

# Using magnitude estimation to investigate the perceptual components of signal detection theory

DALE J. COHEN and LEN LECCI

*University of North Carolina, Wilmington, North Carolina*

Several experiments suggest that the relation between detection, signal energy, and perceived tone intensity is very different from the relation between detection, signal energy, and perceived tone duration. We propose a new task, the magnitude estimation and detection (MED) task, that allows one to assess the relation between the psychological dimensions of a stimulus and detection. In Experiment 1, we used the MED task to assess the relation between perceived tone intensity and detection in a yes/no task. The data show a strong relation between the two. In Experiment 2, we used the MED task to assess the relation between perceived tone duration and detection in a yes/no task. The data show a relatively weak relation between the two. Our data suggest that tone intensity is a less perceptually noisy dimension than tone duration. We present the advantages and disadvantages of the MED task with the hope that this task can aid researchers in better understanding the perceptual and decisional mechanisms underlying various cognitive processes.

In a yes/no task, subjects are asked to detect the presence or absence of a specific stimulus. Because all stimuli are multidimensional, the subjects may use their phenomenological impression of any or all dimensions when making their judgments. Signal detection theory (SDT) addresses the issue of multiple stimulus qualities influencing the detection process by interpreting the decision axis as values of the likelihood ratio,  $\lambda(A_i) = f_{SN}(A_i)/f_N(A_i)$  (i.e., the likelihood that point  $A_i$  came from signal + noise distribution [SN] relative to noise distribution [N]; Swets, Tanner, & Birdsall, 1964). Although this may accurately characterize a subject's decision process, it sidesteps the issue of how particular stimulus qualities combine to produce  $A_i$ .

$A_i$  likely involves both perceptual and cognitive factors. The perceptual factors ( $P_i$ ) include the observer's phenomenological experience of multiple qualities of the stimulus (e.g., tone frequency, duration, clarity, intensity, etc.). The cognitive factors ( $C_i$ ) include the observer's calculation of the relative importance of each perceptual quality of the stimulus to the detection of the target.<sup>1</sup> Therefore, each point on the decision axis for a particular trial represents the sum of all the weighted perceptual qualities of the stimulus:

$$A_i = \sum_{i=1}^N C_i P_i. \quad (1)$$

Although the composition of  $A_i$  is unimportant if one only wishes to get a measure of the observer's sensitivity

to a specific target, it becomes important when one attempts to explain detection. That is, signal detection researchers often explain detection as a function of the signal energy of the target. Signal energy, however, is a physical quality of the stimulus and subjects must make their responses on the basis of the perceptual qualities of the stimulus. When discussing auditory detection, signal energy translates into the physical dimensions of sound pressure over time. Strictly speaking, subjects could be basing their judgments on perceived tone intensity (i.e., loudness), perceived tone duration, or a weighted combination of the two. Several experiments suggest that the relation between detection, signal energy, and perceived tone intensity is very different from the relation between detection, signal energy, and perceived tone duration. This implies that subjects treat the two pieces of perceptual information differently when judging the presence or absence of a target.

The relation between detection, signal energy, and perceived tone intensity is likely straightforward. In a signal detection experiment, Green, Birdsall, and Tanner (1964) presented subjects with signals that varied in signal energy but were of a constant duration (by adjusting the signal to noise ratio,  $E/N_0$ ). They found that detectability was positively correlated with signal energy. In this experiment, signal energy translates into sound pressure alone because time was held constant. This suggests that there is a strong relation between auditory target detection and perceived tone intensity. Interestingly, Durlach and Braida (1969) proposed that intensity resolution could be described in terms of a signal/noise model similar to that of SDT. Although they did not directly assess the relation between perceived intensity and detection, their data suggest that a strong relation exists (Braida & Durlach, 1970).

The relation between detection, signal energy, and perceived tone duration is less straightforward. Creelman

---

The authors thank S. Batson, J. Cohen, and J. Skipper for their helpful comments and support at various stages of this project. Correspondence should be addressed to D. J. Cohen, Department of Psychology, University of North Carolina, Wilmington, NC 28403-3297 (e-mail: cohend@uncwil.edu).

(1964) required subjects to discriminate between two tones of different durations. He showed that when the auditory stimuli were easily heard, signal energy was not correlated with discrimination. Furthermore, Green et al. (1964) showed that when a signal varied in duration but was of constant signal energy (this was achieved by adjusting both duration and  $E/N_0$  such that as duration increased, total signal energy remained constant), detectability remained constant within a large range of durations. To further complicate matters, Dai and Wright (1995, 1999) present evidence that the listeners' expectations of signal duration affect detectability of a tone. Specifically, listeners detect tones of expected duration better than tones of unexpected duration. Finally, in those cases in which tone duration does influence detection (durations less than 200 msec), longer tones are perceived as more intense (Buus, Florentine, & Poulsen, 1999; Munson, 1947). In this case, it appears that increased detection is a result of perceived intensity rather than perceived duration. Together, these data suggest that perceived tone duration at best weakly correlates with signal energy and, therefore, is likely to be only weakly correlated with tone detection.

One cannot use a traditional signal detection task to assess the relation between tone detection and perceived tone intensity and duration because SDT provides the experimenter no access to the perceptual and cognitive components that make up  $A_i$ . To directly assess the relation between tone detection and perceived tone intensity and duration, we developed a new task, the *magnitude estimation and detection* (MED) task.

### Magnitude Estimation and Detection Task

To assess the relation between the perceptual qualities of a stimulus and target detection, we developed the MED task, which links signal detection and magnitude estimation. The MED task is essentially a yes/no experiment in which we record both the subject's magnitude estimates (MEs) of the stimulus quality of interest (either tone intensity or duration) and their judgments of target presence. We then calculate two measures of sensitivity: one based on the subject's MEs, and one based on the subject's judgment data. Finally, we assess the strength of the relation between these two measures of sensitivity. The strength of this relation should provide some insight about the composition of  $A_i$ .

In the present study, we used the traditional  $d'$  as the measure of sensitivity derived from the subject's judgments ( $d'_{SDT}$ ). There are, however, several possible methods for deriving a measure of sensitivity from the subject's MEs. These measures are based on the theoretical similarity between the underlying distributions derived from SDT and those obtained from the ME data. Specifically, on every trial in a yes/no task, SDT assumes that the subject should experience some magnitude of any specified stimulus quality—even on trials in which no target is present. Therefore, the subject's MEs on the target-absent trials of a yes/no task should specify the N distribution of the stimulus quality being assessed. Sim-

ilarly, the subject's MEs on the target-present trials of a yes/no task should specify the SN distribution of the stimulus quality being assessed. From these distributions, one can derive several measures of sensitivity solely on the basis of the subject's perception of the stimulus quality in question.

