(NASA-CR-174462) [HUMAN ANNOYANCE RESPONSE TO AIRCRAFT NOISE] Final Technical Report (Columbia Univ.) 106 p HC A06/MF A01 CSCL 05J
N83-36672 THRU N83-36675
Unclas
G3/53 15136

Psychophysics Laboratory

Columbia University
in the City of New York

SEP 1983
RECEIVED
NASA STI FACILITY
ACCESS DEPT.

This report and the research herein described has been prepared with the partial support of the Office of Naval Research, the U.S. Army Medical Research and Development Command, and other agencies of the United States Government. Inquiries concerning the contents of this report should be made to Professor Eugene Galanter, Department of Psychology, Columbia University, New York, New York 10027.

Reproduction in whole or in part is permitted for any purpose of the United States Government.

Columbia University
Psychophysics Laboratory
324 Schermerhorn Hall
New York, NY 10027

NASA-7

FINAL REPORT -- NSG-1550

31 July 1983

prepared by

Eugene Galanter

The findings of this report are not to be construed as an official National Aeronautics and Space Administration position unless so designated by other authorized documents.

The NASA Technical Officer for this grant is Mr. David Stephens; NASA-Langley Research Center.

N83 36673 D/

Final Report -- NSG-1550

Eugene Galanter

Psychophysics Laboratory, Columbia University

The Psychophysics Laboratory and the ANRL of the NASA-Langley Research Center have been engaged in a cooperative research enterprise since 1975. We have both been trying to untangle and explicate the social-psychological, acoustic, and perceptual aspects of the human annoyance response to aircraft overflight noise. This work has included both laboratory and field research. The field work was designed to assess the basis of community response to aircraft noise. The laboratory studies served to develop models of such response tendencies that might identify aspects of the (single and multiple) events that give rise to the annoyance responses.

The objective of these efforts was to find ways to control the salient features of the annoyance function, and so reduce community annoyance responses. We expected that changing identified parameters of a model could lead to altered response. Environmental variables that can be located by a theoretical model may offer a high benefit/cost ratio in effecting a reduction of annoyance.

Our work over the past several years has concentrated on two aspects of the general problem: 1) How do the effects of many separate exposures to aircraft overflights cumulate within a person's judgmental frame to yield a long-term measure of individual annoyance. 2) How can we assign a numerical value to the magnitude of a person's annoyance that will capture a variety of behavioral consequences, such as complaint behavior, community action, etc.

The first question has lead to the development of multiple exposure models of human annoyance that broke with mere acoustic additivity of one or another kind. During the past one and one half years, these various models have been subjected to a variety of experimental tests. The culmination of this work is described in the dissertation presented to the Graduate School of Arts and Sciences, Columbia University by Nancy Haber. Segments of that dissertation are included in this Final Report as Appendix I.

The second question has resulted in the development of a new scaling technique, named Utility Comparison Scaling, that permits the assignment of numerical values to aversive or desirable events. This work has been prepared for publication, and is being submitted in revised form and with supplementary data to the journal Public Opinion Quarterly. We expect an early editorial decision on this revision of PPL Technical Report NASA-5, dated.

Finally, the general problem of comparative judgment scale construction, of which the Utility Comparison Scale is a special case, has formed the basis for a dissertation for the Ph.D. degree deposited in June 1983, by Richard Popper with the Graduate School of Arts and Sciences. Appropriate sections of that dissertation are included as Appendix II.

On looking back at the variety of activities that this research support has generated I can only conclude that this contribution by NASA has aided the ongoing activities of the Laboratory as well as the educational mission of the University. The total number of students, both undergraduate and graduate that have had their research skills sharpened under the aegis of this grant is 19 undergraduate research assistants, and 5 graduate students of which three have received Ph.D. degrees and one has received an M.A.

During the course of the work, several collateral activities were undertaken that were of direct aid to the major thrust of this research program. One component was a survey designed and analyzed for the FAA to assess the impact of the inauguration of the Concorde supersonic transport. A second collateral study was conducted for the NASA-Langley Research Center to permit the comparison under laboratory conditions of community residents from the New York City area with residents from Hampton, Va. and near-by communities. A final study conducted for the FAA provided laboratory data for the analysis of helicopter noise, and noise from fixed-wing turbine aircraft in both single and multiple overflight conditions.

In summary, I believe that empirical knowledge about the impact of environmental events was accumulated, and used to develop quantitative models of human behavior in the face of such events. The techniques that evolved have enhanced the capabilities of researchers in the social and behavioral sciences. We have shown that survey research methods may be augmented by psychophysical scaling techniques to the benefit of both disciplines. Although we are still short of the goal of calculating the direct and indirect effects of aircraft overflight noise on community residents, we have improved by a considerable degree our understanding of the nature of these community judgments, and how they can be measured and altered.

Appendix I

D2

N83 36674

4

Columbia University
Psychophysics Laboratory
New York, New York 10027

Temporal Characteristics of Emotional Response: A Decay Model of Annoyance Response to Acoustic Noise

1.0 General introduction

Blatz demonstrated in 1925 that physiological responses to a fear-provoking stimulus increase sharply at the onset of the stimulus and eventually return to their respective resting states. If physiological responses to emotion-provoking stimuli decay after the onset of the stimuli, then it is hypothesized that so will the emotional state (subject, of course, to cognitive and other contextual factors, Buck, 1976; Schachter & Singer, 1962). A series of experiments are proposed to demonstrate the existence of this change in emotional state as well as to estimate temporal parameters for specific emotion-provoking situations.

Stimuli, such as electric shock (Nisbett & Schachter, 1966), photographs of emotionally-loaded stimuli (Frijda & Van der Geer, 1961) and manipulated situations (Ekman, 1965), have been widely used to evoke emotional responses. However, with the exception of electric shock, these stimuli provide no natural units and therefore do not avail themselves to systematic manipulation of the stimulus variable along the intensive dimension. Even though electric shock has natural units (voltage), its use has been limited due to health and safety concerns. Acoustic noise, on the other hand, has very specific descriptive units, as well as a consistent and well-established history of evoking an affective response called "annoyance." Kryter (1970) defined annoyance as "the subjective impression of the unwantedness of a not unexpected, nonpain or fear provoking sound as part of one's environment."

The method most commonly used in assessing the degree of annoyance produced by noise is the method of self-report. Since measures of self-report are more likely to represent behavior than physiological measures, in that both behavioral responses and measures of self-report are modified by display rules, its use is suggested by many researchers (Hilgard, 1969; Mackay, 1980).

The existence of numerous measurements of noise, notably aircraft noise, indicates the lack of a reliable, predictive measure of annoyance. Noise measurements commonly used in the United States and Great Britain represent descriptive models of annoyance. The parameters of these measurements-- number, duration and/or amplitude (either peak or average)-- correlate well with judged annoyance (Rice, 1977a). This would seem to indicate that the parameters of the noise measurements reflect the major variables of the noise that are producing the annoyance levels. However since most noise measurements have been designed in an effort to predict community response to noise rather than to model the individual annoyance response function, certain, perhaps more subtle, parameters have been overlooked.

