a tone is added to a B masker and the nonsignal sound is also drawn from among the B maskers. Conversely, an A masker might be less modulated than a B masker even after a tone is added to a B masker. As a result, an observer who depends on changes in envelope modulation to detect an added tone will perform relatively poorly on trials in which the signal is added to a B masker and the nonsignal interval has an A masker. This means that depending on the decision rule used by the observer, one would expect different responses depending on whether the tone is added to an A masker or a B masker and whether the nonsignal interval is an A or a B masker.

Feedback variation/simulations

The current experiments deviate from traditional experiments in that feedback was altered so as to encourage one or another strategy. In the three experiments, feedback indicated which interval had the added tone, which was more intense, and which had the smallest slope statistic (less modulation). A 2IFC procedure was used, and a signal tone was always present in one of the two intervals. Depending on whether the observer was correct according to the feedback rule applied by the experimenter, the *level of the added tone* was increased, decreased, or unaltered according to a 2-down, 1-up tracking procedure (Levitt 1971). Thus, the same tracking algorithm was used in all three experiments; only the definitions of “correct” and “incorrect” responses varied. The data of primary interest were the proportion of responses corresponding to the interval containing the added tone. Observers’ sensitivity to, for example, changes in level (Experiment II) is also extracted from the data set.

To appreciate the impact of the choice of maskers, consider simulation results for the level and envelope models shown in Figure 2. These results are for the detection of a 1000-Hz signal added to a 120-Hz-wide masker. For each signal level tested, 1000 two-interval, forced, choice trials were simulated. The maskers were chosen at random from the 100 maskers stored to disk (the same maskers as in the experiments). Two decision rules were tested. For the level decision rule

(circles), it is assumed that the more intense (digital RMS value) waveform had the signal. For the envelope decision rule (squares), the waveform with the smaller envelope slope statistic was chosen as having the added tone. Note that the ordinate ranges from 0 to 1, not from 0.5 (chance) to 1.

The leftmost panel of Figure 2 shows the psychometric functions across all trials. The middle and right panels show results extracted from that data set. For the middle panel, by chance the signal was added to a masker drawn from set A and the other masker was drawn from set B (indicated as “A+S, B Trials”). The right panel shows results for another subset of the data; trials in which the signal is added to a B masker and the no-signal masker were drawn from set A (indicated as “B+S, A Trials”).

Across all trials, the simulated psychometric functions indicate that the level model predicts the presence of a tonal signal slightly better than envelope modulation (left panel; circles above squares). The middle panel shows that when the signal is added to an A masker and a B masker is presented in the other interval, the envelope rule always provides correct responses, and for low signal levels, the level model leads to many wrong answers. The opposite pattern occurs when the signal is added to a B masker and the non-signal interval has an A masker.

Whatever the results for the detection of a tone added to noise (Experiment I), it is of interest to determine whether observers are capable of altering their decision strategies. If level- and envelope-based feedback leads to changes in strategies, changes in psychometric functions will result. Indeed, because the feedback provided to the observers was used to generate the level and envelope predictions shown in Figure 2, precise matching to feedback would lead to psychometric functions like those shown in Figure 2.

METHODS

Stimuli and Presentation

Masker generation, fits to psychometric functions, and simulations were prepared using Matlab (Math Works, Natick, MA) software. Maskers were generated by band-limiting Gaussian noise. The maskers, of bandwidths nominally 60 and 120 Hz, were generated using an FFT and setting the magnitude of frequencies outside the bandpass to zero. The center frequencies of the maskers were 1000 and 5500 Hz and the signals were cosine phase. Signal attenuation was achieved digitally. The stimuli were synthesized using a 25,000 cycles/s sample rate, presented via a 16-bit digital-to-analog converter and passed through a 8-kHz low-pass filter with attenuation skirts of approximately 90 dB

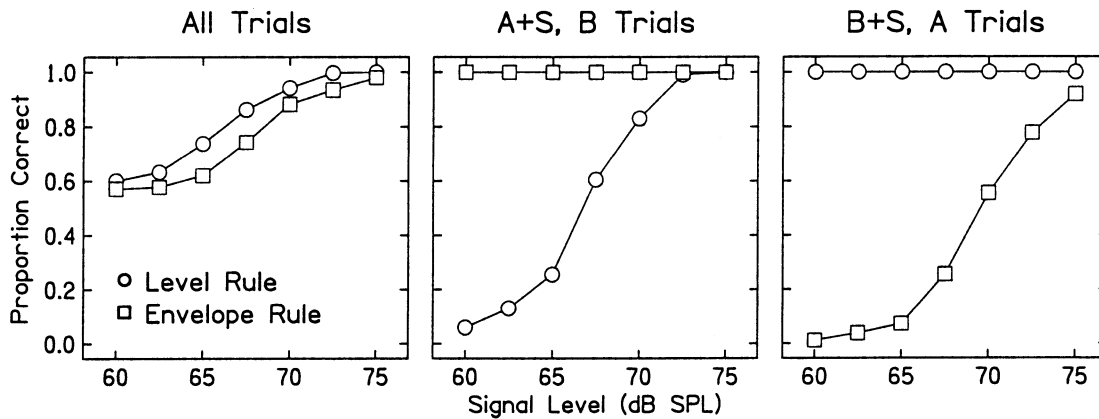


FIG. 2. Simulation-based predictions of psychometric functions for level (circle) and envelope (square) decision rules are shown in three ways. The left panel shows psychometric functions based on all trials. The middle panel is for trials in which the signal is added to an A

masker and the no signal stimulus is a B masker. The right panel is for trials in which the signal is added to a B masker and the no signal stimulus is an A masker.

per octave (Kemo VBF 8). After filtering, the stimuli were passed through two attenuators and presented diotically via Sennheiser HD410 headsets. The signal and masker had 5-ms raised-cosine ramps and a total duration of approximately 165 ms. The maskers were presented at an average spectrum level of 50 dB SPL.