One measure of sensitivity that can be derived from the MEs is the standardized distance between the N and SN distributions:

$$d_{ME} = \frac{(\mu_{SN} - \mu_N)}{\sigma}, \quad (2)$$

where  $\mu_{SN}$  is the subject's mean ME for the target-present trials,  $\mu_N$  is the subject's mean ME for the target-absent trials, and  $\sigma$  is the average standard deviation of the target-absent and target-present trials:  $\sigma = (\sigma_{SN} + \sigma_N)/2$  (generalized from Durlach & Braida, 1969). This measure,  $d_{ME}$ , is a direct measure of the standardized distance between the means of the N and SN distributions. Theoretically,  $d_{ME}$  should be equal to  $d'_{SDT}$  when the stimulus quality measured is the only quality the subject uses to make his or her judgment of the target's presence. However, because  $d_{ME}$  has no practical boundaries and  $d'_{SDT}$  has a practical boundary of about 5.0,<sup>2</sup> one expects a non-linear relation between the two measures, at least at the higher values.

One can also calculate a robust measure of sensitivity that is invariant across transformations by bootstrapping a  $d'$  from the ME data ( $d'_{ME}$ ).<sup>3</sup> Specifically, one can randomly draw, with replacement, one ME from the target-present data or the target-absent data. If the value of this ME is above a set criterion, we assume that the subject said that the target was present. If the value of this ME is below the set criterion, we assume that the subject said that the target was absent. If the value of this ME is equal to the set criterion, we assume the subject guessed, and we assign a .5 probability that the subject responded that the target was present. If the ME was drawn from the SN distribution, and the subject responded that the target was present, then we classify the response as a hit. If the ME was drawn from the N distribution, and the subject responded that the target was present, then we classify the response as a false alarm. This can be repeated for the number of trials (e.g., 200), and  $d'_{ME}$  can be calculated from the hit and false alarm rates. For each subject, this whole cycle can be repeated a large number of times (e.g., 1,000) with the mean  $d'_{ME}$  of the (1,000) cycles representing the sensitivity estimate and the standard deviation of the (1,000) cycles representing the standard error around the  $d'_{ME}$  (Efron, 1982). For each cycle of the bootstrap procedure, we randomly set the criterion between the mean of the N distribution and the mean of the SN distribution of the subject's ME data. Because  $d'_{ME}$  is estimated using the same procedure as the  $d'_{SDT}$ , it should have the same properties as the  $d'_{SDT}$  (e.g., practical boundary of about 5.0). Thus, one expects a linear relation between  $d'_{ME}$  and  $d'_{SDT}$  when the stimulus quality measured is the only quality the subject uses to make his or her judgment of the

target's presence. Although other measures of sensitivity can be derived from the ME data,<sup>4</sup> we will concentrate on these two measures.

In summary, to directly assess how subjects' perceptions of tone intensity and tone duration are correlated with target detection, we set out to (1) collect subjects' MEs of perceived tone intensity and tone duration in two yes/no experiments, (2) collect subjects' judgments of target presence in the yes/no experiments, and (3) assess the strength of the relation between the measures of sensitivity on the basis of the MEs and the  $d'$  in a traditional signal detection task ( $d'_{SDT}$ ).

## EXPERIMENT 1

Experiment 1 was performed to assess the relation between subjects' perceptions of tone intensity and target detection.

### Method

**Subjects.** Forty-one subjects from the general psychology subject pool volunteered to participate in exchange for class credit. All subjects reported normal hearing.

**Apparatus and Stimuli.** All stimuli were presented through a Calrad 15-135B stereo headphone controlled by an 80486 microcomputer using the Windows operating system.

There were six stimuli presented to the subject: the target tone, the target-absent stimulus, and four target-present stimuli. All stimuli were sampled at 22,050 times/second and stored in 16-bit format. The target tone was presented to the subjects only in the instructions and during their rest breaks. The target tone was a 4440-Hz tone in a gaussian envelope that lasted 2 sec. The peak amplitude of the target tone was 41.5 dB SPL and occurred 800 msec into the target stimulus. The target-absent stimulus consisted of 4 sec of generated white noise at 57 dB SPL. The four target-present stimuli were created by adding the target tone to the target-absent stimulus in four equally spaced points along the 4-sec stimulus (beginning at 500 msec, 1,000 msec, 1,500 msec, and 2,000 msec). The four target-present stimuli were created to discourage the subjects from anticipating when the target tone would appear in the noise.

The target-present and target-absent stimuli were chosen on the basis of the results of a pilot experiment run on a separate set of subjects. In the pilot experiment, we systematically adjusted the amplitude of the target tone until most subjects could detect the target-present stimulus with a  $d' = 1.5$  in 200 trials. The resulting stimuli were used in the present experiment with the expectation that there would be variance in the subjects' sensitivities to the stimuli, but the stimuli would be neither below nor above threshold for the majority of the subjects tested.

**Procedure.** Each subject was tested individually in a  $3 \times 6$  foot room that contained a chair, a desk, and a computer. The experiment was fully automated and was presented on the computer. The subject sat in front of the computer, wearing stereo headphones. The experimenter initiated the experiment and left the room. The computer presented the instructions both visually and auditorily (see the Appendix). The instructions explained that, on each trial, the computer would present a stimulus consisting of white noise that may or may not contain a 4440-Hz tone. The subject was to judge whether the 4440-Hz tone was present. At this point in the instructions, the target tone alone was played for the subject, followed by the target-absent stimulus, followed by a randomly selected target-present stimulus. The yes/no response screen was then shown and explained to the subject. The subject made his or her yes/no response on a screen that contained two buttons, one labeled "Signal Present" and the other "Signal Absent." The subject was to indicate whether the tar-

get was present or absent using a computer mouse to "click" on one of two buttons. Finally, the ME response screen was shown and explained to the subject (see the Appendix).

The ME response screen contained 45 thin buttons arranged horizontally across the screen.<sup>5</sup> The subject indicated the relative magnitude of the perceived tone intensity by clicking on the appropriate button. The ME "standard" in this task was the subject's perceived intensity of the target tone without the background noise (to which the subject was allowed to familiarize him/herself prior to the experiment). The 12th button was labeled "Original Intensity" and was used to represent the standard intensity.<sup>6</sup> If the subject's perceived intensity of the tone was less than the standard, the subject was to press the appropriate button to the left of the 12th button. For example, if the subject perceived the tone as half as intense as the standard, then he/she was to press the 6th button. If the subject heard absolutely no tone, he/she was to press the 1st button, which was labeled "0 Intensity." If the subject's perceived intensity of the tone was greater than the standard, the subject was to press the appropriate button to the right of the 12th button. The subject was instructed to respond solely on the basis of his/her perceived intensity of the tone. Furthermore, the subject was not required to use the entire scale.

After the instructions, the subject was given the opportunity to hear the instructions again (by pressing the appropriate button). The subject was also given the opportunity to hear any or all of the three stimuli described above (by pressing the appropriate button). Finally, when the subject was ready to begin the experiment, he/she pressed a button labeled "Begin Experiment."