In essence, the present models of annoyance production merely correlate the parameters of duration, number and amplitude of the noise events with the community annoyance response, usually to aircraft noise. The dynamic environment, that is how each event in time contributes to the individual's annoyance level, is not considered in these measurements. Although the correlation between the amount of annoyance produced by noise exposure and these measures is good for community responses, the models which underlie the measures for individual judgments are lacking a temporal parameter. The importance of including not only amplitude, number and duration parameters in the annoyance response function, but also a temporal parameter, will be demonstrated in experiments that examine the annoyance response function produced by acoustic noise.

In this section, four models of annoyance will be described and tests of consistency derived. The four models are a peak energy model, an energy averaging model, an energy summation model and an annoyance decay model.

1.1 "Peak" energy model

A simple model of the annoyance response function is one in which the response is represented by the highest energy level present in a stimulus or set of stimuli. An annoyance response is generated solely on the peak energy level presented during a time interval.

Most support for the peak energy model is found in studies on the effect of aircraft noise on community residents. Many investigators have found that the maximum A-weighted sound level (sound weighted to quantitatively reduce the effect of low frequency noise, hence approximating the response of the human ear to sound) occurring during a single aircraft overflight correlates with the noisiness produced by the individual aircraft flyover (Berglund, Berglund & Lindvall, 1975; Young & Peterson, 1969). Rylander proposed that the peak dB(A)— maximum A-weighted sound level— of the noisiest type of aircraft that occurs more than three times a day on an airfield that has more than 63 take-offs in 24 hours provides an adequate measure of the community noise limit (Rylander & Sorenson, 1973; Rylander, Sorenson, Alexandre & Gilbert, 1973; Rylander, Sorenson & Berglund, 1974; Rylander, Sorenson & Kajland, 1972; Sorenson, Berglund & Rylander, 1973). Rylander et. al. found a linear relationship between the percentage of people who were "highly annoyed" and the peak dB(A) of the area. This model does not apply to areas with less than 63 take-offs in a 24 hour period, nor does it provide an explanation as to how the characteristics of as few as 3 flyovers can influence the reported annoyance of community residents unless, of course, the peak dB(A) of the remaining 60 or so overflights correlates with the 3 reference flyovers. It follows from this model that, provided two communities have the same peak dB(A) value, the characteristics of the remaining 60 or so overflights have no effect on the judged annoyance levels of the community members. Other investigators have found this not to be the case (Connor & Patterson, 1976; Kryter & Pearsons, 1963; Powell, 1980). Their results support an energy "averaging" model.

1.2 Energy "averaging" model

An energy "averaging" model which has been shown to predict community annoyance judgments to aircraft noise quite well is the "equivalent sound level" or Leq model (Eldred, 1975; Powell, 1980; Rice, 1977b; Shepherd, 1981). This is the average A-weighted noise level on an energy basis for some specified amount of time. This method equates the numerical value of a fluctuating sound to a steady state sound with the same amount of total energy that continues for some specified period of time (Pearsons & Bennett, 1974).

Since the Leq model does not take into account time of presentation, there is no clear definition as to what constitutes the start of a time period, i.e. whether the time starts at some arbitrary point or whether the time starts when the first noise occurs. This model is not concerned with temporal characteristics and therefore, the time period in which energy is intergrated is usually defined by the researcher as either an arbitrary hour or day in community studies or as the session length in laboratory research. Both of these measures are independent of the time of presentation of noise events.

An energy averaging model predicts that: (1) increasing the maximum intensity of a constant number of aircraft overflights while spreading out their occurrence in time results in constant annoyance; (2) annoyance is constant regardless of the number of aircraft overflights as long as the average energy level over time is constant (two or more successive overflights with equivalent peak dB(A) should result in constant annoyance equal to that of a single overflight as long as the intergration time for the two overflights is twice that of the single overflight); and (3) annoyance should not increase as the duration of a signal increases because the average energy level is unchanged.

Although Shepherd (1981) found Leq to be as good as or better than other noise measures in correlating the physical characteristics of aircraft noise with annoyance level, it has been shown that increasing the maximum intensity of four overflights while spreading out their occurrence in time (so that Leq remains constant) does not result in constant annoyance, increasing the number of overflights while keeping the peak dB(A) level constant does not result in constant annoyance (Galanter & Popper, Note 1), and increasing the duration of a signal does not result in constant annoyance as the Leq model would predict (Kryter & Pearsons, 1963; Molino, 1979). Many investigators have found an energy "summation" model to predict annoyance responses better than either the peak dB(A) or the Leq model.

1.3 Energy "summation" model

Connor and Patterson (1976) have shown that annoyance increases steadily with increases in the total energy level for constant daily aircraft operations and with numbers of operations. This result implies that annoyance judgments are dependent not only on the energy level of the single overflights, but also on the number and duration of events. This result introduces the importance of number on the calculation of annoyance judgments.

A measure of this type is Ls. Ls is the energy summation of some noise level (A-weighted) for some specified period of time. Annoyance accumulates continuously according to Ls. Several investigators have found that the Ls model predicts annoyance responses in laboratory settings better than the Leq model (Balin, Haber & Popper, Note 2; Galanter & Popper, Note 1; Haber & Karsten, Note 3; Perera, Popper & Galanter, Note 4).

It appears that for groups of aircraft overflights occurring over relatively short time intervals, an energy summation model predicts annoyance level judgments better than the peak dB(A) or Leq models. However, none of these proposed models account for temporal variations within a fixed time period. Researchers have found that signals characterized as approaching, i.e. peak amplitude near the end of the signal, are judged more annoying than receding signals assuming constant average and peak amplitude over the signal duration (Haber, Note 5; Nixon, von Gierke & Rosinger, 1969). Peak dB(A), Leq and Ls would predict no difference in annoyance judgments between approaching and receding signals because the overall and peak energy levels are the same for the duration of the signals.

Molino (1979) reviewed experiments in which only rise times were varied. Noises having rise times greater than three seconds were judged to be more annoying than those that reached their maximum level in about one second, even though the total duration and total energy were the same. This result may indicate that when more auditory energy occurs at the end of a time period (long rise time), the annoyance response is greater than when it occurs at the beginning (short rise time) provided the overall energy is the same. The possibility exists that annoyance level may decrease over a period of time, so even though noises with higher peak levels are more annoying than noises with lower peak levels at the time of noise presentation, the annoyance level decays as time between noise events increases and time after noise events progresses.

1.4 Annoyance "decay" model

A decay model of annoyance level assumes that annoyance decreases as time passes. A burst of noise results in extreme annoyance immediately following the event, but hours and certainly days later the annoyance level is greatly reduced.

The decay model consists of an emotion-provoking event and a response function. Suppose the annoyance response system is viewed as a general linear filter. A general linear filter consists of a set of weights (g_r, g_{r+1}, ..., g_s), such that, if the input to the filter is some function (y_t), the output is

$$z_t = \sum_{u=r}^{s} g_u y_{t-u}.$$

If the input is an impulse (a single, non-zero value), the output consists of the weights (Bloomfield, 1976). The weights are also known as the impulse response function. To account for a decrease in annoyance level with the passage of time, as the decay model would predict, the impulse response function should be a decreasing function.

Since other researchers have failed to confirm a decay model through either a *post hoc* (Shepherd, 1981) or an *a priori* design (Powell, Note 6), an explanation of their results is warranted. Shepherd designed his experiment to examine "cumulative annoyance due to multiple aircraft flyovers with differing peak noise levels." He did not set out to examine an annoyance decay model, but considered the annoyance decay model by calculating new Leq values for test sessions by using a range of decay rates of 0 to 20 dB per hour for each session. The correlation coefficients between mean annoyance judgments and these new Leq values did not improve over the correlation between mean annoyance and the unadjusted Leq value. Therefore he concluded that the data did not support the decay hypothesis.