Stimuli were presented using a 2IFC procedure. On any interval, the masker was randomly chosen, so A and B maskers were equally likely. Signal levels were altered using a 2-down, 1-up tracking procedure. The tracking procedure initially altered signal levels by 4 dB, and the step size was reduced to 2 dB following the third turnabout. The track was terminated after 50 trials.

Observers and Procedures

Four observers, aged 19–25 years, participated in the experiments. All had thresholds in quiet of 15 dB HL or better for frequencies ranging from 500 to 8000 Hz, except Obs 3 whose threshold at 500 Hz was 25 dB HL. Observers 2–4 were paid for participation; Obs 1 participated to gain laboratory experience. In order to provide data for the detection of a tone added to noise in a relatively “naïve” group of observers, all observers ran in Experiment I first. Observers 1 and 2 then ran in Experiment II followed by Experiment III; Obs 3 and 4 ran in Experiment III and then Experiment II. Observers 3 and 4 had previously participated in psychophysical experiments, including the detection of a tone added to wideband noise. The other observers had no past experience in auditory psychophysical experiments, although Obs 2 had recently participated in a vision experiment.

Before data collection in each experiment, observers practiced for 3–4 hours during which time all conditions were tested. Oral instructions were provided

prior to practice in each experiment. Prior to Experiment I, observers were told that the task was to detect an added tone. Prior to Experiment II, observers were told their task was to choose the louder of two sounds, regardless of sound quality. Prior to Experiment III, observers were told that the signal interval was the one that was clearer, or less rough.

Data collection was blocked by condition, i.e., observers completed one of the four conditions (two signal frequencies and two bandwidths) before moving on to the next. For Experiment I, at least 25 tracks were run (1250 total trials). For Experiments II and III, the tracks ran to lower signal levels than encountered in Experiment I, and so 30 tracks were run. Before moving on to the next condition, standard errors of the means for the ultimate and penultimate 10 track-based thresholds were evaluated to test for practice effects. On rare occasions practice effects were present. In these instances, at least 5 additional tracks were obtained.

Psychometric functions

The data from the adaptive tracking procedure were analyzed as psychometric functions using the function and procedure suggested by Dai (1995). Reported thresholds are in terms of the signal level required for detection regardless of feedback contingencies. The psychometric function $d' = (x/\alpha)^\beta$, where x is the signal level in dB, requires two free parameters α and β . Threshold is the value of α at which $d' = 1$. The d' value was estimated from the percent correct scores using a numerical approximation to the p -to- z transform (Abramowitz and Stegun 1964). The value minimized in the fitting procedure was a weighted sum of the squared errors, where the weights were associated

TABLE 1

Threshold estimates ^a												
SF BW	Tone				Level				Envelope			
	1000 60	1000 120	5500 60	5500 120	1000 60	1000 120	5500 60	5500 120	1000 60	1000 120	5500 60	5500 120
Obs 1	67.3	68.9	67.4	69.3	66.7	67.8	66.2	67.9	69.4	69.9	70.1	70.5
Obs 2	66.4	68.6	66.4	67.9	66.4	68.3	66.9	68.1	68.3	68.2	68.3	69.1
Obs 3	67.1	67.8	66.9	68.4	66.9	67.6	67.6	70.1	66.3	68.1	68.9	69.7
Obs 4	70.6	69.3	72.0	75.1	70.0	72.8	72.7	73.3	72.6	72.0	75.8	76.1
AVG	67.9	68.6	68.2	70.2	67.5	69.2	68.3	69.8	69.2	69.6	70.8	71.3
SEM	0.9	0.3	1.3	1.7	0.8	1.2	1.5	1.2	1.3	0.9	1.7	1.6

^aRows show the estimated thresholds for each observer, the averaged thresholds, and the standard errors of the means across 4 observers. The columns are grouped by type of feedback, signal frequency (SF) and masker bandwidth (BW).

with binomial variance [Dai (1995) described the procedure for the function used in the current study]. In most cases the fits accounted for at least 95% of the available variance.

In addition to psychometric functions relating signal level and percent correct signal detection, psychometric functions appropriate to feedback were also generated. For Experiment II, the psychometric functions related percent correct level discrimination and ΔL in dB. For Experiment III the psychometric functions plotted percent correct modulation discriminations as a function of changes in modulation depth (using the slope statistic). Binning the independent variable and then determining the percent correct for each bin yielded psychometric functions. Then the method suggested by Dai was used. Although the goodness of the fit varied with bin size, threshold estimates were relatively invariant. The thresholds shown below are for 10 bins, which led to good fits to the data; at least 90% of the variance in the data is accounted for by the fitted psychometric functions.

RESULTS

Table 1 shows threshold estimates for Experiments I–III. The columns refer to the different experiments (Experiments I–III, respectively) and experimental conditions. The rows show results for different observers with the final two rows indicating the means and standard errors of the means across observers. The dominant result is that threshold estimates are remarkably consistent across observers and experiments. The most notable deviation from this pattern is the relatively poorer sensitivity exhibited by Obs 4. Feedback appropriate to both level and envelope cues led to thresholds approximately equal to those estimated using the traditional tone-in-noise task. Bos and de Boer (1966) also report that for bandwidths similar to those used here, thresholds for the detection of a tone

added to a narrowband masker and the detection of a narrowband noise added to a narrowband masker (intensity discrimination task) are approximately equal. The current data are also consistent with the results of Richards and Nekrich (1993), who used stimulus manipulations to evaluate observers' sensitivity to changes in level and level-invariant cues in a tone-in-noise task.