The experiment consisted of 200 trials (100 target-present and 100 target-absent trials) and lasted about 1 h. Each trial began with the random presentation of either the target-present stimulus or the target-absent stimulus. For each target-present trial, the target-present stimulus was randomly selected from the four target-present stimuli described above. After the stimulus was presented, the yes/no response screen (described above) was presented to the subject. Immediately following the subject's yes/no response, the ME response screen (described above) was presented to the subject. One second after the subject responded to the ME response screen, the next trial began. After every 25 trials, the subject received a self-timed rest break. During the break, the subject had the opportunity to refamiliarize him/herself with the target tone, a randomly selected target-present stimulus, and the target-absent stimulus. To hear a particular stimulus, the subject pressed a button that corresponded to that stimulus and was allowed to do so as often as he/she liked during the break. When the subject was ready to begin the experiment again, he/she pressed a fourth button labeled "Begin Experiment."

### Results and Discussion

For each subject, we calculated  $d'_{SDT}$  from his or her yes/no data and  $d_{ME}$  from his or her ME data (as described in the introduction). We also bootstrapped  $d'_{ME}$  from each subject's ME data as described in the introduction. When calculating the  $d'_{SDT}$ , hit rates and false alarm rates equal to one were converted to .999 ( $N = 7$ ;  $d'_{SDT}$  values of these subjects ranged from 0 to 5.55). There were no hit or false alarm rates equal to zero.

To determine the relation between  $d'_{SDT}$  and  $d_{ME}$ , a regression analysis, including a quadratic term, was calculated. The following significant function was found:  $d'_{SDT} = 1.06 * d_{ME} - 0.06 * d_{ME}^2 + 0.30$  [ $F(2,38) = 125.3$ ,  $p < .001$ ,  $r^2 = .87$ ]. Both the linear term ( $t = 8.62$ ) and the quadratic term ( $t = -3.22$ ) were significant, but the intercept did not differ significantly from zero ( $t = 1.6$ ). When the same regression was run, while forcing a zero intercept, the following significant function was

found:  $d'_{\text{SDT}} = 1.23 * d_{\text{ME}} - 0.07 * d_{\text{ME}}^2$  [ $F(2,39) = 119.7, p < .001, r^2 = .86$ ]. Again, both the linear term ( $t = 20.1$ ) and the quadratic term ( $t = -5.9$ ) were significant (see Figure 1). Further analyses (see notes 7 and 8) determined that the quadratic term was not an artifact of the high  $d_{\text{ME}}$  values,<sup>7</sup> nor was it because  $d_{\text{ME}}$  and  $d'_{\text{SDT}}$  calculated the variance of the SN and N distributions differently.<sup>8</sup> Because no predictive power was lost when eliminating the intercept, the results will be discussed in terms of the latter function.<sup>9</sup>

The above function falls very close to the identity line (which indicates where the points are located if  $d'_{\text{SDT}} = d_{\text{ME}}$ ; i.e., slope = 1, and intercept = 0) for  $d' < 3.75$ . When  $d' > 3.75$ , the function falls below and to the right of the identity line. Assuming that the subject faithfully performed the yes/no task, this pattern can mathematically occur only if the subject's perception of the particular stimulus quality better predicts the target's presence than the decision made by the subject in his or her yes/no responses. This is likely the case at the high end because  $d'_{\text{SDT}}$  has a practical boundary, whereas  $d_{\text{ME}}$  does not.

It is important to note that if a subject's responses on the ME task are not predictive of the subject's  $d'_{\text{SDT}}$ , the data would fall above and to the left of the identity line. In this instance, we assume that  $d_{\text{ME}}$  is less responsive to the subject's sensitivity to the target than  $d'_{\text{SDT}}$ . Because the less responsive measure should always (except by chance) indicate a lower sensitivity than the more responsive measure, the more responsive measure provides an upper boundary for target sensitivity. Therefore,  $d_{\text{ME}}$  should be less than  $d'_{\text{SDT}}$ . Because this is a restricted, triangular-shaped area, a significant relation would still occur between  $d'_{\text{SDT}}$  and  $d_{\text{ME}}$ . To determine the strength

of the relation that would result from such an instance, we ran a Monte Carlo simulation in which pairs of data points were generated ( $d_{\text{ME}}, d'_{\text{SDT}}$ ). In this simulation, 60 pairs of random data were generated such that  $d'_{\text{SDT}} < 4.0$ , and  $d_{\text{ME}} < d'_{\text{SDT}}$ . A linear regression was then fitted to the data points. Five hundred runs yielded an average  $r^2 = .41$  ( $SE = .09$ ). The same simulation shows that a quadratic function fitted to 60 data points will yield an average  $r^2 = .44$  ( $SE = .1$ ). It is therefore necessary to consider the increment in the  $r^2$  over that of the Monte Carlo simulation when the data points fall within this region. To assess whether our data accounted for significantly more variance than that of the Monte Carlo simulation of semirandom data described above, we calculated a Fisher's  $z$  transformation (see J. Cohen & P. Cohen, 1983). The results show that the above equation accounted for significantly more variance than that of the Monte Carlo simulated data ( $z = 3.77, p < .001$ ).

For each subject,  $d'_{\text{ME}}$  was bootstrapped from the MEs. A regression was computed for the following equation:  $d'_{\text{SDT}} = f(d'_{\text{ME}})$ . The following significant function was found:  $d'_{\text{SDT}} = 1.0 * d'_{\text{ME}} + 0.10$  [ $F(1,38) = 264.5, p < .001, r^2 = .87$ ]. The intercept ( $t = 0.62$ ) was not significant, whereas the linear term ( $t = 16.3$ ) was significant (see Figure 2). This relation is nearly equivalent to the identity line. Using Fisher's  $z$  transformation we were again able to determine that the above equation accounted for significantly more variance than that of the Monte Carlo simulated semirandom data ( $z = 3.92, p < .001$ ).