Powell, on the other hand, set out to study the effect of time-of-occurrence on annoyance judgments. His experiment consisted of 1, 2 or 4 high-noise level flyovers occurring at the beginning, middle or end of a 30-minute test session containing a total of eight flyovers. The peak dB(A) of the high- and low-noise level flyovers was, respectively, 79.2 dB and 67.2 dB. The results of an analysis of variance with annoyance responses as the dependent measure indicate that the time-of-occurrence of the high-noise level flyovers was not a significant factor. Although an increase in annoyance is reported when the high-noise level flyovers occur at the end of the session, there is no consistent trend as time between the occurrence of the high-noise level flyovers and the end of the session decreases. Because the experimental design is incomplete, Powell could not test for interaction between time-of-occurrence and number of high-noise level flyovers. Therefore he concludes "the experiment does not provide conclusive information to justify completely discounting any time-of-occurrence effect (Powell, Note 6, p. 7)."

Powell's second experiment tested the effect of session length and number of flyover noises on annoyance judgments. The durations of the sessions that he examined were 15, 30 and 60 minutes and the sessions contained 1, 4 or 16 flyovers evenly distributed in time with a fixed peak noise level of either 79.3 dB or 61.3 dB. It should be noted that when one flyover occurred, Powell placed the occurrence of that one flyover directly in the middle of the session. His results indicate that annoyance decreases as session length increases. This result supports an Leq (energy averaging) model and does not support an [illegible]s (energy summation) model. As in the first experiment because of his experimental design, Powell was unable to examine the interaction between number of flyovers and session duration. Therefore, he is unable to conclude as he was unable to conclude in his first experiment that the annoyance decay model cannot be supported. In fact his results in this experiment do support an annoyance decay model. As session length increases, the time between overflights increases. Therefore, the time in which the annoyance to each overflight decays increases. If the annoyance to each overflight decays as time passes, the annoyance judgments made at the conclusion of the test session should be much less at the end of a 60-minute session than they would be at the end of a 15-minute session. Subjects in these experiments, as is the case in most experiments which test models of emotional response, are asked to attend to the noise either exclusively or while engaged in passive activities. Subjects are usually asked to "relax and read or do any needlework [they] may have brought with [them]" (Powell, 1980, Note 6; Shepherd, 1981). It is hypothesized that this lack of activity does not encourage "reported" annoyance to decay. The request of subjects to attend to noise exclusively implies a sort of accounting procedure where subjects hear a noise, evaluate the noise and store this information for future reference. As we know in real-life as noise occurs around us, we are not usually attending exclusively to the noise. Thus the noise occurs and the event passes without any conscious analysis of the event. Therefore, by engaging subjects in a task, the occurrence of noise and the evaluation of the annoyance produced would simulate the disturbance of noise in a natural setting.

In his research, Borsky engaged subjects in a task of watching a color television program so as to recreate the disturbance of aircraft flyovers in a natural setting (Borsky, 1977; Borsky & Leonard, 1973). So as not to confuse annoyance with auditory masking effects, subjects should participate in a non-auditory task. Reviews on the effects of noise on performance conclude that the majority of published studies indicate that noise is likely to have no effect on performance for vigilance or compensatory tracking tasks (Coates, Adkins & Alluisi, 1975; Coates, Alluisi & Adkins, 1977; Loeb, 1975). Therefore, it is suggested in this research that subjects be asked to judge their degree of annoyance to noise events that occur while they are engaged in a compensatory tracking task.

This research is conducted to test the models of annoyance response to noise and estimate parameters for the response function for individual noise events. In the first experiment, subjects are asked to make annoyance judgments at the end of time periods of various durations in which bursts of white noise are presented. According to the decay model, the placement of these noises in time affects the annoyance judgments of individual subjects. For example, noise events presented at the beginning of a time period are perceived as less annoying than noise events presented at the end of a time period provided that the judgment is asked for at the end of the time period, and the total energy levels during the time periods are the same. The peak dB(A), Leq and Ls models would not support the temporal effect as long as the peak, average and total energy levels were the same. The final experiments are designed to test the form of the decay function for single bursts of noise and to compare the decay constants for bursts of various amplitudes.

The four factors examined here are the session length, the number of noise bursts, the amplitude of the bursts and the time of presentation of these noise bursts. Because these four factors are relevant to the four models presented in the introduction, inclusion of these factors simultaneously in one experimental design provides a test of these models.

This experiment is designed in a way so that the size of the block (number of sessions/subject) can be reduced although that means sacrificing information on certain high-order interactions. By partially confounding rather than completely confounding high-order interactions, information is available on the confounded interactions. Since these effects are confounded, special calculations are required (for a detailed discussion of the experimental design and data analysis, see Appendix A).

An analysis of variance can determine which model best represents the annoyance response function. The peak dB(A) model would be supported if a significant difference existed for amplitude only. Other factors such as session length, number of events and time of presentation have no bearing on response level according to the peak dB(A) model. The energy averaging (Leq) model would be supported if session length, number of events and amplitude were significant factors, but not time of presentation. The energy summation (Ls) model relies only on number and amplitude. However, a significant difference on time of presentation would support the hypothesis that a temporal parameter increases the predictability of response measures of annoyance in a dynamic environment, that is the annoyance decay model.

To verify that indeed no noise effects on task performance exist, some subjects will not experience any noise interference while engaged in the compensatory tracking task. Performance is compared between noise and no-noise experimental conditions to determine what, if any, noise effects on task performance exist. In order to test the hypothesis that exclusive attention to noise encourages subjects to "hold on" to annoyance levels in anticipation to response time, some subjects are asked to make annoyance judgments while sitting quietly in a room reading magazines. Comparisons of annoyance response functions will test this hypothesis.

2.0 Method - Experiment 1

2.1 Subjects

The experimental design for Experiment 1 requires 36 *Ss* (see Appendix A for details). The sessions for two *Ss* were run again using two other *Ss* because the original two *Ss* used "percentage annoyed during session" responses instead of magnitude estimation responses as the instructions stated (their responses were given in terms of percentiles). Also, one *S* had difficulty with the compensatory tracking task, so his sessions were also replaced by another *S*. Most of the *Ss*, 26, were male. All *Ss* were between the ages of 17 and 35. A subset of the 36 *Ss* (Replication 2: nine *Ss*, see Appendix A, Table A2) were called back to run in a "no noise" condition and an additional nine *Ss* ran in a single replication (Replication 2, see Appendix A, Table A2) of the experimental design in a "no task" condition. *Ss* were paid $4.00 for participation. All *Ss* reported normal hearing.

2.2 Stimuli

Ss were presented with six sessions (3, 6 or 9 minutes in length) in which ensembles consisting of 1, 2 or 4 bursts of white noise with peak amplitude of 74, 80 or 86 dB(A) were presented either early (beginning 30 seconds after the start of the game session) or late (ending 10 seconds before response time). Bursts of noise remained at their peak amplitude for 10 seconds (250 msec. rise and decay times). Bursts within an ensemble were separated by a 10-second interstimulus interval measured from the end of one ramp to the start of the next. Slow rise times were used in this experiment so as to reduce the "startle" effect for late presentations. Since the time from the start of the session varied for late sessions making the time of occurrence dependent on session length, this precaution equated the late presentations by eliminating the addition of startle to the annoyance judgment. The noise bursts in the early sessions started 30 seconds after the start of the session as opposed to being presented at the onset of the session. This was done to insure that Ss were actively engaged in the task for presentations of noise in both early and late sessions.