Although thresholds appear similar, an ANOVA revealed several statistically significant effects. The within-subjects design was analyzed using observers as a random variable and experiment, signal frequency, and bandwidth as fixed variables. The main effects of bandwidth ($F_{(1,3)} = 106, p < 0.005$) and experimental condition ($F_{(2,6)} = 7.6, p < 0.05$) were significant, as was the bandwidth-by-experiment interaction ($F_{(2,6)} = 7.2, p < 0.05$). The main effect of frequency and the remaining interactions did not approach significance.

Figure 3 shows the significant effects revealed by the ANOVA. The data are collapsed across observers and frequency. Thresholds in Experiments I–III (tone-in-noise, level, and envelope feedback, respectively) are shown using asterisks, circles, and squares. The significant bandwidth-by-experiment interaction reflects the smaller impact of bandwidth on thresholds in Experiment III compared with Experiments I and II. That the function relating threshold and bandwidth is flat for Experiment III (envelope) is largely meaningless because the bandwidths were chosen, to a large degree, to assure this result. Overall, while the effects of bandwidth, experiment, and their interactions are statistically significant, there is little information to be gleaned from this analysis.

Figure 4 shows the results of Experiments I–III (tone-in-noise, level, and envelope, respectively) for Obs 1 when a 1000-Hz signal was added to a 120-Hz-wide masker. The top, middle, and bottom rows show results for Experiments I–III, respectively. Note that the ordinate indicates proportion correct "tone" detections regardless of feedback contingencies. The format

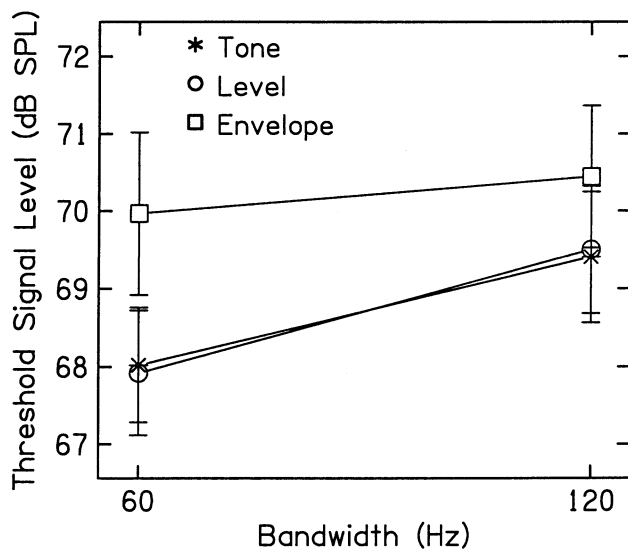


FIG. 3. Data averaged across signal frequencies and observers. Masker bandwidth is indicated on the abscissa and threshold signal level is indicated on the ordinate. Results for Experiments I–III are shown using asterisks, circles, and squares, respectively. Standard errors of the mean are across observers and frequencies.

follows that of Figure 2. The leftmost panels show psychometric functions across all trials. The middle panels show results for trials in which the signal was added to masker type A and the nonsignal interval contained a B masker. The right panels show results for trials in which the signal was added to masker type B and the other interval contained an A masker. In the left panels a dark line shows the psychometric function fitted to Obs 1's data. Error bars indicate estimates of the standard deviations assuming Bernoulli trials. Proportion correct is plotted only for signal levels for which there were at least 6 trials contributing to the estimate of proportion correct.

Simulation-based predictions for the level and envelope models were generated for the exact trials encountered by Obs 1. The predictions are not plotted because the results are very similar to those shown in Figure 2. Also, it should be noted that although psychometric functions are not plotted, trials in which both intervals contained A maskers and trials in which both intervals contained B maskers also occurred in these experiments.

For Experiments II and III the tracking procedure reduced the signal level well below the signal levels encountered in Experiment I. A plausible explanation for the smaller range of signal levels encountered in Experiment I vs. Experiments II and III is as follows. When the signal levels are small, the two intervals have detectable differences in level and/or envelope modulation, but those differences do not reliably indicate which interval has the added tonal signal. One implication of the increase in the range of signal levels

tested in Experiments II and III compared with Experiment I is that Obs 1 altered his responses depending on feedback. Moreover, comparing the results for Obs 1 with the predictions plotted in Figure 2, changes in feedback produce extracted psychometric functions that tend toward the expected pattern. When feedback was based on differences in level, Obs 1 often performed below chance levels when the signal to be detected was added to an A masker and the nonsignal interval contained a B masker (second row, middle panel). When the signal was added to a B masker and the nonsignal interval contained an A masker, Obs 1 performed above chance levels (second row, right panel). Obs 1's data from Experiment III also conform to the predictions of the envelope model shown in Figure 2 (compare bottom row of Fig. 4 and the squares of Fig. 2). Another way to appreciate the impact of feedback is to note that changes in the pattern of responses as one progresses down the panels of the second and third columns. The results shown in Figure 4 are as representative of the observers' data as any one observer's results. Next, more condensed plots are shown for all observers.

Figures 5–8 show a subset of the results for Obs 1–4, respectively. Only psychometric functions for conditions in which the signal was added to masker A and the other interval contained a B masker, or the opposite, are shown. Asterisks, circles, and squares indicate the results of Experiments I–III (tone present, level, and envelope feedback), respectively. Error bars indicate standard deviations using an assumption of Bernoulli trials. The data are plotted only for signal levels similar to those tested in Experiment I. The rows are for the different conditions: signal frequency/masker bandwidth values of 1000/60, 1000/120, 5500/60, and 5500 Hz/120 Hz running from top to bottom. Thus, the second row of Figure 5 collapses the data of Figure 4 onto two graphs.