In sum, the data indicate that the two measures of sensitivity derived from the MEs ( $d_{\text{ME}}$  and  $d'_{\text{ME}}$ ) are very consistent with the  $d'_{\text{SDT}}$ . Importantly, the  $d'_{\text{ME}}$  derived from estimates of tone intensity provides information almost

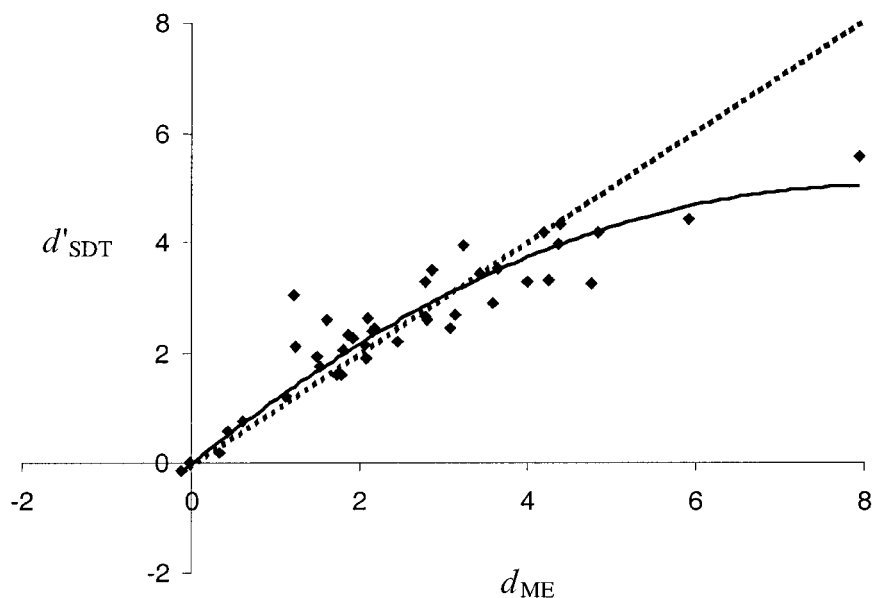


Figure 1. The  $d_{\text{ME}}/d'_{\text{SDT}}$  data pairs from Experiment 1. Each datum represents a subject's data. The pattern indicates that  $d_{\text{ME}}$  is highly related to  $d'_{\text{SDT}}$ .

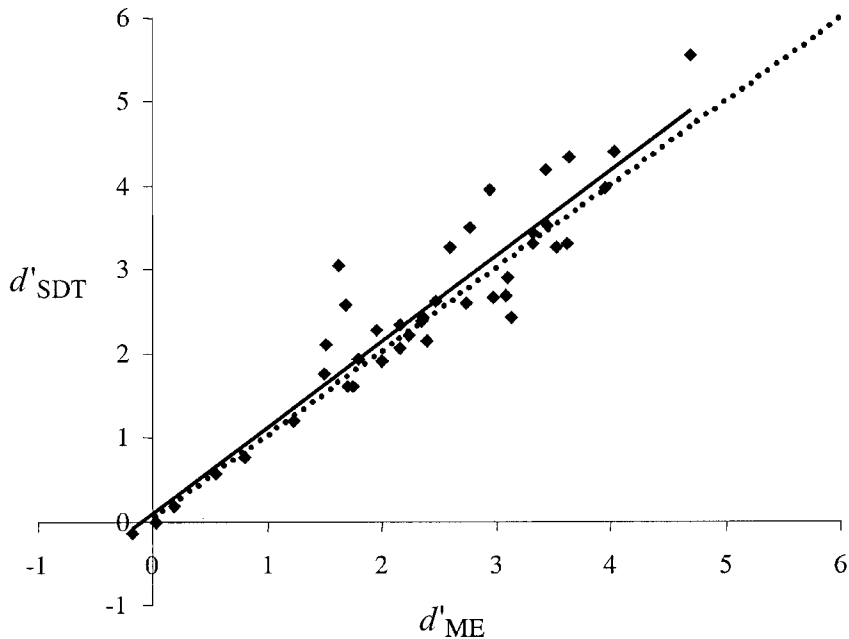


Figure 2. The  $d'_{ME}/d'_{SDT}$  data pairs from Experiment 1. Each datum represents a subject's data. The pattern indicates that  $d'_{ME}$  is highly related to  $d'_{SDT}$ .

identical to that provided by the  $d'_{SDT}$ . The strength of the relation between the measures suggests that (1) a subjective measure can provide an accurate estimate of subject sensitivity and (2) auditory detection in a yes/no task is in large part a function of fluctuations in perceived stimulus intensity. These conclusions will be discussed in the General Discussion section.

## EXPERIMENT 2

Experiment 2 was performed to assess the relation between subjects' perceptions of tone duration and target detection.

### Method

**Subjects.** Forty-one subjects from the general psychology subject pool volunteered to participate in exchange for class credit. All subjects reported normal hearing. No subject participated in Experiment 1.

**Apparatus and Stimuli.** The stimuli and apparatus were identical to those of Experiment 1.

**Procedure.** The procedure was identical to that of Experiment 1, with two exceptions. First, in the automated instructions, the word "Duration" replaced "Intensity" throughout (with one appropriate correction in sentence structure). Second, the word "Duration" replaced "Intensity" in the ME response screen.

### Results and Discussion

For each subject, we calculated  $d'_{SDT}$  from his or her yes/no data and calculated  $d_{ME}$  and  $d'_{ME}$  from his or her ME data (as in Experiment 1). When calculating the  $d'_{SDT}$ , hit rates and false alarm rates equal to one were converted to .999 ( $N = 5$ ;  $d'_{SDT}$  values of these subjects ranged

from 0 to 4.97). There were no hit or false alarm rates equal to zero.

To determine the relation between  $d'_{SDT}$  and  $d_{ME}$ , a regression analysis, including a quadratic term, was calculated. The following significant function was found:  $d'_{SDT} = 1.69 * d_{ME} - 0.20 * d_{ME}^2 + 0.36$  [ $F(2,39) = 49.1$ ,  $p < .001$ ,  $r^2 = .70$ ]. Both the linear term ( $t = 6.69$ ) and the quadratic term ( $t = -3.47$ ) were significant, but the intercept did not differ significantly from zero ( $t = 1.51$ ). When the same regression was run, while forcing a zero intercept, the following significant function was found:  $d'_{SDT} = 2.0 * d_{ME} - 0.26 * d_{ME}^2$  [ $F(2,40) = 46.5$ ,  $p < .001$ ,  $r^2 = .67$ ]. Again, both the linear term ( $t = 12.9$ ) and the quadratic term ( $t = -5.66$ ) were significant (see Figure 3). Because no predictive power was lost when eliminating the intercept, the results will be discussed in terms of the latter function.

The quadratic function falls far from the identity line for  $d' < 4.0$ , and it returns to the line only because of two points at the far end of the distribution. Figure 3 shows that the data are fairly randomly dispersed above and to the left of the identity line. Recall that even random data will produce a significant relation in this region. Using Fisher's  $z$  transformation, we were able to determine that the above equation did not account for significantly more variance than that of the Monte Carlo simulated semi-random data ( $z = 1.80$ , n.s.). In addition, the above equation accounted for significantly less variance than the equation derived from the tone intensity data ( $z = 1.97$ ,  $p < .05$ ). These data indicate that the MEs of stimulus duration do not provide a consistent measure of the subject's sensitivity to the target stimulus.