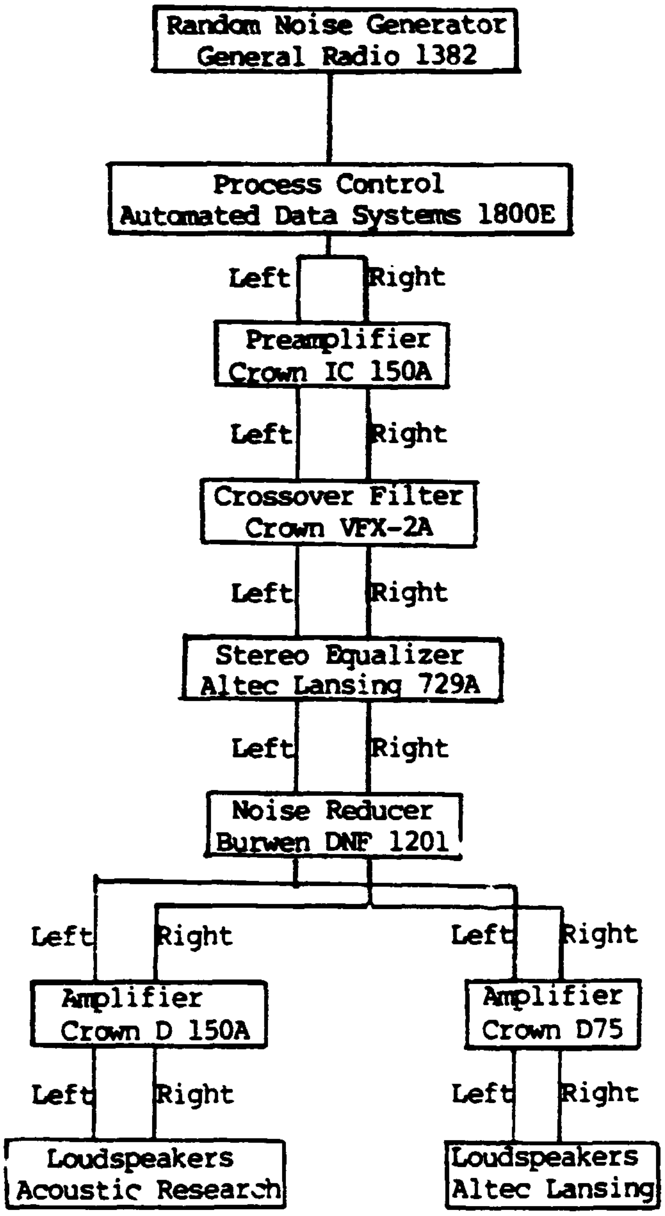
2.3 Apparatus

A block diagram of the audio-control system is presented in Figure 1. White noise was generated by a General Radio 1382 random noise generator and fed to an audio switch. An Automated Data Systems (ADS) 1800E process control system gated the noise and controlled the amplitude of the signal by the use of a programmable attenuator. The system, equipped with an internal clock, also timed the start of presentations of ensembles and the interstimulus interval times. The output was split and fed into a Crown IC 150A preamplifier with calibrated step-attenuator intensity control. The output was then fed into a Crown VFX-2A crossover filter set at lowpass 20 Hz. All frequencies greater than 20 Hz were passed to the loudspeakers. Frequencies less than 20 Hz were passed to a dead-end output. The signal was then fed into an Altec Lansing 729A stereophonic equalizer which was previously calibrated to correct for room acoustics. The equalizer was calibrated by feeding pink noise into the room (70 dB(A)) and adjusting the specific frequencies to result in a "flat" octave band analysis, i.e. equal energy across all bands. The signal was fed into a Burwen DNF 1201 noise suppressor which filters out tape noise. Since the signals were generated at the time of presentation and not recorded on tape, the Burwen was switched "out" of the system. The signal was then split: one-half fed into a Crown D 75 amplifier which drove two Altec Lansing Voice of the Theater speaker systems (22 feet in front of S), the other half was fed into a Crown D 150A amplifier which drove two Acoustic Research 3-A loudspeaker systems (8 feet to the right and left of S).

The compensatory tracking task and the instructions were generated on a Commodore PET 2001 Series computer which included a Commodore CBM Model 2040 dual floppy disk drive. The task was a professionally-supplied game program packaged under the name "Demon game" (for a copy of the computer program, see Jeffries & Fisher, 1981). The program was modified to allow S to control the game from a peripheral device. The video output from the computer was viewed by S on a cathode-ray tube (CRT) placed directly in front of him. An Atari joystick control allowed S to play the game as well as advance the instructions. Communication between S and E during the game sessions was permitted via an audio/visual intercom system. The experiment took place in a 16' X 26' room furnished in the style of a typical middle-class living room (Figure 2).

