

A signal detection theory analysis of gap detection in the rat

DONALD S. LEITNER

Saint Joseph's University, Philadelphia, Pennsylvania

DENNIS P. CARMODY

Saint Peter's College, Jersey City, New Jersey

and

EILEEN M. GIRTEN

Saint Joseph's University, Philadelphia, Pennsylvania

An innocuous sensory event (a prestimulus) briefly preceding a startle-eliciting stimulus (SES) reduces the amplitude of the elicited reflex. This study used signal detection theory (SDT) techniques to quantify the effects of gaps (pauses in otherwise continuous noise) on the rat's acoustic startle reflex. Sixteen rats were given four identical test sessions consisting of the randomized presentation of 150 trials of the SES alone and 150 trials of a gap-and-SES combination. Gap duration (1, 2, 4, and 8 msec) varied between sessions. Data analyses based on amplitude, difference scores, percentage scores, and SDT techniques identified similar patterns. The three longest gaps, but not the shortest, were reliably detected, and differences among these three were identified with percentage and SDT analyses. Analyses of amplitude changes over test sessions yielded different patterns for each measure. The results demonstrate that an SDT analysis is a sensitive index of prestimulus effects.

Many reflexes, including the whole-body startle reflex of the rat and the human eye-blink, can be modulated in amplitude if the reflex-eliciting stimulus is preceded by a detectable change in the sensory environment (a prestimulus). Prestimuli do not themselves elicit startle or any other overt response, yet they can greatly reduce the amplitude of a subsequently elicited reflex. (For reviews, see Hoffman & Ison, 1980; Ison & Hoffman, 1983.) In rats, prestimuli with lead times of between 50 and 150 msec will reduce the amplitude of the startle reflex, although the effect is present to a lesser extent over a wide range of lead intervals (see, e.g., Hoffman, Stitt, & Leitner, 1980).

Reflex modulation by prestimuli makes an excellent paradigm for assessing sensory functioning in both animals and humans. Through comparison of the amplitude of the reflex on trials on which the prestimulus precedes the startle-eliciting stimulus (SES) with the amplitude of the reflex on trials on which SES is elicited without the prestimulus, the detectability of the prestimulus can be monitored objectively. This procedure is superior to other

commonly used techniques that rely either on the cooperation of the subject being tested, as in the verbal reports of human subjects following instructions, or on the training of the subject, as in the motivated, operant baseline performance of other species that is maintained as a response measure by appetitive or aversive schedules (e.g., Kelly & Masterton, 1977).

Sensory assessment using reflex modulation avoids these problems. It does not require training or the cooperation of the subject. It does not require learning; the amplitude of the reflex is reduced on the very first trial on which it is preceded by the prestimulus (Ison, Hammond, & Krauter, 1973). It is not based on a motivated operant baseline performance as a response measure, eliminating the confounding of sensory and motivational changes that may be caused by some independent variables as well as the logistical difficulties of training large numbers of subjects.

An unresolved issue that has stirred much debate is the mathematical procedure used to quantify the effect of a prestimulus. A common method is simply to see whether the mean amplitude of the elicited reflex on trials on which the prestimulus is absent (the baseline reflex amplitude) is reliably larger than the mean of the trials on which the prestimulus is present (see, e.g., Leitner et al., 1993). This works well as long as other variables do not alter the amplitude of the elicited reflex. If a variable under study results, for example, in an increase in reflex amplitude on the prestimulus trials but the baseline reflex amplitude remains constant, then interpretation is straightforward. Since the variable has disrupted the prestimulus' ability

The authors would like to thank Donald Dorfman, Kevin Berbaum, and Charles Metz for generously supplying the software used in the signal detection theory analyses detailed in this manuscript. The authors would also like to thank Elizabeth Mitchell for her help in data collection. Portions of the analyses were supported by a Faculty Fellowship from Saint Peter's College and by a Scholar-in-Residence Award from the Faculty Resource Network, New York University, granted to D.P.C. Reprint requests and other correspondence concerning this manuscript should be addressed to D. S. Leitner, Department of Psychology, Saint Joseph's University, 5600 City Avenue, Philadelphia, PA 19131-1395 (e-mail: dleitner@sjuphil.sju.edu).

to reduce startle amplitude, the variable has also disrupted the processing of the prestimulus.

Unfortunately, many variables tend to alter the amplitude of the baseline startle reflex over the course of an experiment. Habituation (Geyer & Braff, 1987; Geyer, Swerdlow, Mansbach, & Braff, 1990), the effects of drugs (Crofton, 1992; Geyer et al., 1990), and aging (Krauter, Wallace, & Campbell, 1981), to name but three, can alter the amplitude of a repeatedly elicited reflex over the course of an experiment. In such situations—when baseline reflex amplitude changes at the same time as does the amplitude on trials where on which prestimulus is present—interpretation becomes difficult. If, for example, one wishes to study a drug that, when administered, may disrupt sensory processing, and if the drug is found to alter both baseline amplitude and the amplitude on prestimulus trials when compared with a placebo, what does this indicate about the drug's effect?

Hoffman and Ison (1980), in reviewing the reflex modulation literature, noted that the "presence of an inhibitory prepulse subtracts the same amount from the small or moderate response to a startle-eliciting stimulus of weak or medium intensity as it does from the relatively large response to an eliciting stimulus of great intensity. In other words, the amount of inhibition is independent of the intensity of the stimulus that elicits the to-be-inhibited response" (p. 179). This is also the position taken by Davis (1988).

Thus, if a variable changes baseline reflex amplitude, an appropriate way of quantifying the effects of a prestimulus would be, for each subject, to subtract the mean of the trials on which the prestimulus was present from the mean of the trials on which the SES was presented alone. The resulting difference score would be constant in size and independent of any changes in baseline reflex amplitude. If the administration of a drug alters both baseline amplitude and the amplitude on prestimulus trials, difference scores should allow one to compare the effectiveness of the prestimulus under drug and placebo conditions, and to infer the drug's effect on the processing of the prestimulus.