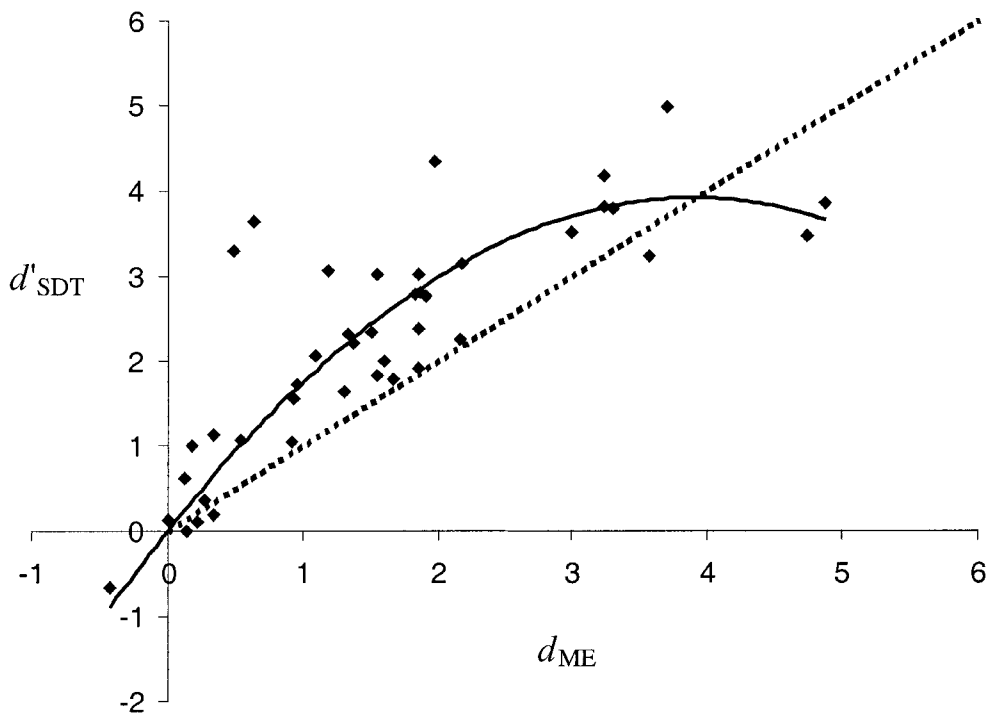


Figure 3. The  $d_{ME}/d'_{SDT}$  data pairs from Experiment 2. Each datum represents a subject's data. The pattern indicates that  $d_{ME}$  is weakly related to  $d'_{SDT}$ .

When the  $d'_{ME}$  is bootstrapped from the ME data (as in Experiment 1), the data are fairly randomly dispersed in the upper left-hand section of the graph. This again indicates that  $d'_{ME}$  is a poor indicator of the target's presence. A regression was calculated for the following equation:  $d'_{SDT} = f(d'_{ME})$ . The following significant function was found:  $d'_{SDT} = 0.88 * d'_{ME} + 0.63$  [ $F(1,39) = 67.0$ ,  $p < .001$ ,  $r^2 = .63$ ]. Both the intercept ( $t = 2.68$ ) and the linear term ( $t = 8.19$ ) were significant (see Figure 4). Using Fisher's  $z$  transformation, we were able to determine that the above equation accounted for significantly less variance than the equation derived from the tone intensity data ( $z = 2.57$ ,  $p = .01$ ) and did not account for significantly more variance than that of the Monte Carlo simulated semirandom data ( $z = 1.37$ , n.s.). These data further indicate that the MEs of stimulus duration do not provide a consistent measure of the subject's sensitivity to the target stimulus. Thus, the amount of useful information about the presence of the target inherent in tone duration is far less than that in tone intensity.

## GENERAL DISCUSSION

We conducted two experiments to determine the relation between the psychological dimensions of tone intensity and duration and an observer's ability to detect an auditory target in a yes/no paradigm. Traditional signal detection experiments have described how detection is related to the physical dimension of signal energy. Unfortunately, traditional signal detection experiments cannot

determine how the psychological dimensions of the stimulus relate to detection. We designed our MED task for this purpose.

Our data show a strong relation between the measures of sensitivity based on the subjects' MEs ( $d_{ME}$  and  $d'_{ME}$ ) and the  $d'_{SDT}$  when the subjects were asked to estimate the magnitude of tone intensity. Conversely, our data show a relatively weak relation between these measures when the subjects were asked to estimate the magnitude of tone duration. Importantly, for each stimulus quality (i.e., intensity and duration), the strength of the relation between the two measures of sensitivity was consistent with the implications of previous findings (Buus et al., 1999; Creelman, 1964; Green et al., 1964; Munson, 1947). Together, these findings suggest that the both  $d_{ME}$  and  $d'_{ME}$  accurately reflect the subject's sensitivity to a target stimulus, along a specific stimulus dimension (e.g., tone intensity).

The results of Experiment 1 suggest that, when judging the presence of an auditory target, subjects favor a pure measure of perceived stimulus intensity over an aggregate of multiple perceptual dimensions. If one assumes that the perceptual noise from each dimension of the stimulus (e.g., frequency, duration, clarity, etc.) is equal, random, and independent from other dimensions, then the optimal detection strategy would be to sum the perceptual magnitudes of all stimulus dimensions. If any or all of these assumptions are incorrect, then a narrower approach may be optimal. For example, if the perceptual noise from stimulus dimensions other than intensity were greater than that for intensity, then the optimal strategy

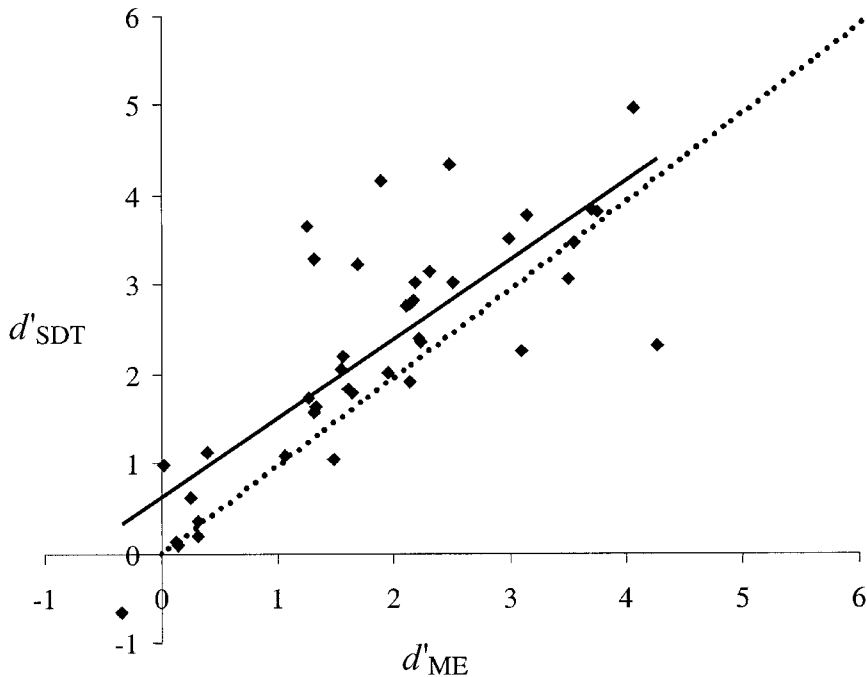


Figure 4. The  $d'_{ME}/d'_{SDT}$  data pairs from Experiment 2. Each datum represents a subject's data. The pattern indicates that  $d'_{ME}$  is weakly related to  $d'_{SDT}$ .

would be to weigh perceptual intensity higher than the remaining dimensions.