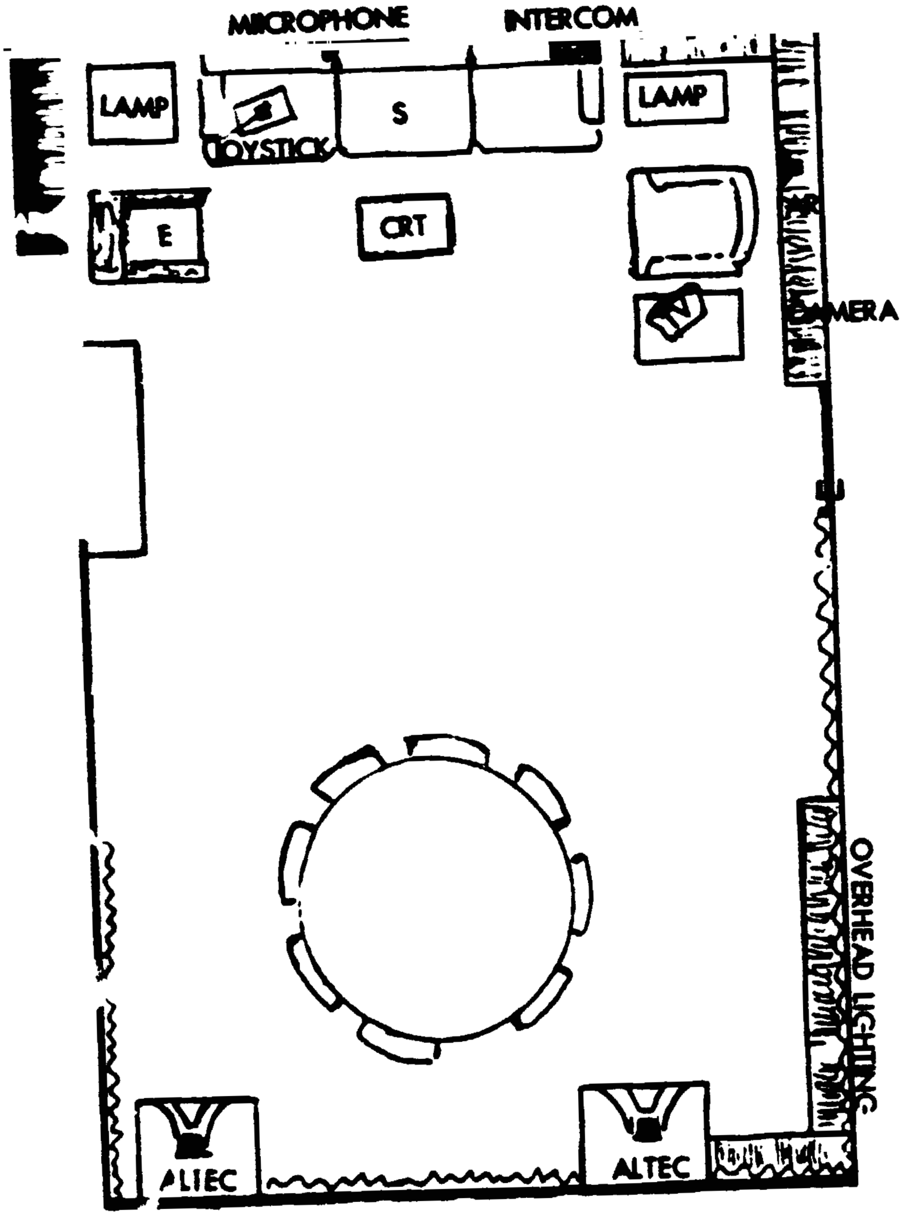
ORIGINAL PAGE IS
OF POOR QUALITY

FIGURE 1


Random Noise Generator
General Radio 1382
Process Control
Automated Data Systems 1800E
Left
Right
Preamplifier
Crown IC 150A
Left
Right
Crossover Filter
Crown VFX-2A
Left
Right
Stereo Equalizer
Altec Lansing 729A
Left
Right
Noise Reducer
Burwen DNF 1201
Left
Right
Left
Right
Amplifier
Crown D 150A
Amplifier
Crown D75
Left
Right
Left
Right
Loudspeakers
Acoustic Research
Loudspeakers
Altec Lansing

ORIGINAL PAGE IS
OF POOR QUALITY

FIGURE 2


MICROPHONE
INTERCOM
LAMP
S
JOYSTICK
LAMP
E
CRT
AIR
TV
CAMERA
OVERHEAD LIGHTING
ALTEC
ALTEC

2.4 Procedure

S was greeted as he entered the laboratory and was led into the simulated living room, seated on the couch and asked to sign the Psychophysics Laboratory Subject Informed Consent Form (Appendix B). E then read the instructions (Appendix C) while S followed along on the video monitor (CRT). Ss were told that the experiment was designed to assess the reaction of people to various noise environments-- how annoying noise is perceived to be and how noise affects work performance. An exercise in the use of the joystick was followed by an abbreviated version of the Demon game. The abbreviated game ended when either five demons were captured or when three minutes expired, whichever came first (always the former). Ss were next informed that the person who captured the most "demons" during the six sessions received a bonus of $25.00. All Ss wrote their names, addresses and telephone numbers on a sheet of paper. This was done in the belief that receiving this information would reassure Ss of Es intent to pay the prize. In fact, E did pay the person who captured the most demons at the conclusion of this series of experiments. Next, the magnitude estimation procedure by which S used to communicate his degree of annoyance was explained (Stevens & Galanter, 1957). S was asked to "rate the annoyance of the ... noises in comparison to the annoyance produced by the standard [noise] ... [and to] be sure that the numbers represent the annoyance produced during each session in proportion to the annoyance produced by the standard ..." E then left the room and presented S with the standard stimulus (one 10-second burst of white noise at 80 dB(A)) followed by six game sessions. S controlled the start of a game session by pressing a button on the Atari joystick. E determined when the session was over (either 3, 6 or 9 minutes) and caused the game to terminate. The next session started with a new game when S pressed the button on the Atari joystick. At the end of each game session, S was asked to verbally give his annoyance response to E. At the end of the six game sessions, a short questionnaire (Appendix D) was presented and completed by S. An experimental session lasted one hour. S was paid a participation fee of $4.00, debriefed and dismissed.

The procedure for the no-noise condition was identical to the procedure for the previous experiment except that instructions regarding magnitude estimation were omitted and no judgments on noise were made. The standard noise burst as well as noise during sessions were omitted. The questionnaire contained only items that did not refer to noise.

The procedure for the no-task condition was the same as the procedure for the general experiment. E read the instructions while S followed along on the CRT. No mention was made of the task. The magnitude estimation procedure was explained. S was told that he could read magazines that were left for him if he wished (current issues of *Newsweek*, *People* and *U.S. News*

& World Report). Noise was then presented as in the task condition. Items in the questionnaire referring to the game were omitted.

3.0 Results and Discussion - Experiment 1

The main effects in this experiment were of primary interest. A significant effect solely on amplitude would support the peak dB(A) model; significance on number and amplitude would support the Ls model; and significance on session length, number and amplitude would support the Leq model. However, only the decay model would predict a significant difference on time of presentation.

A logarithmic transformation was performed on the raw data responses because error variances around large numerical responses are greater than those around small numerical responses. A logarithmic transformation results in more equivalent error variances.

The geometric mean responses for all main effects and interactions are presented in Table 1. The means of confounded interactions are pooled across blocks in this table. The results of an analysis of variance (Table 2) support the hypothesis that number, amplitude and time of presentation are important parameters in the determination of annoyance level ($p<.01$). Peak dB(A) is the most important factor in determining annoyance level (Figure 3). The straight lines in the figures are the result of a regression analysis using the geometric means in Table 1 as the dependent variable. In addition, the number of bursts (Figure 4) and the time of presentation are also important parameters, but to a lesser degree. Bursts of noise presented early in a session were reported to be much less annoying than bursts presented late in a session ($p<.01$) when subjects were asked to respond at the end of a session. This result supports a decay model of annoyance level.

If an annoyance level decay process was to be supported, one would expect that the interaction between session length and time of presentation would be a significant one. That is, the longer the session length, the smaller the response should be for early signals (the more time for the annoyance level to decay), whereas session length should have no effect on late signals. The results indicate that annoyance responses to bursts of noise presented late in a session remain fairly constant as session length increases (Figure 5, solid line); whereas the annoyance to bursts presented early decay as the length of the session increases (dotted line). In fact, the decay for annoyance responses to early presentations accounts for the session length by amplitude interaction. However, as can be seen in Figure 6 (interaction between session length, amplitude and time of presentation), annoyance responses for bursts with amplitudes of 86 dB(A) and 80 dB(A) decay with time, whereas annoyance responses for bursts with amplitudes of 74 dB(A) do not appear to decay.