More recently, Bowen, Barlow, and Ison (1992) presented data indicating that the situation is not this simple. Consistent with previous research, they found that if reflex amplitude is increased by the use of a more intense SES, the difference between the means of the SES-alone trials and of the prestimulus trials is indeed a constant amount; thus difference scores would seem to be a useful index of prestimulus effects. But their data also demonstrate that when baseline reflex amplitude is altered by the presence of factors that affect the more general state of the organism, such as diurnal rhythmicity, a prestimulus will subtract a larger amount from a large response than from a small response.

This would mean that difference scores may not be the best index of prestimulus effectiveness in such situations, because these scores would vary with the amplitude of the baseline reflex and would therefore be confounded with whatever variable was altering baseline amplitude. In these

situations, changes in prestimulus effectiveness quantified with difference scores would be artifacts of a variable's effects on baseline amplitude. Bowen et al. (1992) suggest that because a prestimulus subtracts a larger amount from a large response than from a small response, a better measure of the effectiveness of a prestimulus would be obtained by expressing the amplitude of the prestimulus trials as a ratio of the SES-alone trials—that is, as a proportion or a percentage score. Such a measure would be constant in size, and independent of any changes in baseline reflex amplitude.

Thus, there is currently no agreement on the best way to quantify and interpret the effects of a prestimulus. Neither untransformed reflex amplitude nor transformed scores such as difference scores or percentage scores can serve as a "pure" measure of sensory functioning. This controversy has hindered the utility of this otherwise powerful and objective tool for sensory assessment.

An alternative approach to analyzing prestimulus effects is the powerful technique of signal detection theory (SDT). In a typical experiment in which a prestimulus is used in a reflex modulation paradigm, the prestimulus is usually of low intensity and presented repeatedly, and the research question under study is how the organism's processing of such a hard-to-detect stimulus is altered. This is all very similar to a typical signal detection experiment.

Analysis of prestimulus effects using SDT offers some advantages over analyses of amplitude data or percentage or difference scores. SDT separates sensory processing into two independent indices: sensitivity, which reflects the acuity of an organism's sensory system; and criterion, which reflects the bias present in the organism's decision about whether or not a stimulus is present (Gescheider, 1985; Macmillan & Creelman, 1991). Because SDT allows the calculation of a separate, relatively pure index of sensory sensitivity, applying an SDT-based analysis to prestimulus effects would permit an objective assessment of changes in sensory processing.

A further advantage of an SDT-based analysis of prestimulus effectiveness would be that it avoids the necessity of using the mean baseline amplitude in its calculations. As outlined above, this is a major problem when one uses difference or percentage scores. Because of this, an SDT-based analysis would not be as sensitive to fluctuating response baselines and would thus avoid the potential confound inherent in the other approaches.

The research reported here was an attempt at applying SDT techniques to the analysis of prestimulus effects on the rat's acoustic startle reflex. The purpose was to see whether mathematical analyses of prestimulus effects based on SDT were feasible, and if so, whether the resulting data would parallel the patterns found with the use of more conventional analyses. If successful, it would permit future use of SDT-based analyses in the study of variables that affect baseline amplitude.

Gaps—brief pauses in otherwise continuous background noise—were used as prestimuli to reduce the amplitude of a subsequently elicited acoustic startle reflex

in rats. Gap prestimuli are effective in reducing startle amplitude and are believed to reflect the ability of the auditory system to detect and process rapidly changing acoustic transients (Ison, 1982; Ison, O'Connor, Bowen, & Bocirnea, 1991; Leitner et al., 1993; Wecker & Ison, 1984). As such, they provide an index of the temporal resolution of the auditory system, an important factor in human speech comprehension (Buus & Florentine, 1985; Corso, Corso, & Corso, 1981; McCroskey & Kidder, 1980; Trinder, 1979).

METHOD

Subjects

The subjects were 16 male albino Sprague-Dawley rats, approximately 90 days old at the beginning of testing. They were individually housed and allowed ad-lib food and water except during testing. They were maintained on a 10:14-h reversed dark:light cycle and were tested within a few hours of the midpoint of the "night" portion of the cycle.

Apparatus

An Apple IIe microcomputer was used to control all experimental variables. Two subjects were tested simultaneously. Startle amplitude was assessed for each rat in one of two identical startle stations. Each station consisted of a plastic tub ($20 \times 10 \times 10$ cm) fitted with a lid that snapped onto the top, containing the subject. A rectangular section (15×7.5 cm) cut out of the lid and replaced with an aluminum mesh allowed good acoustic penetration.

The dependent variable was the amplitude of the rat's whole-body acoustic startle reflex. During testing, the tub was placed on a Coulbourn Instruments movement transducer platform (Model E45-12). The voltage from the platform's strain gauge bridge (Coulbourn Instruments, Model S72-25) was filtered, rectified, integrated, and passed to a Computer Continuum 8-bit analog/digital-digital/analog converter. The microcomputer sampled the incoming voltage at a frequency of 1000 Hz for 150 msec, beginning with the onset of the SES; the peak value during this interval served as the rat's response amplitude on that trial.

The two startle stations were located in a sound-attenuated, anechoic chamber. Located 22.5 cm above each platform were three ultra-high-frequency tweeters (Radio Shack, Model 40-1377). Two of these transducers were used to deliver the SES; the third was used for the white noise in which the gap was embedded.

The white noise was set at 60 dB SPL(A). The rise-fall time of the gap was controlled by increasing or decreasing the gain of a voltage-controlled amplifier (Coulbourn Instruments, Model S77-05); the controlling voltage was supplied by the digital/analog converter, which varied its output voltage according to a cosine function. The SES was a 120-dB SPL(A) burst of white noise, with a 100- μ sec rise-fall time and a duration of 50 msec. Both the background noise and the SES were produced by Coulbourn Instruments white noise generators (Model S81-02), modified so that their output spectra were flat to approximately 100,000 Hz. The stimuli were amplified with a Crown solid-state stereo amplifier (Model DC300).