The data from the two experiments suggest that the perceptual noise associated with each perceptual dimension is not equal. Our subjects' phenomenological impressions of stimulus duration were only weakly related to their sensitivity to the target. Because tone duration and tone intensity theoretically carry equivalent information about the presence of the target, the usefulness of each perceptual dimension to identify the presence of the target is likely a function of the perceptual noise within each dimension. Although standard SDT techniques cannot directly assess the perceptual noise associated with each dimension of the stimulus, our ME data can. The average standard deviation of our subjects' signal and noise distributions of MEs for tone duration ( $M = 6.44$ ,  $SD = 2.98$ ) was significantly greater than that for tone intensity ( $M = 4.06$ ,  $SD = 1.92$ ) [ $t(164) = 6.07$ ,  $p < .001$ ]. This indicates that our subjects' perceptions of tone duration were less stable than their perceptions of tone intensity.

The present experiments demonstrate the usefulness of the MED task when assessing the relation between subjective impressions of stimulus qualities in a yes/no task and detection of the stimulus. Furthermore, the experiments demonstrate the reliability of the ME responses and make a strong case for their validity as well. Thus, one may cautiously conclude that both  $d_{ME}$  and  $d'_{ME}$  provide good measures of detectability of a stimulus along the particular dimension assessed (i.e., as if subjects *only* perceived the dimension assessed). Specifically, the data

suggest that  $d'_{ME}$  has the same numerical properties as  $d'_{SDT}$ , whereas  $d_{ME}$  may be more responsive than  $d'_{SDT}$  when subjects are highly sensitive to a target.

The MED task could be used to answer many questions that standard signal detection techniques cannot. For example, Green et al. (1964) showed that when a signal varied in duration but not in signal energy, there was a large range of durations where detectability remained constant. This finding raises some interesting questions. For instance, do the observers perceive all of the stimuli as the same intensity (as may be predicted by energy summing theories), or do they perceive the longer stimuli as less intense but hear the tone more consistently (as may be predicted by multiple look theories)? The MED task can answer these questions by assessing perceived intensity across different tone durations.

More generally, when measuring the perceptual space of a stimulus quality using the subject's ME responses, the axis on which the distributions fall is fully specified. This allows one to compare distribution shifts within and across conditions. Such comparisons can be useful in answering important theoretical questions. For example, there is currently a debate in the literature about whether reports of false memories are the result of a shift of one's criterion (Miller & Wolford, 1999) or a shift in the SN distribution (Wixted & Stretch, 2000). Standard signal detection techniques cannot resolve this issue because they do not provide direct access to the underlying distributions. However, because one may obtain a direct measure of memory strength using the MED task, shifts in the

subject's criterion and/or his or her SN distribution can be observed. This is just one example of many in which defining the decision axis may be useful.

The assessment of the distributions underlying SDT can also be useful in adjudicating between theories in which the shape of the underlying distributions is central. For example, recent data suggest that color information and form information are not processed independently, as was once thought (D. J. Cohen, 1997; Kubovy, D. J. Cohen, & Hollier, 1999). Unfortunately, when assessing this issue, one must make critical assumptions about the shape of the underlying distributions (Thomas, 1999). If these assumptions are not valid, then the conclusions drawn from the data may be incorrect. Because the MED task may provide direct access to the underlying distributions for individual perceptual qualities, the MED task may prove useful in assessing the validity of these assumptions.

There are also interesting issues associated with the task that need to be verified. For example, on the surface, the ME response in the MED task seems similar to a confidence rating response in a standard yes/no task. In a yes/no confidence rating task, subjects are asked to rate their confidence that a target is present, whereas in the MED task, subjects describe their phenomenological impression of a specific stimulus quality. One empirical issue is the relation between the two responses. Research has shown that when subjects are asked to judge multiple stimuli, measures of sensitivity derived from labeling procedures (e.g., scaling) are attenuated when compared with those derived from confidence ratings (Irwin, Hautus, Dawson, Welch, & Bayly, 1994; Irwin & Whitehead, 1991). The attenuation in the labeling procedure likely occurs for two reasons: (1) the added variance inherent in labeling a range of stimuli, and (2) the labeling procedure's sensitivity measure is derived from the subject's perception of a single perceptual dimension, whereas the confidence rating's sensitivity measure is derived from an aggregate of all perceptual dimensions. This suggests that confidence ratings and ME responses are generally not equivalent measures.<sup>10</sup>

A second issue that needs to be addressed is the degree to which the ME response is susceptible to task demands. Our data suggest that this is of minimal concern. Recall that the two experiments were identical except for the instructions regarding the relevant stimulus quality to which one should attend. If artificial task demands unduly influenced the subject's responses, one would expect that the results of Experiment 2 would more closely parallel those of Experiment 1. Nevertheless, a series of experiments are needed to fully understand this issue.

It should be noted that the MED task suffers from the same limitations as the traditional magnitude estimation procedure. Because the limitations of traditional magnitude estimation procedures have been discussed extensively elsewhere (e.g., Gescheider, 1988; Marks & Algom, 1998; Stevens, 1975), we will not delineate them here. Despite the limitations of the ME procedure, our data suggest that the effect of these limitations was negligible. As-

suming that the subject's responses in the signal detection task are, at worst, only minimally affected by the same biases that may affect the subject's ME response, the high concurrent validity between  $d_{ME}$ ,  $d'_{ME}$  and  $d'_{SDT}$  in Experiment 1 and the discriminant validity between  $d_{ME}$ ,  $d'_{ME}$ , and  $d'_{SDT}$  in Experiment 2 suggest that the effects of these biases on  $d_{ME}$  and  $d'_{ME}$  were also minimal.

In conclusion, two experiments showed that the subject's ME responses in the MED task may accurately quantify the capacity of a perceived stimulus quality to predict the presence of a target stimulus, as determined by  $d'_{SDT}$ . The data further reveal that perceived tone intensity is highly predictive of a target's presence, whereas perceived tone duration is not. Our data suggest that this is the case because tone intensity is a less perceptually noisy dimension than tone duration. We presented the advantages and disadvantages of this task with the hope that this task can aid researchers in better understanding a variety of perceptual and cognitive processes and yield exciting new insights into the detection process.