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 1
GEOMETRIC MEANS OF ANNOYANCE RESPONSES FOR MAIN EFFECTS AND INTERACTIONS FOR EXPERIMENT 1

Session Length (A)		Number (B)		Amplitude (C)		Presentation Time (D)	
3 minutes:	98	1:	84	74 dB(A):	57	Early:	84
6 :	85	2:	83	80 :	86	Late :	101
9 :	92	4:	110	86 :	157		

A X B*

	3	6	9
1:	102	68	86
2:	90	84	76
4:	102	107	122

A X C*

	3	6	9
74:	55	51	67
80:	92	80	87
86:	186	150	138

A X D

	3	6	9
E:	98	80	75
L:	98	91	115

B X C*

	1	2	4
74:	55	47	73
80:	87	80	92
86:	126	154	198

B X D

	1	2	4
E:	77	70	110
L:	92	100	111

C X D

	74	80	86
E:	55	82	131
L:	60	91	187

A X B X D*

	3		6		9	
	E	L	E	L	E	L
1:	99	105	65	72	71	104
2:	97	84	66	107	53	109
4:	99	106	120	96	112	134

A X C X D*

	3		6			
	E	L	E	L	E	L
74:	52	59	52	51	63	71
80:	102	83	80	81	67	112
86:	181	190	124	181	100	191

B X C X D*

	1		2		4	
	E	L	E	L	E	L
74:	50	60	45	50	75	72
80:	87	86	63	102	99	86
86:	103	153	122	194	178	221

* means pooled with block effect

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 2
ANALYSIS OF VARIANCE OBTAINED USING A LOGARITHMIC TRANSFORMATION OF ANNOYANCE RESPONSES FOR EXPERIMENT 1

SOURCE OF VARIATION	SS	DF	MS	F
Between Blocks	5.5088	35		
Replications	0.2213	3	0.0738	0.44
Blocks w/in Reps	5.2875	32	0.1652	
Within Blocks	19.3510	180		
A(Session Length)	0.1373	2	0.0687	2.54
B(Number)**	0.6824	2	0.3412	12.65**
C(Amplitude)**	6.9030	2	3.4515	127.99**
D(Time)**	0.3377	1	0.3377	12.52**
A X B	0.1860	4	0.0465	1.72
A X C**	0.3845	4	0.0961	3.56**
A X D**	0.3329	2	0.1665	6.17**
B X C	0.0498	4	0.0125	0.46
B X D*	0.1994	2	0.0997	3.69*
C X D	0.1622	2	0.0811	3.00
A X B X D**	2.4097	4	0.6024	22.33**
A X C X D**	1.7723	4	0.4431	16.43**
B X C X D**	1.9008	4	0.4752	17.62**
A X B X C X D	0.2525	8	0.0316	1.17
Error	3.6405	135	0.0270	

** indicates significance at 0.01 level
* indicates significance at 0.05 level

The significant interactions in Table 2 that pertain to number of bursts reflect the similiarity between one and two bursts and their contrast with four bursts. The effect of decay on number of noise events is most easily seen in the interaction between session length, number and time of presentation (Figure 7). The slopes of the best-fitting lines for late bursts as a function of session length (solid lines) are approximately zero for one, two and four bursts. Annoyance responses decay for one and two bursts presented early as was predicted by the decay model, but the annoyance produced by four bursts does not decay at all (dotted lines). The interaction between number and time and the interaction between number, amplitude and time support this result as well. It appears that either four bursts leave a very strong impression and do not decay at all or four bursts decay less quickly than one or two noise bursts. The case may be that the time between the end of the noise ensemble and the response time in the nine minute sessions (approximately seven minutes) is still not sufficiently long to result in a decay of reported annoyance, just as there is no significant difference between early and late presentations in the three minute sessions.

In order to determine whether the presence of a task had any effect on annoyance response and whether the task prevented the subject from rehearsing his response and thereby basing his judgment on his emotional response at response time, an additional group of nine subjects ran in a single replication of the experimental design (Replication 2) in a no-task condition. One of the questions in the questionnaire answered by all subjects asked them to recall the number of noise events in each session. A t-test was performed on the absolute error totals for subjects in the task and no-task conditions and was not found to be significant. Although the memory for the events in both conditions was not significantly different, the annoyance responses as a function of delay to response time are different. As was hypothesized, the annoyance response function does not decay for bursts of noise presented in the no-task condition. Figures 8 and 9 represent the interactions between session length and time of presentation for the task and no-task conditions, respectively. Figure 8 is based on nine of the subjects who ran in Replication 2 in the task condition (the appropriate subset of the 36 subjects who ran in the full Experiment 1 task-noise condition). As can be seen in Figure 8, the annoyance responses for bursts in the task condition presented early decay as session length increases (dotted line). The annoyance responses for late bursts remain relatively constant. On the other hand for subjects who were not involved in a task, as Figure 9 demonstrates, annoyance responses do not decay. The slopes in the task and no task condition for the early presentations are significantly different from each other ($p<.05$). In conclusion, the reason why researchers have not found decay effects in the past may have been that they were not optimizing the decay process by introducing the presence of a task. This means that any value of the decay parameter that one calculates is context specific.

As was discussed in the introduction, since the Leq model does not take into account time of presentation, there is no clear definition as to what constitutes the start of a session, i.e. whether the time starts when the trial begins regardless of when the first noise occurs or whether the time starts when the first noise occurs. If the latter is the case, then one could argue that the results of the analysis of variance support this version of the Leq model because sessions in which the noise burst occurred late in the session would have a higher Leq value than those in which the noise burst occurred early. If Leq were calculated using the start of the first burst in this experiment as the start of the session, then this version of Leq would predict that annoyance responses to the late bursts would be much more annoying than responses made for the early bursts even in the three minute sessions. Using the three minute session as an example, in the early sessions energy would be averaged over three minutes, whereas in the late sessions energy would be averaged over anywhere from 20 to 80 seconds. Therefore, the Leq value for the early sessions would be much lower than the Leq value for the late sessions (for single bursts there is a nine dB difference). The results of this experiment do not support this definition of Leq. This can be seen in the A X B X D interaction (see Table 1). The geometric mean response for the single early burst presented during the three-minute session is 99, whereas the geometric mean response for the late single burst is 105. Since the Leq values for the early and late session differ by 9 dB, one would expect a doubling of annoyance between the early and late sessions. This is not the case in this experiment.

In order to determine whether the interest level or past experience with video games had any effect on annoyance judgments, correlations between the geometric mean of annoyance response for each subject and the amount of time subjects reported they spent playing video games in the past month and the interest level of the Demon game were calculated. The results indicate no significant correlation between amount of time spent playing video games in the past month and annoyance responses nor between the interest level of the Demon game and the annoyance responses. These results would indicate that the range of interest levels in the game had no significant effect on annoyance judgments. However, there was a significant negative correlation between the amount of time spent playing video games in the past month and the interest level of the Demon game ($p<.01$). This result is not surprising. Those subjects who played video games reported that they played rather sophisticated ones and found this game to be of less interest to them. However, those that never played found the Demon game to be more interesting. The reason for choosing the Demon game over other games that were available was that familiarity with video games seemed to have little effect on the ability to play this game. The results of a correlation between the number of demons caught and the amount of time subjects reported they spent playing video games is significant at the .01 level. Though this result implies a strong correlation between number of demons caught and time spent playing other video games, in fact the only subjects whose responses supported this effect were those who spent a great deal of time playing video games (they captured many demons) and those who never had played any other video games (they captured the least number of demons). However, the vast majority of subjects in this experiment were somewhere in the middle (n=31) in which the effect of prior exposure to video games is minimal. Since there is no significant correlation between number of demons caught and annoyance response, the Demon game appears to be a suitable background task for the study of temporal effects on annoyance response.

To confirm that noise did not affect task performance, nine subjects were called back to run in a no-noise condition. An analysis of variance on number of demons caught resulted in session length as the only significant factor (Table 3). That is, given more time, subjects capture more demons. The presence of noise in this experimental paradigm had no significant effect on task performance.