For Obs 1–3 (Figs. 5–7), three features are apparent. First, observers alter their responses depending on feedback. This is apparent in the fact that the results of Experiments I–III generate different psychometric functions (asterisks, circles, and squares, respectively). Second, the way in which Obs 1–3 alter their responses is consistent with feedback contingencies. That is, the results follow the predictions shown in Figure 2. Reading from the first column of Figures 5–7, when a signal is added to an A masker and the nonsignal interval contains a B masker, feedback consistent with an envelope decision rule (squares) leads to a higher proportion of correct signal detections than when the feedback is consistent with the level decision rule (circles). When the signal is added to a B masker and the nonsignal interval contains an A masker, the opposite pattern occurs. The third feature of note is that neither the level nor the envelope model can account for

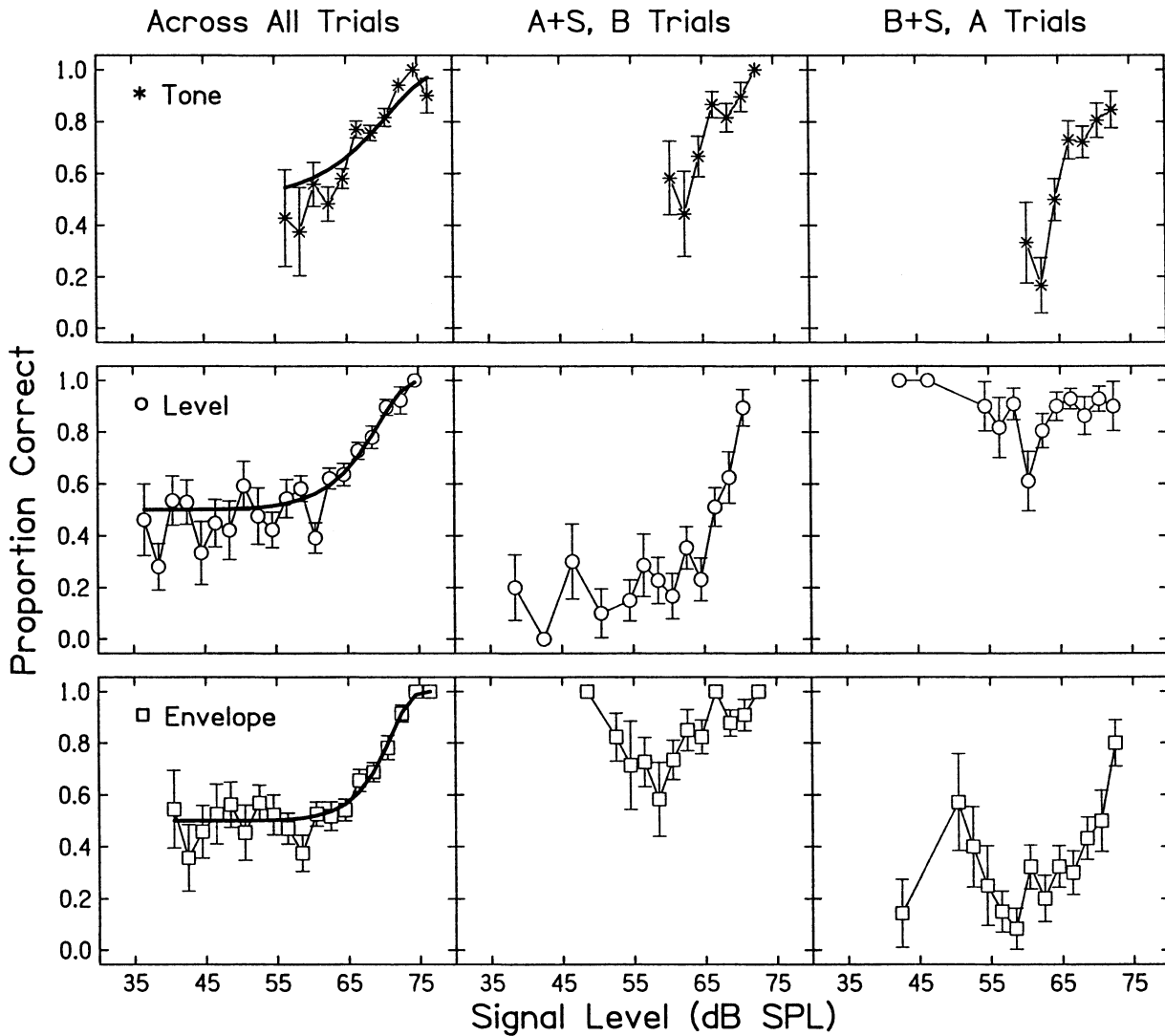


FIG. 4. Using the format of Figure 2, psychometric functions for Obs 1 are shown. The signal frequency is 1000 Hz and the masker bandwidth is 120 Hz. The first row is for Experiment I (asterisks), the second row is for Experiment II (circles), and the third row is for Experiment III (squares). Psychometric functions fitted to all trials are shown in the leftmost panels.

observers' responses when the observer was instructed to detect a tone added to noise (asterisks). For the results shown in Figures 4–7, the asterisks tend to lie between the level and envelope psychometric functions.

The results for Obs 4 (Fig. 8), who is less sensitive than the other observers, do not conform to this pattern. Note that compared with Figures 5–7, the abscissa of Figure 8 is shifted upward by 5 dB. The signal levels required for Obs 4 to perform above chance in the tone-in-noise task are sufficiently high that the level and envelope models both predict near-perfect performance (Fig. 2, signal levels above 70–75 dB SPL). Thus, for the maskers used in the current study, the models make the same prediction *vis-à-vis* the effects of adding the signal to A vs. B maskers. For signal levels lower than those shown in Figure 8 (in

Experiments II and III), however, Obs 4's data do not segregate as predicted (see Fig. 2). Rather she performed at near-chance levels regardless of feedback. Obs 4 may not have altered her decision strategy in a feedback-dependent way. By testing maskers with even larger differences in envelope and level properties, effects of feedback might be elicited. Alternatively, Obs 4 may have simply ignored feedback and/or the experimenter's instructions.