Procedure

Each rat was exposed to four test sessions, spaced at least 48 h apart. The duration of the gap prestimulus used (1, 2, 4, or 8 msec) was varied between test sessions. They were presented to each rat in a sequence determined by a Latin square matrix. These gap durations were selected because previous research had demonstrated that 2-msec-long gaps are close to detection threshold for rats (Ison, 1982; Ison et al., 1991; Leitner et al., 1993). The time between the end of the gap and the onset of the SES was fixed at 50 msec for all gap durations.

Each 90-min test session consisted of the repeated presentation of the SES, with an intertrial interval that varied randomly among 16, 18, 20, 22, and 24 sec. Present throughout testing was the background white noise in which the gap, if present, was embedded. A trial consisted of an average of 20 sec of white noise, and ended with the presentation of the gap (if present) and the SES and the assessment of startle amplitude. The next trial began immediately. Each session included 150 trials on which a gap preceded the SES and 150 trials on which the SES was presented alone, in a random sequence.

RESULTS

Analyses of Startle Amplitude

For these analyses, mean startle amplitude was calculated across trials for each subject for each condition, and grand means were then calculated across subjects for each condition. These data are depicted in Figure 1. It can be seen that mean startle amplitude on the control trials on which the SES was presented alone are all about the same across test sessions. It can be noted further that the 1-msec-long gap did not appreciably reduce startle amplitude, but that the other gap durations did.

The validity of these trends was examined with a 2 (startle-alone trials vs. gap-and-startle trials) \times 4 (gap duration) analysis of variance (ANOVA), with repeated measures on both factors. The analysis revealed a reliable effect of trial type [$F(1,15) = 80.82, p < .001$], a reliable effect of gap duration [$F(3,45) = 6.43, p < .01$], and a reliable interaction [$F(3,45) = 46.85, p < .001$]. Further analysis with Tukey's honestly significant difference (HSD) test demonstrated that there were no reliable differences among the four conditions in which the SES was presented alone, and that the mean amplitude for the condition with the shortest gap duration, 1 msec, was not reliably different from the SES-alone trials with which it was paired. Further testing with Tukey's HSD test demonstrated that the mean amplitude for the conditions with the three longest gap durations were reliably different from the SES-alone trials with which each was paired ($p < .05$). The mean amplitude for the condition with the shortest gap was also found to be reliably different from the means of the other three gap conditions ($p < .05$), which were not reliably different from each other.

These data are consistent with previous research (Ison, 1982; Ison et al., 1991; Leitner et al., 1993) in which a 1-msec-long gap did not reliably reduce startle amplitude, whereas a gap that was 2 msec in duration did. Also, these data are consistent with the findings of Leitner et al. (1993) that a 2-msec-long gap was not reliably less inhibitory than longer gaps.

Analyses of Difference Scores

For these analyses, mean startle amplitude was calculated across trials for each subject for each condition. Difference scores were then calculated by subtracting the mean startle amplitude for each of the four gap-and-startle conditions from the corresponding SES-alone condition, for each subject. This method of calculation results in small difference scores representing large responses

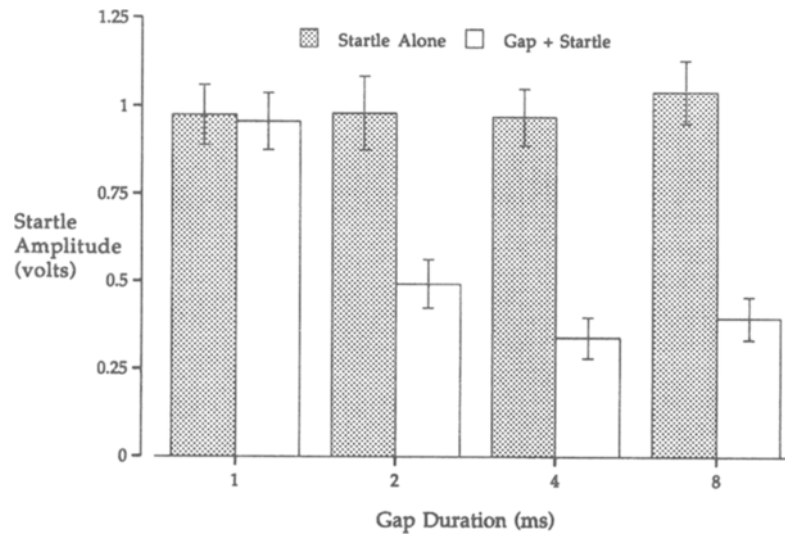


Figure 1. Mean startle amplitude in volts ($\pm 1SE$) for each SES-alone condition and its corresponding gap-and-SES condition.

made on the prestimulus trials (i.e., a lesser prestimulus effect), and large scores representing small responses made on the prestimulus trials (i.e., a greater prestimulus effect). Four grand means were then calculated across these mean difference scores.

These data appear in Table 1. It can be seen that the difference score for the 1-msec-long gap condition is almost zero, indicating that this short gap did not reduce startle amplitude. For the other gap durations, the difference scores are much larger, with the 2-msec-long gap producing somewhat less of a difference than either of the two longer gaps.

Data analysis consisted of a one-way repeated measures ANOVA, supplemented with post hoc Tukey's HSD tests. The ANOVA revealed a reliable difference among mean difference scores [$F(3,45) = 46.85, p < .001$]. Post hoc analyses indicated that the mean difference score for the 1-msec-long gap was reliably different from that for the other three ($p < .05$), and that the mean difference scores for the 2-, 4-, and 8-msec-long gaps were not reliably different from each other. Again, these data are consistent with previous research (Ison, 1982; Ison et al., 1991; Leitner et al., 1993).