Though the results of Experiment 1 support a decay model of annoyance level as a function of time to response, several adjustments to the experimental procedure are suggested before tests of the form of the function may proceed. First, the instructions regarding the annoyance response appear to be ambiguous. Though they replicate the instructions in Powell's and Shepherd's experiments, the annoyance responses made by subjects were probably one of two types. One is retrospective in nature and the other is truly based on the subject's current emotional response. During short time periods, the memory for auditory events is rather good, and hence subjects can respond with their annoyance level at the time of the noise bursts. Therefore, late presentations are always analyzed at response time and memory trace of the noise event is excellent. However, as time between event and response time increases, other types of annoyance judgments may be made. Some subjects may respond with their current, decayed emotional response (very low) while others might respond with either their recollection of their emotional response following the event or their memory of the sensory event and their translation of that event into an emotional response (much higher). In the latter case, subjects are saying, "If I heard a noise that loud, I would be very annoyed." Therefore, he reports a high level of annoyance. In fact, this experiment relies on these distinctions between annoyance responses, as do all experiments which use magnitude estimation procedures. Subjects are asked to compare their current annoyance levels with the standard annoyance level, an annoyance level they felt in the past. Their recollection of the emotional response following the emotion-provoking event is excellent. Otherwise, their annoyance responses made to the sixth session would be significantly higher than their responses made to the first session. Afterall, the annoyance produced by the standard has decayed significantly by the end of the experiment. The annoyance responses made by subjects indicate that their

TABLE 3
ANALYSIS OF VARIANCE OBTAINED USING NUMBER OF DEMONS CAUGHT AS DEPENDENT MEASURE, NOISE VERSUS NO-NOISE CONDITIONS

SOURCE OF VARIATION	SS	DF	MS	F
A(Noise Condition)	15.5740	1	15.5740	0.25
S(Subjects)	4871.3703	8	608.9213	
A X S	484.2593	8	60.5324	
B(Session Length)	54868.4815	2	27434.2408	2675.43**
B X S	1702.1852	16	10.2541	
A X B	43.8148	2	21.9074	0.44
A X B X S	790.8519	16	49.4282	

** indicates significance at 0.01 level

recollection of the standard annoyance does not decay significantly with time. Therefore, one would expect responses to vary greatly in conditions where the time between event and response time is large because the distinction between these two types of responses is greater. The results of a Wilcoxon Matched Pairs Signed Ranks Test for the variances of responses for each experimental cell indicate that, indeed, there exists a significant difference between variances of responses made for early presentations and variances of responses made for late ones ($p<.05$), where variances are greater for early presentations. Since an explanation has been found to account for the non-significant results of Powell's and Shepherd's experiments, i.e. presence of task, more specific instructions regarding the time of the analysis of response are included in Experiment 2. In order to focus on current annoyance response as opposed to remembered annoyance level, subjects are asked to respond to the annoyance level they feel at the time of the response, not the annoyance level they felt at the time of the noise presentation.

Since this study is comparative and not absolute, it does not set criterion for predicting annoyance judgments. The study establishes which parameters are suitable to include in criteria. Since the decay parameter is specific to the presence or absence of a task, the parameter is estimated using a specific context, i.e. the Demon game. The generalizability of the estimate of this parameter remains to be seen, but this context maximizes the decay parameter so that estimates may be made over relatively short time periods. Only the parameters of amplitude and decay time are examined in the next experiment. By adding other decay times to the annoyance function, the form of the annoyance response function may be tested. By viewing the burst of noise as an impulse, the responses to bursts of noise after a period of time would represent the impulse response function. Since the rate of decay may be dependent on amplitude and may be "bottoming out" at 74 dB(A), several amplitudes greater than 86 dB(A) are introduced. The standard noise should have an amplitude in the center of the range so as to avoid "range effects" (Poulton, 1968). Poulton discovered that when the physical magnitude of the standard is near the lower end of the range, values smaller than the standard give a steeper slope than values larger than the standard. By placing the standard in the center of the range, this effect diminishes.

Since session length had no significant effect on annoyance response in the range of three to nine minutes, session length in Experiment 2 will be confounded with decay time. All noise bursts will start after 50 seconds of game time and last for 10 seconds. The session length will be equal to the decay time plus one minute (the time needed to present the burst of noise). Since experimental design considerations limit the study of number of noise bursts at this time, the remaining studies are only concerned with the annoyance response function for single bursts of noise.

ORIGINAL PAGE IS
OF POOR QUALITY

4.0 Method - Experiment 2

4.1 Subjects

Nine Ss were recruited by placing posters in and around the Columbia University campus. Most of the Ss, 8, were male. Ss were paid $20.00 for participation— $4.00 per hour for a total of 5 hours. All Ss reported normal hearing.

4.2 Stimuli

The time between presentation and response, and the amplitude of a 10-second noise burst within a game session varied in a 9X5 factorial experiment with repeated measures on all factors. All noise bursts occurred 50 seconds after the start of the game session. The time between the conclusion of the burst and the end of the game session was 0, 1, 2, 3, 4, 5, 6, 7 or 8 minutes; and the peak amplitude of the burst was 75, 80, 85, 90 or 95 dB(A). In this way, several decay times and peak amplitudes were analyzed in a within-subject design which in turn reduces error variances. Because the number of days a subject is required to participate in this design is five, it is not necessary to use a confounded design as was the case in Experiment 1.

Because the time at which the noise burst starts with regard to the start of the game session is constant for all trials, the noise bursts need not be ramped on and off. These bursts of noise may now be viewed as an impulse of noise rather than the sum of impulses which ramp on in 250 msec. to the peak and ramp off in 250 msec. after 10 seconds. The resulting annoyance function may be viewed as the impulse response function as discussed in the introduction.

Also, since it is no longer necessary to replicate the procedures of Powell and Shepherd, the standard noise was presented within a game session. The standard was a 10-second burst of white noise with peak amplitude of 85 dB(A) which occurred 50 seconds from the start of a 1-minute game session. This change in procedure makes judgments between the standard and the test sessions more comparable. Both judgments are now being made in the same context.

4.3 Apparatus

The audio-control system is represented by the block diagram of Figure 1 and described in the Method section of Experiment 1.

4.4 Procedure

S was greeted as he entered the laborato [illegible] led into the

ORIGINAL PAGE IS
OF POOR QUALITY

simulated living room, seated on the couch and asked to sign the Psychophysics Laboratory Subject Informed Consent Form (Appendix E). Changes were necessary in the Subject Informed Consent Form used in Experiment 1 (Appendix B). The terminology "For your protection..." was ambiguous enough so that many Ss questioned what they were being protected from. For Experiment 2, the reference was omitted and S was given the form in Appendix E to sign.

The major difference in procedure between Experiments 1 and 2 is the instructions regarding annoyance response. Emphasis is made in Experiment 2 on the annoyance felt at the time that the response is reported, not on the annoyance felt at the time of the noise events. The instructions are contained in Appendices F and G.