Rough categorizations of the observers' data in the tone-in-noise task are as follows. When detecting a tone added to a narrow band of noise, for Obs 1 the psychometric functions fall roughly between the level and envelope detector data (Fig. 5, asterisks about midway between squares and circles). For the tone-in-noise task, Obs 3 is nearer an envelope detector than a level detector (Fig. 7, asterisks tend to be nearer

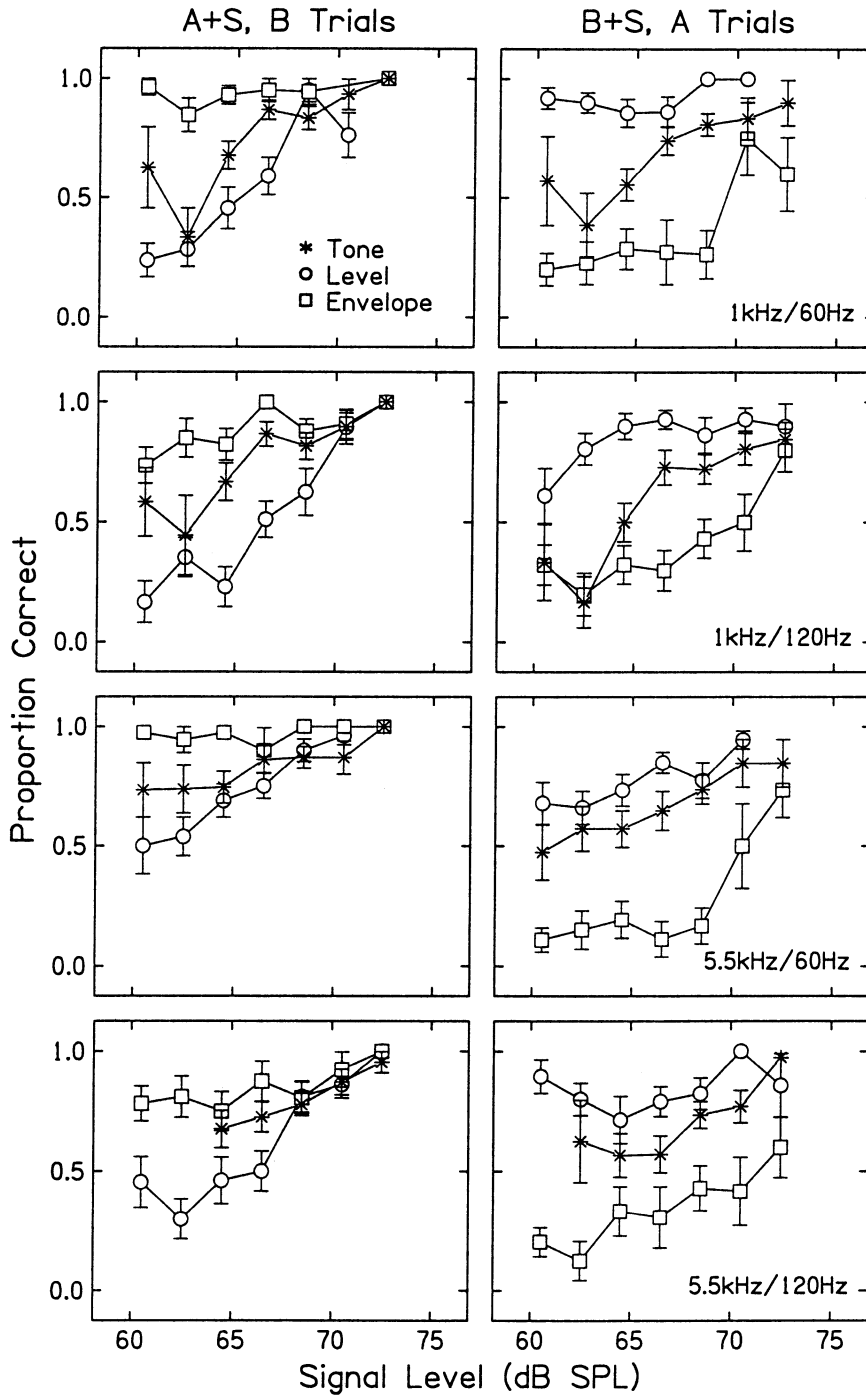


FIG. 5. Psychometric functions for Obs 1 are shown for trials in which the signal is added to the A masker and the nonsignal contains a B masker (left panels) and the opposite (right panels). Data for different conditions are plotted in different rows; signal frequency/masker bandwidth of 1000/60, 1000/120, 5500/60, and 5500 Hz/120 Hz from top to bottom. Asterisks are for Experiment I (tone-present feedback), circles are for Experiment II (level feedback), and squares are for Experiment III (envelope feedback). Error bars indicate an estimate of the standard deviation at each signal level based on Bernoulli trials. Data based on fewer than 6 trials are not plotted.

squares than circles). Observer 2 is intermediate. For the 1000-Hz signal, her data tended toward the predictions of the level detector (Fig. 6, top two rows, asterisks nearer circles), and for the 5500-Hz signal, her data tended toward the predictions of the envelope detector (Fig. 6, bottom two rows, asterisks nearer squares). Possibly due to her lack of sensitivity, Obs 4 defies categorization. For Obs 1–3, the data do not support either the level or the envelope detection model for

the detection of a tone added to noise. For Obs 4 neither model can be rejected.