Analyses of Percent Amplitude Reduction

For these analyses, percent amplitude reduction scores were calculated for each subject by dividing each difference score for each of the four gap conditions (as calculated for the second analysis, above) by the subject's mean amplitude for the corresponding SES-alone trials, and then subtracting the quotient from 1. The resultant from this was then multiplied by 100. Four mean percent amplitude reduction scores, one for each of the four gap conditions, were then calculated across subjects. This method of calculation results in small percent amplitude reduction scores representing small responses made on the

prestimulus trials (i.e., a greater prestimulus effect), and large scores representing large responses made on the prestimulus trials (i.e., a lesser prestimulus effect).

These data are also presented in Table 1. It can be seen that a pattern similar to the second analysis is present; the percent score for the 1-msec-long gap condition is virtually 100, indicating that this gap did not reduce startle amplitude. For the other gap durations, the percentage scores are much smaller, with the 2-msec-long gap again not as effective in reducing startle amplitude as either of the two longer gaps.

These data were analyzed with a one-way repeated measures ANOVA, supplemented with post hoc Tukey's HSD tests. The ANOVA revealed a reliable difference among mean percent scores [$F(3,45) = 84.87, p < .001$]. Post hoc analyses indicated that the mean percent amplitude reduction score for the 1-msec-long gap was reliably different from those for the other three ($p < .05$); that the mean percent amplitude reduction score for the 2-msec-long gap was reliably different from those for the two longer durations ($p < .05$); and that the mean percent amplitude reduction scores for the 4- and 8-msec-long gaps were not reliably different from each other.

Table 1
A Comparison of Three Different Methods Used to Quantify the Effects of Gap Prestimuli of Four Different Durations

Method	Gap Duration (in Milliseconds)							
	1		2		4		8	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Difference	0.02 _a	0.03	0.49 _b	0.08	0.62 _b	0.07	0.64 _b	0.08
Percent	99.10 _a	2.94	50.09 _b	5.33	33.88 _c	4.15	37.19 _c	4.39
<i>A_z</i>	.51 _a	.01	.76 _b	.03	.85 _c	.02	.84 _c	.02

Note—Means with different subscripts differ significantly at $p < .05$ in the Tukey HSD comparison.

Signal Detection Analyses

The data were examined by means of a receiver operating characteristic curve (ROC) rating analysis—the multireader, multicase approach (MRMC, Version 1.55) developed by Dorfman, Berbaum, and Metz (1992). This approach permits generalizations to the population of readers and cases. Based on the ideas of Swets and Pickett (1982), Hanley and McNeil (1983), and Metz, Wang, and Kronman (1984), it involves an analysis of variance of the pseudovalues computed by the Quenouille–Tukey version of the jackknife (Dorfman et al., 1992). Decision accuracy is measured by A_z , the area under the ROC curve. The two underlying distributions are the responses to SES-alone trials and to the gap-and-SES trials. The ability of the subject to detect the gap is reflected in the value of A_z , with an A_z of .50 representing no detection and an A_z of 1.0 representing perfect detection.

As applied to the gap detection paradigm, the standard response was a large-amplitude startle response, a response typically exhibited when no gap was either available or detected. A low-amplitude startle response represented either inhibition of the response due to a detection of a gap or to a failure to respond. Startle amplitude was thus used as a strength of decision, with larger amplitudes suggesting more likelihood of gap detection. It was assumed that for each subject in each test session, the startle amplitudes covered a range of magnitudes from the complete amplitude reduction possible for the gap duration used to the maximum startle amplitude demonstrated by the subject within the block of 300 trials.

It has been shown that the amplitude of the startle reflex for rats varies over time (Chabot & Taylor, 1992a, 1992b; Davis & Sollberger, 1971; Horlington, 1970; Krauter et al., 1981). Furthermore, subject-to-subject variability of reflex amplitude is such that many researchers working with rats in startle paradigms use a matching procedure to roughly equate reflex amplitude between groups of subjects (see, e.g., Davis, 1988). Because of this between- and within-subjects variability, and the Latin square design used in the present study in which each condition for each subject occurred on different days, performance was assessed for each block of 300 trials for each rat. In this manner, neither day-to-day variation for each rat nor rat-to-rat variation contributed to estimates of performance. The advantage of this approach is to avoid the comparison of gap trials on one day with baseline data collected on a different day, thus avoiding the day-to-day source of variability. Each estimate of A_z is thus calculated for 1 rat observer on one day of performance.

The 300 trials of each test session were used to create a frequency distribution, and cut points were calculated to divide the distribution into six categories of 50 trials each. Cut points corresponded to the amplitudes defining the 16.7th, 33.3rd, 50th, 66.7th, and 83.3rd percentiles. The startle amplitudes for the 16 subjects in the four conditions were thus transformed into six categories, with 1 representing the *smallest startle amplitudes*, and 6 representing the *largest startle amplitudes*. The categorized responses for each block of 300 trials were further

identified as SES-alone (*noise*) or gap-and-SES (*signal*). The 150 SES-alone and the 150 gap-and-SES trials and the six category designations were used to create a 2×6 frequency matrix. If there was no detection of the gap, indicated by no inhibition of startle, then a random arrangement of frequencies would yield an A_z estimate near .5. A perfect separation of SES-alone from gap-and-SES trials would yield an A_z of 1.0.

The MRMC program accommodated the 19,200 ratings (16 subjects \times 4 conditions \times 300 trials) in a single analysis of A_z values. As shown in Table 1, on the average, a 1-msec-long gap is not detectable, a 2-msec-long gap is reliably detected above chance, and detection performance peaks with a 4-msec-long gap, which is not different from the 8-msec-long gap.