## REFERENCES

- BRAIDA, L. D., & DURLACH, N. I. (1970). Intensity resolution. II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America*, **51**, 483-502.
- BUUS, S., FLORENTINE, M., & POULSEN, T. (1999). Temporal integration of loudness in listeners with hearing losses of primarily cochlear origin. *Journal of the Acoustical Society of America*, **105**, 3464-3480.
- COHEN, D. J. (1997). Visual detection and perceptual independence: Assessing color and form. *Perception & Psychophysics*, **59**, 623-635.
- COHEN, J., & COHEN, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- CREELMAN, C. D. (1964). Human discrimination of auditory duration. In J. A. Swets (Ed.), *Signal detection and recognition by human observers: Contemporary readings* (pp. 265-290). New York: Wiley.
- DAI, H., & WRIGHT, B. A. (1995). Detecting signals of unexpected or uncertain durations. *Journal of the Acoustical Society of America*, **98**, 798-806.
- DAI, H., & WRIGHT, B. A. (1999). Predicting the detectability of tones with unexpected durations. *Journal of the Acoustical Society of America*, **105**, 2043-2046.
- DURLACH, N. I., & BRAIDA, L. D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, **46**, 372-383.
- EFRON, B. (1982). *The jackknife, the bootstrap, and other resampling plans*. Philadelphia: Society for Industrial and Applied Mathematics.
- GESCHEIDER, G. A. (1988). Psychophysical scaling. *Annual Review of Psychology*, **39**, 169-200.
- GREEN, D. M., BIRDSALL, T. G., & TANNER, W. P. (1964). Signal detection as a function of signal intensity. In J. A. Swets (Ed.), *Signal detection and recognition by human observers: Contemporary readings* (pp. 243-264). New York: Wiley.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- IRWIN, R. J., HAUTUS, M. J., DAWSON, N. J., WELCH, D., & BAYLY, M. F. (1994). Discriminability of electrocutaneous stimuli after topical anesthesia: Detection-theory measurement of sensitivity to painful stimuli. *Perception & Psychophysics*, **55**, 125-132.
- IRWIN, R. J., & WHITEHEAD, P. R. (1991). Towards an objective psychophysics of pain. *Psychological Science*, **2**, 230-235.
- KUBOVY, M., COHEN, D. J., & HOLLIER, J. (1999). Feature integration that routinely occurs without focal attention. *Psychonomic Bulletin & Review*, **6**, 183-203.
- LUCE, R. D., & KRUMHANS, C. L. (1988). Measurement, scaling, and psychophysics. In R. C. Atkinson et al. (Eds.), *Stevens' Handbook of ex-*

- perimental psychology: Vol. 1. Perception and motivation (pp. 3-74). New York: Wiley.
- MACMILLAN, N. A., & CREELMAN, C. D. (1990). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- MARKS, L. E., & ALGOM, D. (1998). Psychological scaling. In M. H. Birnbaum (Ed.), *Measurement, judgment, and decision making* (pp. 81-178). New York: Academic Press.
- MILLER, M. B., & WÖLFORD, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review*, **106**, 398-405.
- MUNSON, W. A. (1947). The growth of auditory sensation. *Journal of the Acoustical Society of America*, **19**, 584-591.
- STEVENS, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.
- SWETS, J. A., TANNER, W. P., & BIRDSALL, T. G. (1964). Decision processes in perception. In J. A. Swets (Ed.), *Signal detection and recognition by human observers: Contemporary readings* (pp. 3-57). New York: Wiley.
- THOMAS, R. D. (1999). Assessing sensitivity in a multidimensional space: Some problems and a definition of a general  $d'$ . *Psychonomic Bulletin & Review*, **6**, 224-238.
- WIXTED, J. T., & STRETCH, V. (2000). The case against a criterion-shift account of false memory. *Psychological Review*, **107**, 368-376.

## NOTES

1. The cognitive process likely influences the signal-to-noise ratio by shifting the SN distribution. To give an extreme example, suppose an observer ignores his/her perceptual experience of a stimulus quality that is highly predictive of the presence of the target. In such a case, only less predictive stimulus qualities would be combined to provide the activation level on the decision axis for a particular trial. This would result in a smaller  $d'$  than if the highly predictive stimulus quality was included in the observer's calculation.

2. To obtain a  $d'_{STD} > 5.0$ , a subject would have to be correct on about 497 of 500 trials.

3. We chose to label the bootstrapped measure  $d'_{ME}$  because it is an estimated distance, whereas one directly measures the distance when calculating  $d_{ME}$ .

4. One can calculate a robust measure of sensitivity that makes no distributional assumptions and is invariant across transformations by converting the ME data into the area under the ROC. This measure is equivalent to the predicted percent correct in a two-alternative forced-choice task (Green & Swets, 1966; Macmillan & Creelman, 1990). One can also convert  $d'_{SDT}$  to the area under the ROC, if one assumes that the SN and N distributions have equal variance. These two measures can then be compared. Throughout the paper, the results based on this measure of sensitivity are equivalent to those of  $d'_{ME}$ . We therefore only present those of  $d'_{ME}$ .

5. Although a true magnitude estimation task should allow for any nonnegative real number, subjects generally use between 10 and 20 numbers (Luce & Krumhansl, 1988). Furthermore, when a set of restricted

responses is used, the results are essentially unchanged if that set is sufficiently large (Durlach & Braida, 1969). Given that we are assessing subjects' perceived intensity of a single target stimulus (and no stimulus) with a response set that is more than double that generally used by subjects to estimate many target stimuli, our response set should be sufficiently large so as not to affect the results.

6. The "Original Intensity" button was not placed in the center of the scale, so that the subjects could express intensities greater than twice that of the original intensity. The scale also allowed the subjects to express intensities less than half that of the original intensity.

7. To assess whether the quadratic term is necessary to account for only the very high  $d'$ s, several regression analyses [ $d'_{SDT} = \delta(d_{ME}) + \epsilon(d_{ME}^2)$ ] were computed on truncated ranges of the data. Specifically, separate regressions were computed for  $d'_{SDT} < \{2, 2.5, 3, 3.5, 4, 5\}$ , and the entire data set. The results of these analyses indicated that the quadratic term is significant at  $\alpha = .05$  for all subsets of the data except  $d'_{SDT} < 2.0$ . The findings for  $d'_{SDT} < 2.0$  ( $p = .06$ ) appear to be a function of the low power ( $n = 11$ ), because the magnitude of the effect is not compromised. Thus, the data indicate that (1) the quadratic term is not merely a function of the very high values of  $d'$  and (2) the MEs of stimulus intensity provide a very consistent measure of the subject's sensitivity to the target stimulus, as measured by the  $d'_{SDT}$ .