Table 2 shows thresholds for sensitivity to changes in level (ΔL in dB) and changes in normalized slope. The thresholds are based on fits to psychometric functions generated using the data of Experiments II and III, respectively. The rows show results for different observers and the final two rows indicate the mean and standard error of the mean across observers. For

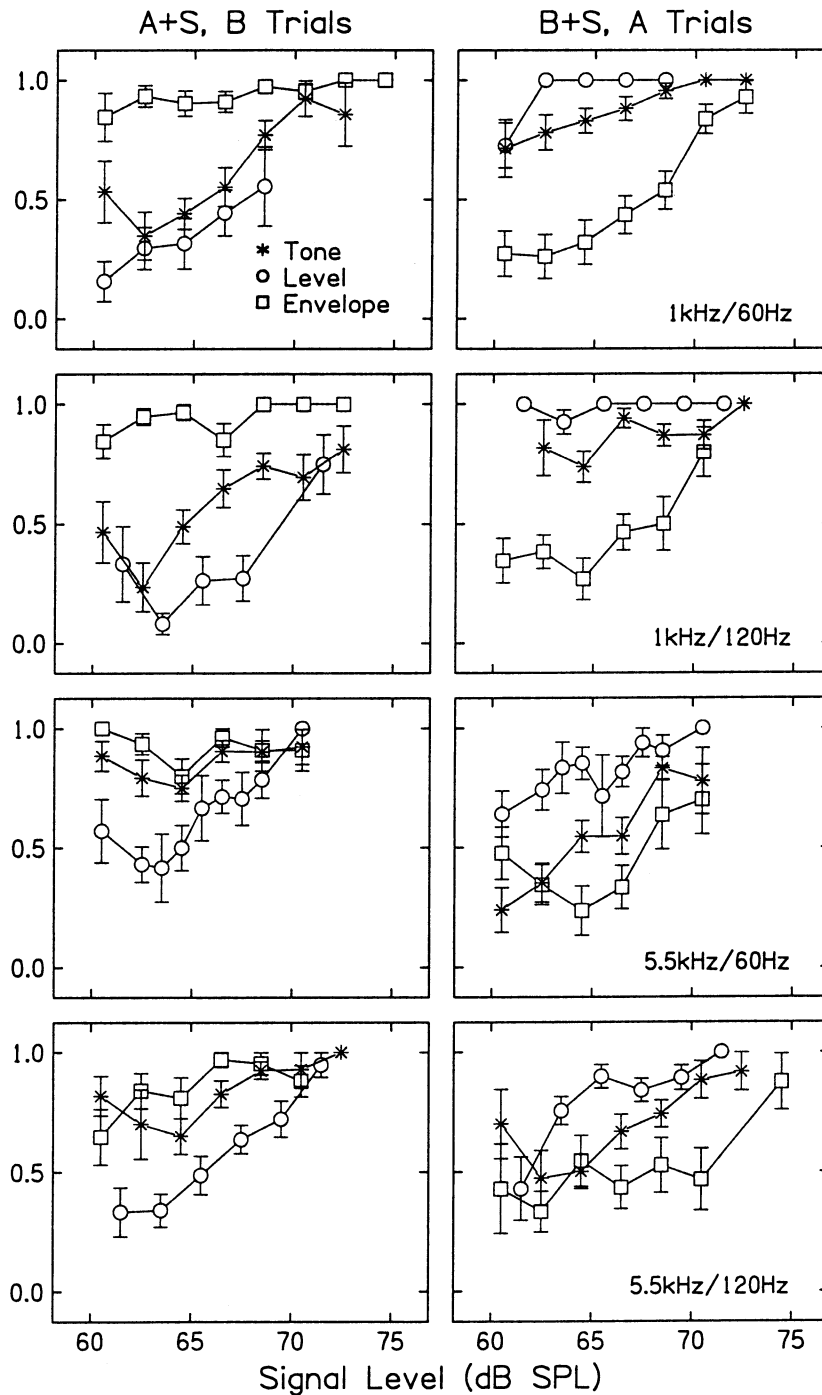


FIG. 6. The data for Obs 2 are plotted using the format of Figure 5.

the 60- and 120-Hz bandwidths, simulations indicate that the signal level associated with the average threshold, ΔL , is approximately 67 dB SPL. For the normalized envelope slope statistic, simulations indicate that the threshold is associated with a signal level of 71 dB SPL. The only notable effect of condition on this estimate is that when a 120-Hz-wide masker masks the 5500-Hz signal, the signal level required for threshold is approximately 0.5 dB higher than for the other conditions.

DISCUSSION AND CONCLUSIONS

For Obs 1–3, the results of Experiments I–III indicate (a) variations in feedback contingencies lead to changes in observers' strategies and provide a method of testing alternative psychophysical models (see also Nachmias 2001); (b) when a tone is added to noise, observers can make discriminations based on either level or envelope cues (see also Bos and deBoer 1966;