These trends were examined by using A_z as the dependent variable in a 4 (gap duration) \times 16 (subject) \times 300 (trial) ANOVA, with repeated measures on the first and third factors. There was a significant effect for gap duration [$F(3,50) = 113.74, p < .001$]. Reliable effects were also found for subject [$F(15,4485) = 24.69, p < .001$] and trial [$F(299,4485) = 7.59, p < .001$]. Significant interactions were found for gap duration \times subject [$F(45,13455) = 5.17, p < .001$] and gap duration \times trial [$F(897,4485) = 1.27, p < .001$].

Post hoc testing using Tukey's HSD test revealed reliable differences between 1- and 2-msec-long gaps ($p < .01$) and between 2- and 4-msec-long gaps ($p < .01$), but not between 4- and 8-msec-long gaps ($p > .05$). Other effects of interest included the reliable trial effect. This indicates that habituation, learning, or perhaps fatigue (of either the sensory or motor systems) processes occurred within each block. The reliable subject \times trial and gap duration \times trial interactions indicate that this process affected each subject differently in each condition tested.

It may be argued that the conversion of continuous data into ordinal data results in a loss of information and precision of measurement. To examine this possibility, three additional analyses were performed on the data. These analyses were identical to the one described above, except that the number of categories created was varied. For the first, 5 categories were created; for the second, 10 categories; and for the third, 20 categories were created.

The mean values of A_z for each gap condition for each of these four analyses were virtually identical, falling within $\pm .005$ of each other, and with *SEs* that fell within $\pm .001$ of each other. That no differences were present among these four types of categorizations was confirmed by a one-way ANOVA, with number of categories as the factor under test [$F(3,12) < 1.00, p > .05$]. These results support the conclusion that the findings reported above were not dependent on the number of categories used in the analyses.

Although startle amplitude was not intentionally manipulated in this study, the reliable trial effect noted above indicated that this issue might be examined by analyzing the first 50 trials (25 startle-alone trials and 25 gap-and-startle trials) and the last 50 trials separately for each subject. If habituation or some other trial-correlated variable

caused a reduction of startle amplitude over the duration of a test session, then such analyses would allow the assessment of the stability of the several measures in the face of fluctuating baseline startle amplitudes. Figure 2 shows the mean for each condition, calculated across subjects, for each of the four methods used to quantify gap inhibition of startle computed for the first 50 (early) and the last 50 (late) of the 300 trials.

To see whether there was a change in startle amplitude from the beginning to the end of a test session, the untransformed startle amplitude data were analyzed in a 2 (early vs. late) \times 2 (SES-alone vs. gap-and-SES) \times 4 (gap duration) ANOVA, with repeated measures on all factors. Reliable main effects were found for early/late [$F(1,15) = 14.81, p < .01$], SES-alone/gap-and-SES [$F(1,15) = 56.04, p < .01$], and gap duration [$F(3,45) = 3.92, p < .01$]. In addition, reliable interactions were found for early/late \times SES-alone/gap-and-SES [$F(1,15) = 32.64, p < .01$] and SES-alone/gap-and-SES \times gap duration [$F(3,45) = 23.14, p < .01$]. The interaction between early/late and gap duration was not reliable [$F(3,45) = 2.59, p > .05$], nor was there a reliable three-way interaction [$F(3,45) = 1.51, p > .05$].

These analyses demonstrate that there was a reliable change in startle amplitude from early to late in a test ses-

sion. Means computed for only the SES-alone trials across gap durations for the first and last 50 trials demonstrated that the reduction in startle amplitude was 44% (first 25 trials, $M = 1.34$ V, $SE = 0.1$ V; last 25 trials, $M = 0.75$ V, $SE = 0.06$ V). The reduction in amplitude across test sessions may possibly have been caused by the subjects' habituation to the repeatedly presented SES. If so, it is not clear whether this reduction reflects habituation of the sensory or motor systems involved.

The inhibitory effect of gaps on startle also varied from the early to the late trials, as demonstrated by the reliable interaction found for early/late \times SES-alone/gap-and-SES. It is not evident, though, whether this effect is reflective of the subjects' altered perception of the gaps over the course of a test session or merely an artifact of measurement caused by the large change in startle amplitude that occurred across a test session, as discussed above.

Further analyses were conducted on the three transformed measures of startle inhibition. For difference scores, a 2 (early vs. late) \times 4 (gap duration) ANOVA with repeated measures on both factors revealed reliable main effects for early/late [$F(1,15) = 32.64, p < .01$] and gap duration [$F(3,45) = 23.14, p < .01$]. The interaction was not reliable [$F(3,45) = 1.51, p > .05$]. Post hoc Tukey's HSD tests revealed that for both the early and late trials,