8. The  $d_{ME}$  is calculated using the average  $SD$ s of the SN and N distributions,  $(\sigma_N + \sigma_{SN})/2$ , whereas the  $d'_{SDT}$  assumes that the  $SD$ s are equal,  $\sigma_N/\sigma_{SN} = 1$ . Because  $\sigma_N$  and  $\sigma_{SN}$  of the ME data were not equal,  $\sigma_N/\sigma_{SN} = 0.8$  [ $t(40) = 3.06, p = .004$ ],  $d'_{SDT}$  and  $d_{ME}$  might not be expected to be linearly related. To assess this possibility, we compared  $d_{ME}$  to  $d'_c$ , which is a measure of sensitivity that is computed from the yes/no data using  $(\sigma_N + \sigma_{SN})/2$ . Specifically,  $d'_c = [2/(1+s)][z(H) - s_z(F)]$ , where  $s = \sigma_N/\sigma_{SN}$  (i.e., the slope of the ROC when plotted on  $z$ -coordinates), and H and F are the hit and false alarm rates from the yes/no data, respectively (Macmillan & Creelman, 1990). If the quadratic component emerges primarily because the  $SD$ s of the distributions underlying  $d_{ME}$  are different from those underlying  $d'_{SDT}$ , then the quadratic component should be reduced or eliminated if we calculate  $d'_c$  using the the  $SD$ s of the ME data. This analysis revealed that the function relating  $d_{ME}$  to  $d'_c$  was almost identical to that relating  $d_{ME}$  to  $d'_{SDT}$ ,  $d'_c = 1.26 * d_{ME} - 0.075 * d_{ME}^2$  [ $F(2,39) = 88.9, p < .001, r^2 = .82$ ]. Thus, the quadratic component is not a function of the unequal  $SD$ s of the SN and N distributions in the ME data.

9. The relationship between any two measures is influenced by the reliability of each measure. Specifically, if one or both measures have low reliability, the slope of the function will be shallower than the true slope. Our reliability estimates ranged from .92 to .99. Therefore, the reliability of the measures did not significantly affect the slopes of our functions.

10. The added variance inherent in labeling a range of stimuli was not a factor in our experiments because we asked subjects to label a single stimulus. Furthermore, the attenuation of the labeling procedure's sensitivity measure was small when subjects estimated perceived intensity and large when subjects estimated perceived duration. As we have stated, this pattern of results would occur if subjects primarily used perceived intensity when judging the presence of a target.

## APPENDIX

### Instructions for Experiment 1 (Changes for Experiment 2 Are Italicized)

In this experiment you will be presented with a sound. This sound will be simply random noise. Random noise is the sound that comes out of your radio when it is not tuned to a specific station. Listen now to the noise stimulus. [play noise stimulus]

On half of the trials there will be a high pitched tone embedded in the noise. I will call this tone a "signal." Listen now to the Signal. [play signal stimulus]

This signal will be short, very soft, and placed randomly somewhere within the noise. Thus it will be very hard to hear. Listen now to the signal embedded in the noise. [play noise + signal stimulus]

You will have an opportunity to listen to the signal + noise sound again after the instructions.

Your task is to determine whether or not the noise contains the signal. After each trial you will see a response box. The response box looks like this. [show response box; see Figure A1]

The response box contains two buttons. One button is labeled signal present and the other is labeled signal absent. If you think that you heard the tone embedded in the noise, please press the signal present button. Other-

## APPENDIX (Continued)

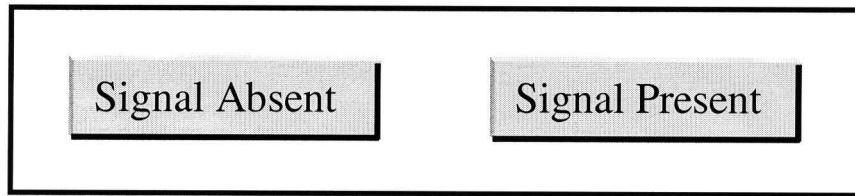


Figure A1. The judgment response screen for Experiments 1 and 2.

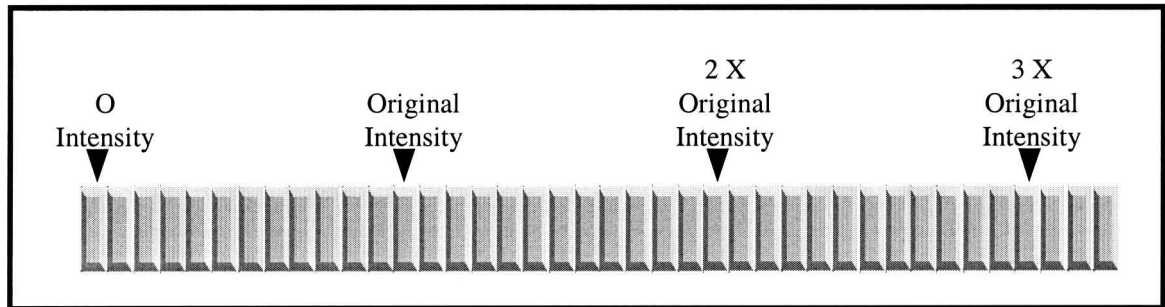


Figure A2. The magnitude estimation response screen for Experiment 1.

wise please press the signal absent button. After you respond a different response box will appear that looks like this. [hide present buttons and show rating; see Figure A2.]

This response box is used to assess the perceptual intensity of the signal on each trial. That is, on each trial the signal may appear to be more or less intense. Even on trials in which you do not believe that a signal was present, you may think that you hear a tone. Therefore, for every trial, please rate the perceptual intensity of the tone you think that you may have heard in the noise, even if you believe that tone is not the signal.

On the scale you will see that there is a continuum of buttons ranging from no tone heard to an intensity greater than three times the original signal's intensity. You should choose the button that accurately represents the intensity of tone you heard in relation to the original signal tone. Please do not feel that you have to use the entire scale. Simply rate the intensity of the tone, as you hear it, in relation to the original signal tone.

[For the duration experiment (Experiment 2), the previous two paragraphs were replaced with the following two paragraphs.]

*This response box is used to assess the perceptual duration of the signal on each trial. That is, on each trial the signal may appear to play for more or less time. Even on trials in which you do not believe that a signal was present, you may think that you hear a tone. Therefore, for every trial, please rate the perceptual duration of the tone you think that you may have heard in the noise, even if you believe that tone is not the signal.*

*On the scale you will see that there is a continuum of buttons ranging from no tone heard to a duration greater than three times the original signal's duration. You should choose the button that accurately represents the duration of tone you heard in relation to the original signal tone. Please do not feel that you have to use the entire scale. Simply rate the duration of the tone, as you hear it, in relation to the original signal tone.*

If you would like to hear the test sounds again, you may press the buttons below. The button labeled signal, will play the signal tone. The button labeled noise will play the noise stimulus, and the button labeled signal + noise, will play the stimulus in which the signal is embedded in the noise. Remember, when the signal is embedded in the noise it will be very soft, short, and its placement will be unpredictable. When you are familiar with the test sounds, please press the button labeled Start Experiment to begin the experiment.