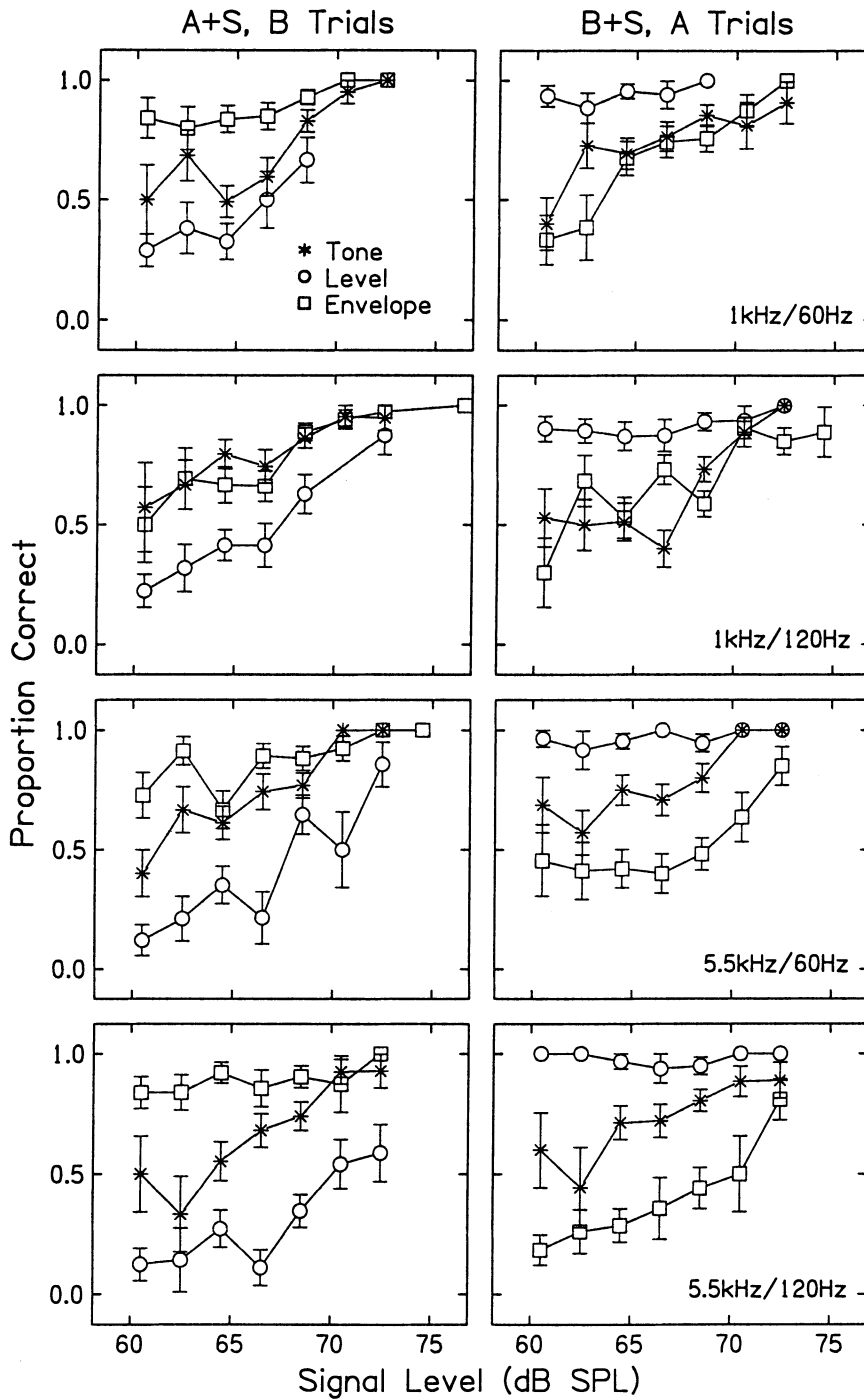


FIG. 7. The data for Obs 3 are plotted using the format of Figure 5.

Richards and Nekrich 1993); (c) sensitivity to differences in level and envelope modulation rivals that for the detection of a tone added to noise (see also Richards and Nekrich 1993); and (d) observers do not use just changes in level nor just changes in modulation depth to detect a tone added to a narrow band of noise. It is important to note that the terms “level” and “envelope” should be taken to mean “level models and other models that provide similar predictions,” etc. The results of Obs 4 do not support the above

conclusions, possibly because the different models tested generated similar predictions for the high signal levels required by Obs 4 for detection.

Except for Obs 2, the psychometric functions were similar regardless of whether the signal and maskers were presented in the 1000- or the 5500-Hz region. Thus, a reliance on fine structure information is not apparent in the current data set. Keeping in mind that changes in fine structure are correlated with changes in envelope pattern and not changes in level, Obs 2’s