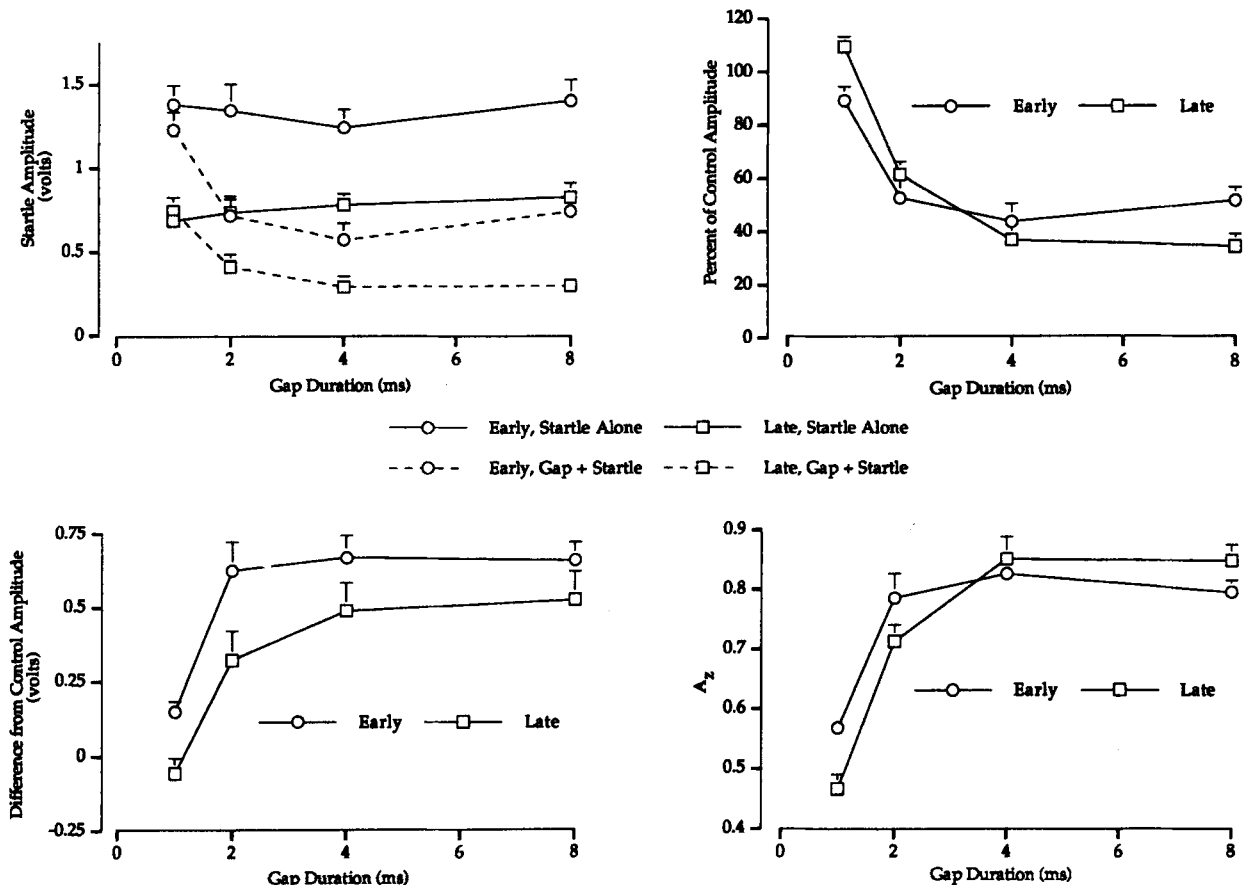


Figure 2. A comparison of four different methods used to quantify the effects of four gap prestimuli, calculated for the first 50 and last 50 trials of a test session for each subject. The values depicted are means (± 1 SE) calculated across subjects.

the 1-msec-long gap produced a reliably smaller difference score (i.e., a lesser effect of gap) than did the other three gap durations ($p < .05$), which were not different from each other. Furthermore, for each gap duration except the 8-msec-long gap, the difference score for the early trials was reliably larger than the difference score for the late trials ($p < .05$), indicating greater gap effectiveness during the early trials. No reliable difference was found between the early and late trials for the 8-msec-long gap.

Analyses identical to those used for the difference scores were conducted for the percent and A_z measures. For percent, no reliable main effect was found for early/late [$F(1,15) < 1, p > .05$]. There was a main effect for gap duration [$F(3,45) = 39.61, p < .01$], and the interaction was reliable [$F(3,45) = 8.33, p < .01$]. Post hoc Tukey's HSD tests revealed that for the early trials, the 1-msec-long gap produced a reliably larger percent score (i.e., a lesser effect of gap) than did the other three gap durations ($p < .05$), which were not different from each other. For the late trials, the 1-msec-long gap again produced a reliably larger percent score than did the other three gap durations ($p < .05$), and the 2-msec-long gap resulted in percent scores that were reliably larger (indicating a lesser effect of gap) than those for the 4- and 8-msec-long gaps ($p < .05$), which were not different from each other.

The 1-msec-long gap resulted in a reliably larger percent score on the late trials than on the early trials ($p < .05$). The early and late percent scores generated by the 2-msec-long gaps were not reliably different, nor were those for the 4-msec-long gap. For the 8-msec-long gap, the early trials produced reliably larger percent scores than did the late trials ($p < .05$). Thus, the interaction can be explained by the differences between the early and late trials noted for the 1- and 8-msec-long gaps.

The analyses for A_z were based on the values derived from the procedure that used six categories. As was found for percent, no reliable main effect was found for early/late [$F(1,15) = 3.94, p > .05$]. Also as for percent, there was a main effect for gap duration [$F(3,45) = 51.49, p < .01$] and a reliable interaction [$F(3,45) = 8.28, p < .01$]. Post hoc Tukey's HSD tests revealed that, as for percent, for the early trials the 1-msec-long gap was reliably less detectable than the other three gap durations ($p < .05$), which were not different from each other. Also as for percent, for the late trials, the 1-msec-long gap was reliably less detectable than the other three gap durations ($p < .05$), and the 2-msec-long gap was reliably less detectable than the 4- and 8-msec-long gaps, which were not different from each other.

Following the trend noted for the percent scores, the 1-msec-long gap had a reliably greater detectability on the early trials than on the late trials ($p < .05$). This was also true for the 2-msec-long gap ($p < .05$). The early and late A_z scores generated by the 4-msec-long gap were not reliably different, nor were those for the 8-msec-long gap. Thus, the interaction can be explained by the reliable differences between the early and late trials noted for the 1- and 2-msec-long gaps, which were not present for the 4- and 8-msec-long gaps.

In summary, gap detectability appears to have changed over the course of a test session, and the three different methods used to quantify startle amplitude reduction resulted in three different patterns. By all three measures, for the first 50 trials a 2-msec-long gap was as detectable as gaps that were 4 and 8 msec long, compared with the 1-msec-long gap. This pattern also held for difference scores for the last 50 trials, and, with the exception of the longest gap used, detectability as quantified by this measure worsened over the course of a test session. The percent scores indicate that late in a test session detectability lessened for the 1-msec-long gap and was greater for the 8-msec-long gap, and that no change occurred for 2- and 4-msec-long gaps. For A_z , detectability of 4- and 8-msec-long gaps during the last 50 trials did not differ from each other or from detectability for the first 50 trials, and this was superior to that for gaps of 1 and 2 msec.

All three measures converged on a lessened sensitivity to the 1-msec-long gap over a test session, and in addition, difference scores and A_z converged on a lessened sensitivity to the 2-msec-long gap over a test session. As indicated by the literature cited above, a 1-msec-long gap is not reliably detected and a 2-msec-long gap is probably near threshold. These gaps might therefore be the most susceptible to alterations in sensory processing that may occur over the course of a test session. Changes in the detectability of the two briefest gaps may explain the trials effect and the interactions noted in the first MRMC analyses.

Explanation of this loss in sensitivity would require another series of experiments in order to attribute it to receptor or subject fatigue, response habituation, or an effect of sampling. It is also unclear what might cause the increase in sensitivity noted for the longest gap over a test session when quantified in percent. Further experimentation would be necessary to determine whether this was a real increase in sensitivity or an artifact of measurement.

DISCUSSION

The purposes of the research described here were to see whether an SDT approach would yield results similar to those of more conventional analyses, and whether the SDT analyses would be applicable in situations in which the baseline amplitude might be affected by independent variables. The SDT analyses indicated that rats did not reliably detect 1-msec-long gaps in acoustic noise, a finding consistent with the three conventional analyses used and with previous reports. Rats reliably detected 2-msec-long gaps, findings also identified by the other three analyses and consistent with previous research (Ison, 1982; Ison et al., 1991; Leitner et al., 1993). The SDT analyses indicated that detection of these two gaps worsened over the course of a test session, whereas detection of the two longer gaps did not. Additional analyses of the data indicated that estimates of A_z were not affected by the number of categories used for the MRMC analysis.

These results are also consistent with research using other techniques to characterize the limits of auditory tem-

poral resolution. These experiments have been conducted with humans (e.g., Penner, 1977; Perrott & Williams, 1971; Plomp, 1964; Williams, Elfner, & Howse, 1979) and several other species, including the house finch (Dooling, Zoloth, & Baylis, 1978) and the chinchilla (Giraudi, Salvi, Henderson, & Hamernik, 1980). Regardless of the species studied or the assessment technique used, the bulk of this research indicates that the minimum reliably detected gap is approximately 2–3 msec.

As an example, Giraudi et al. (1980) studied gap detection in the chinchilla by using operant shock-avoidance techniques. They found that detection thresholds varied with the intensity of the noise in which the gap was embedded; the more intense the noise, the briefer the minimum detectable gap. Although traditional statistical analyses were not performed, the psychometric functions generated by this study showed differences in the detectability of the different durations of gaps used. For the condition most similar to the present study, in which 67-dB SPL noise was used, 2-msec-long gaps were below detection threshold (defined as 50% correct responding). Both 4- and 8-msec-long gaps were above detection threshold, and they were not very different from each other (approximately 80% and 95% correct responding, respectively).

Similar conclusions can be inferred from the psychometric functions generated by studies of gap detection in humans (Penner, 1977; Plomp, 1964). It would appear, therefore, that the reliable difference found between 2- and 4-msec-long gaps identified by analyses of percentage scores and A_z is a real one. This difference has not been previously noted, and it demonstrates that the SDT analyses are at least as sensitive in identifying prestimulus effects as the more conventional percent analyses.

The SDT approach used here avoided the necessity of using the mean baseline amplitude in its calculations. As outlined above, fluctuating baseline amplitude is one of the major problems encountered when quantifying prestimulus effects using amplitude, difference scores, or percentage scores. An SDT-based analysis of sensitivity would not be affected by fluctuating response baseline between test sessions because of the normalization of the data within each test session. This may allow an objective and unconfounded assessment of changes in sensory processing, since effects due to the day-to-day variations in baseline amplitude that have been reported elsewhere (Chabot & Taylor, 1992a, 1992b; Davis & Sollberger, 1971; Horlington, 1970; Krauter et al., 1981) are eliminated.

The SDT approach reduces the potential for confounding sensory processing with changes in reflex amplitude to within-session only. That such changes in reflex amplitude occur over the course of a test session was identified in this study by the reliable effect for trials found in the first SDT analysis conducted. Unlike the other techniques used, the SDT approach called attention to this trial-correlated effect.

Further analyses demonstrated that startle amplitude decreased over the course of a test session. The four measures of gap detectability used also varied over the course of a test session. Although overall a trend toward lesser

gap detectability over the course of a test session was present, each measure displayed a unique pattern of change over time. Because of the differences in trends among measures, it is hard to believe that all are confounded by the within-session variability of the amplitude of the startle reflex. If the within-session changes in gap detectability are real perceptual changes, whether they reflect habituation, learning, fatigue (of either the sensory or motor systems), or some other process also needs to be investigated.