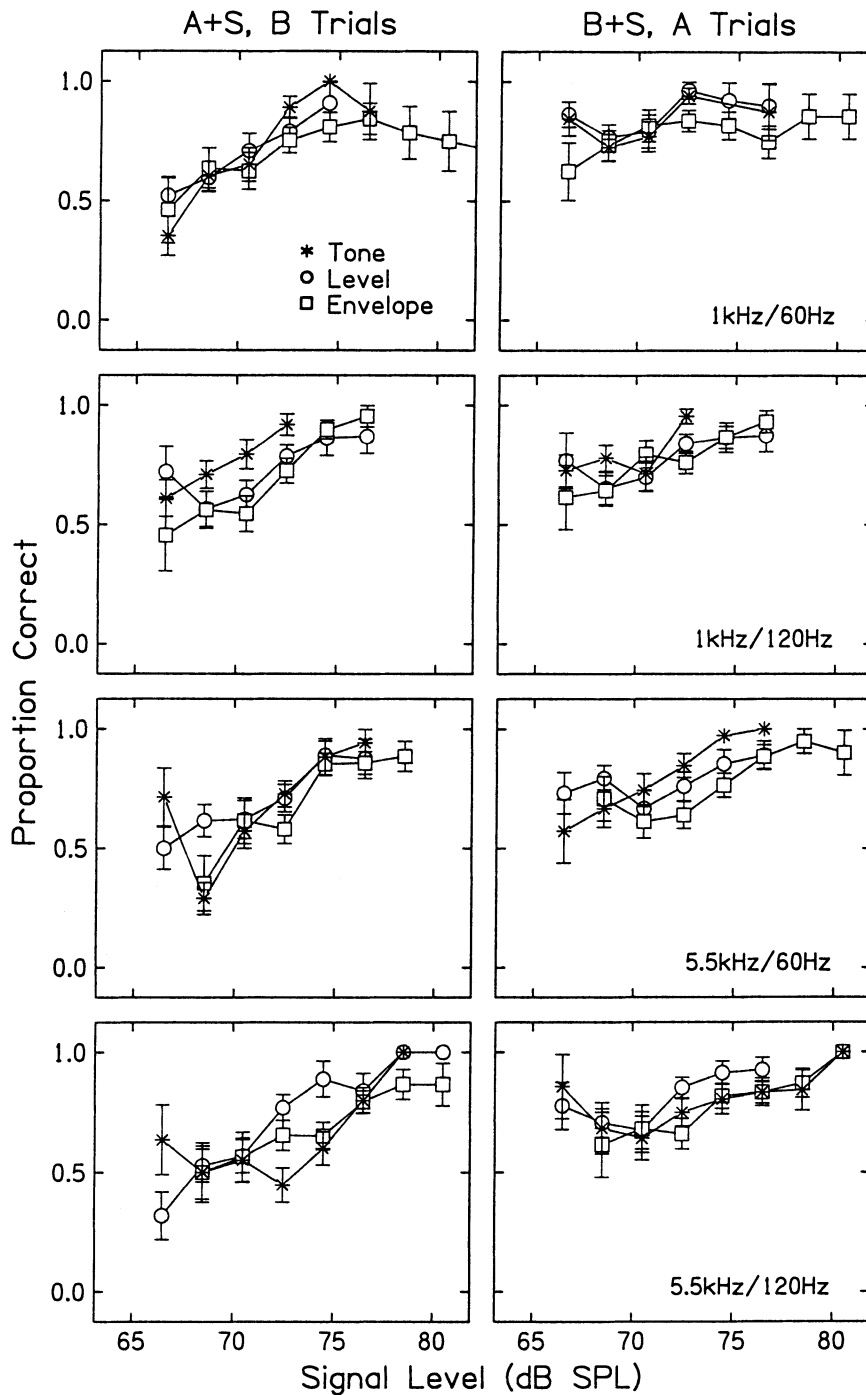


FIG. 8. The data for Obs 4 are plotted using the format of Figure 5. Note that compared with Figures 5–7, the abscissa is translated to higher signal levels.

apparently heavier reliance on envelope patterns at high but not low frequencies is not consistent with a dependence on fine structure at low frequencies.

Overall the data indicate that changes in feedback can unmask potential cues in psychophysical tasks. The data also suggest that the traditional tone-in-noise paradigm lacks constraint. Observers can and do vary their responses depending on feedback contingencies. There exist at least two separate strategies, level- and

envelope-based, that observers can apply to the task. Importantly, both strategies lead to similar thresholds, and thus from the observer's perspective there is little pressure to choose one strategy over another.

This is not to say that one cannot fit a model to account for an observer's behavior. For the current data set there is no reason to doubt that a linear combination of level and envelope cues could be fashioned to account for the data. Rather, for the experiment

TABLE 2

Thresholds for Experiments II and III (LEVEL and ENV) expressed as ΔL in dB and the difference in normalized envelope slope, respectively ^a								
SF BW	Level				Envelope			
	1000 60	1000 120	5500 60	5500 120	1000 60	1000 120	5500 60	5500 120
Obs 1	2.1	1.4	2.6	1.6	0.0007	0.0015	0.0006	0.0011
Obs 2	1.6	0.7	2.1	1.6	0.0010	0.0015	0.0007	0.0011
Obs 3	1.6	1.3	1.3	0.9	0.0009	0.0016	0.0008	0.0008
Obs 4	3.4	3.5	4.8	3.5	0.0018	0.0025	0.0025	0.0031
AVG	2.2	1.7	2.7	1.9	0.0011	0.0018	0.0012	0.0015
SEM	0.4	0.6	0.8	0.5	0.0002	0.0003	0.0004	0.0005

^aThe rows show thresholds for each observer, the averaged thresholds, and standard errors of the means across 4 observers. The different columns are grouped by type of experiment, signal frequency (SF) and masker bandwidth (BW).

reported here, there is little reason to believe that such a model would be meaningful. What drives the observer's choice of strategy? Given that there exist *at least* two possible strategies that lead to similar levels of performance, happenstance might play as large a role as any other factor in determining how observers combine multiple cues present in a tone-in-noise task. It appears that for psychophysical tasks in which multiple cues and/or perceptual dimensions contribute to observers' behavioral responses, substantial complexity in experimental design and/or models may be necessary in order to explain the cue combinations adopted by observers.

ACKNOWLEDGMENTS

This work was supported by grant RO1 DC 02012 from the National Institutes of Health. Drs. Jennifer J. Lentz, Brian C.J. Moore, Jacob Nachimias, and Frederic L. Wightman provided very helpful suggestions on an earlier draft of this manuscript. Keith Diamond, Linda To, and Tao Zeng provided assistance in data collection and analysis.

REFERENCES

- ABRAMOWITZ M, STEGUN IA. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover New York, 1964.
- AHUMADA JR A, LOVELL J. Stimulus features in signal detection. J. Acoust. Soc. Am. 49:1751–1756, 1971.
- BOS CE, de BOER E. Masking and discrimination. J. Acoust. Soc. Am. 39:708–715, 1966.
- DAI H. On measuring psychometric functions: A comparison of the constant-stimulus and adaptive up–down methods. J. Acoust. Soc. Am. 98:3135–3139, 1995.
- GREEN DM, SWETS JA. Signal Detection Theory and Psychophysics. Krieger New York, 1966.
- HARTMANN WM, PUMPLIN J. Noise power fluctuations and the masking of sine signals. J. Acoust. Soc. Am. 83:2277–2289, 1988.
- LEVITT H. Transformed up–down methods in psychoacoustics. J. Acoust. Soc. Am. 49:167–177, 1971.
- MOORE BCJ. Mechanisms of masking. J. Acoust. Soc. Am. 57:391–399, 1975.
- NACHIMAS J. Contrast discrimination with and without spatial uncertainty. Vision Res. (in press).
- RICHARDS VM. The detectability of a tone added to narrow bands of equal-energy noise. J. Acoust. Soc. Am. 91:3424–3435, 1992.
- RICHARDS VM, NEKRICH RD. The incorporation of level and level-invariant cues for the detection of a tone added to noise. J. Acoust. Soc. Am. 94:2560–2574, 1993.
- VIEMEISTER NF. Temporal modulation transfer functions based upon modulation thresholds. J. Acoust. Soc. Am. 66:1364–1380, 1979.