REFERENCES

- BOWEN, G. P., BARLOW, J. A., & ISON, J. R. (1992). Relationships between the modification of acoustic startle reflex behavior and variation in its baseline strength in rodents. *Society for Neuroscience Abstracts*, **18**, 1547.
- BUUS, S., & FLORENTINE, M. (1985). Gap detection in normal and impaired listeners: The effects of level. In A. Michelson (Ed.), *Time resolution in auditory systems* (pp. 159-179). Berlin: Springer-Verlag.
- CHABOT, C. C., & TAYLOR, D. H. (1992a). Circadian modulation of the rat acoustic startle response. *Behavioral Neuroscience*, **106**, 846-852.
- CHABOT, C. C., & TAYLOR, D. H. (1992b). Daily rhythmicity of the rat acoustic startle response. *Physiology & Behavior*, **51**, 885-889.
- CORSO, J. F., CORSO, G. M., & CORSO, C. A. (1981, April). *Prediction of speech discrimination from temporal discrimination measures*. Paper presented at the meeting of the Eastern Psychological Association, New York.
- CROFTON, K. M. (1992). Reflex modification and the assessment of sensory dysfunction. In H. Tilson & C. Mitchell (Eds.), *Neurotoxicology* (pp. 181-211). New York: Raven Press.
- DAVIS, M. (1988). Apomorphine, d-amphetamine, strychnine and yohimbine do not alter prepulse inhibition of the acoustic startle reflex. *Psychopharmacology*, **95**, 151-156.
- DAVIS, M., & SOLLBERGER, A. (1971). Twenty-four-hour periodicity of the startle response in rats. *Psychonomic Science*, **25**, 37-39.
- DOOLING, R. J., ZOLOTH, S. R., & BAYLIS, J. R. (1978). Auditory sensitivity, equal loudness, temporal resolving power, and vocalizations in the house finch (*Carpodacus mexicanus*). *Journal of Comparative & Physiological Psychology*, **92**, 867-876.
- DORFMAN, D. D., BERBAUM, K. S., & METZ, C. E. (1992). Receiver operating characteristic rating analysis: Generalization to the population of readers and patients with the jackknife method. *Investigative Radiology*, **27**, 723-731.
- GESCHIEDER, G. A. (1985). *Psychophysics: Method, theory and application* (2nd ed.). Hillsdale, NJ: Erlbaum.
- GEYER, M. A., & BRAFF, D. L. (1987). Startle habituation and sensorimotor gating in schizophrenia and related animal models. *Schizophrenia Bulletin*, **13**, 643-668.
- GEYER, M. A., SWERDLOW, N. R., MANSBACH, R. S., & BRAFF, D. L. (1990). Startle response models of sensorimotor gating and habituation deficits in schizophrenia. *Brain Research Bulletin*, **25**, 485-498.
- GIRAUDI, D., SALVI, R., HENDERSON, D., & HAMERNIK, R. (1980). Gap detection by the chinchilla. *Journal of the Acoustical Society of America*, **68**, 802-806.
- HANLEY, J. A., & MCNEIL, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, **148**, 839-843.
- HOFFMAN, H. S., & ISON, J. R. (1980). Reflex modification in the domain of startle: I. Some empirical findings and their implications for how the nervous system processes sensory input. *Psychological Review*, **87**, 175-189.
- HOFFMAN, H. S., STITT, C. L., & LEITNER, D. S. (1980). The optimum interpulse interval for inhibition of acoustic startle in the rat. *Journal of the Acoustical Society of America*, **68**, 1218-1220.
- HORLINGTON, M. (1970). Startle response circadian rhythm in rats: Lack of correlation with motor activity. *Physiology & Behavior*, **5**, 49-53.
- ISON, J. R. (1982). Temporal acuity in auditory function of the rat: Reflex modification by brief gaps of noise. *Journal of Comparative & Physiological Psychology*, **96**, 945-954.

- ISON, J. R., HAMMOND, G. R., & KRAUTER, E. E. (1973). Effects of experience on stimulus produced reflex inhibition in the rat. *Journal of Comparative & Physiological Psychology*, **83**, 324-336.
- ISON, J. R., & HOFFMAN, H. S. (1983). Reflex modification in the domain of startle: II. The anomalous history of a robust and ubiquitous phenomenon. *Psychological Bulletin*, **94**, 3-17.
- ISON, J. R., O'CONNOR, K., BOWEN, G. P., & BOCIRNEA, A. (1991). Temporal resolution of gaps in noise by the rat is lost with functional decortication. *Behavioral Neuroscience*, **105**, 33-40.
- KELLY, J. B., & MASTERTON, B. (1977). Auditory sensitivity of the albino rat. *Journal of Comparative & Physiological Psychology*, **91**, 930-936.
- KRAUTER, E. E., WALLACE, J. E., & CAMPBELL, B. A. (1981). Sensory-motor functioning in the aging rat. *Behavioral & Neural Biology*, **31**, 367-392.
- LEITNER, D. S., HAMMOND, G. R., SPRINGER, C. P., INGHAM, K. M., MEKILO, A. M., BODISON, P. R., ARANDA, M. T., & SHAWARYN, M. A. (1993). Parameters affecting gap detection in the rat. *Perception & Psychophysics*, **54**, 395-405.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- MCCROSKEY, R. L., & KIDDER, H. C. (1980). Auditory fusion among learning disabled, reading disabled and normal children. *Journal of Learning Disabilities*, **13**, 18-25.
- METZ, C. E., WANG, P. L., & KRONMAN, H. B. (1984). A new approach for testing the significance of differences between ROC curves measured from correlated data. In F. Deconinck (Ed.), *Information processing in medical imaging* (pp. 432-445). The Hague: Martinus Nijhoff.
- PENNER, M. J. (1977). Detection of temporal gaps in noise as a measure of the decay of auditory sensation. *Journal of the Acoustical Society of America*, **61**, 552-557.
- PERROTT, D. R., & WILLIAMS, K. N. (1971). Auditory temporal resolution: Gap detection as a function of interpulse frequency disparity. *Psychonomic Science*, **25**, 73-74.
- PLOMP, R. (1964). Rate of decay of auditory sensation. *Journal of the Acoustical Society of America*, **36**, 277-282.
- SWETS, J. A., & PICKETT, R. M. (1982). *Evaluation of diagnostic systems: Methods from signal detection theory*. New York: Academic Press.
- TRINDER, E. (1979). Auditory fusion: A critical interval test with implications in differential diagnosis. *British Journal of Audiology*, **13**, 143-147.
- WECKER, J. R., & ISON, J. R. (1984). Acute exposure to methyl or ethyl alcohol alters auditory function in the rat. *Toxicology & Applied Pharmacology*, **74**, 258-266.
- WILLIAMS, K. N., ELFNER, L. F., & HOWSE, W. R. (1979). Auditory temporal resolution: Effects of sensation level. *Journal of Auditory Research*, **19**, 265-269.

(Manuscript received June 29, 1995;
revision accepted for publication August 6, 1996.)