An experimental session lasted approximately one hour. The 45 experimental conditions were randomly assigned. The first experimental session consisted of a practice game session followed by the standard game session and several test game sessions totalling between 33 and 40 minutes of game time. The practice game session was eliminated for the remaining four days. The remaining experimental sessions consisted of the standard and test sessions totalling between 42 and 53 minutes of game time. S was paid a participation fee of $20.00 at the end of the fifth day. After S was paid, E asked S to answer the question: "If the standard annoyance is '100,' what number represents the annoyance level at which any number you gave above that number you would label 'annoying' and any number you gave below that number you would label 'not annoying.'" If S had difficulty understanding this question, E repeated the question slowly (usually with hand-motions) until E was satisfied that S understood that E was asking S for his threshold of annoyance.

5.0 Results and Discussion - Experiment 2

The results of Experiment 2 indicate that indeed the annoyance response function is a decreasing function (Figure 10). The geometric means of the individual annoyance response functions for each amplitude were fitted by means of a standard linear regression analysis to a linear decay function:

$$\text{Annoyance Level} = A_0(1-at),$$

where A_0 is dependent on the initial amplitude and is the value of the function at $t=0$ (zero decay time) and a is the linear rate of decay (Figure 10, dotted lines). The linear annoyance decay function has this form because initial perusal indicated that a was proportional to the initial amplitude A_0. The calculated values of the two parameters and the r^2-values for the linear regressions are presented in Table 4.

Although the r^2-values are significantly different from zero ($p<.01$)

ORIGINAL PAGE IS
OF POOR QUALITY

FIGURE 10A -- 75 DBA

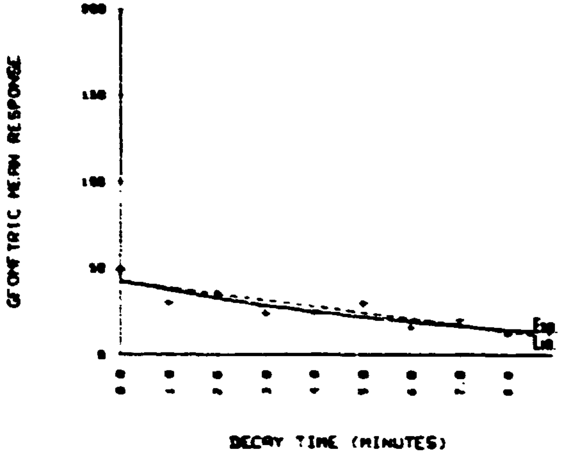


FIGURE 10B -- 80 DBA

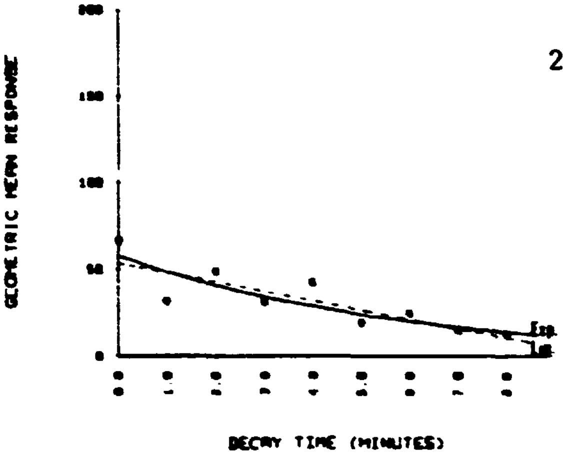


FIGURE 10C -- 85 DBA

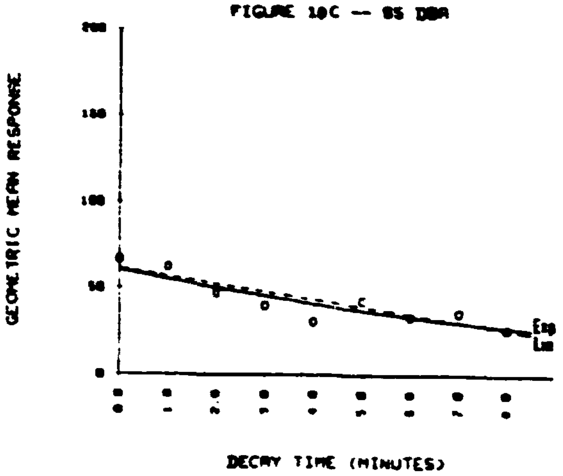


FIGURE 10D -- 90 DBA

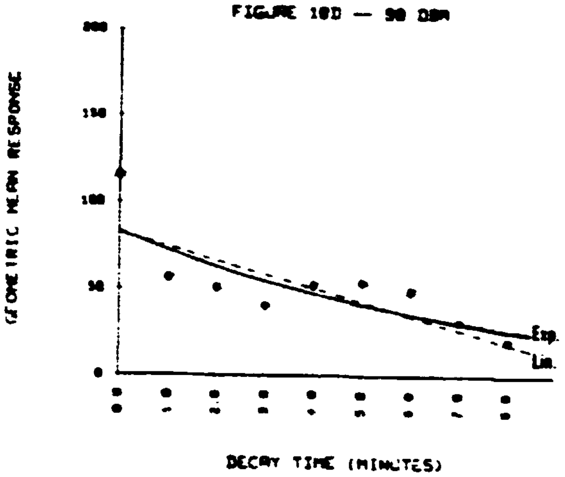
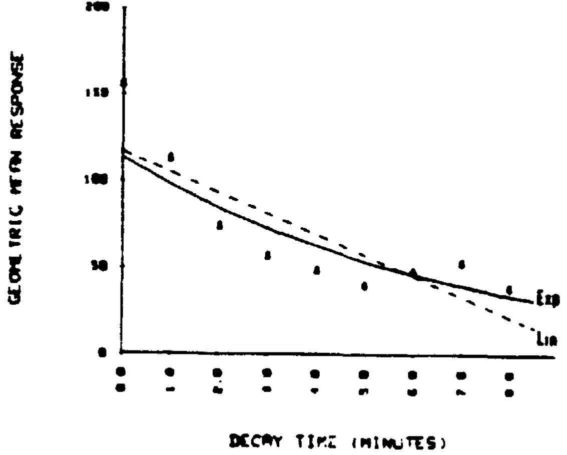


FIGURE 10E -- 95 DBA



ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 4
COMPARISON BETWEEN A LINEAR DECAY MODEL AND AN EXPONENTIAL DECAY MODEL FOR EXPERIMENT 2

Linear Decay Model: Annoyance Response = $A_0(1-at)$

	r^2	a	A_0
75 dB(A)	.747649	11.625235	41.153333
80 dB(A)	.715724	10.038335	54.088222
85 dB(A)	.767034	13.805311	60.391333
90 dB(A)	.586764	11.097884	82.098445
95 dB(A)	.680984	9.910625	117.467333

Exponential Decay Model: Annoyance Response = $A_0 e^{at}$

	r^2	a	A_0
75 dB(A)	.788989	-.135082	42.952180
80 dB(A)	.786966	-.175912	57.836723
85 dB(A)	.778537	-.098288	60.908009
90 dB(A)	.677366	-.139758	83.273811
95 dB(A)	.747910	-.148695	113.676275

indicating that the slope of the function is significantly different from zero, the annoyance responses made between zero and three minutes after presentation appear to decay more rapidly than this model predicts (see Figures 10D and E). An alternative model which could explain this change in the rate of decay is the exponential decay function:

$$\text{Annoyance Level} = A_0 \, e^{at},$$

where A_0 is dependent on the initial amplitude and is the value of the function at t=0 and a is the decay constant. The geometric mean responses are fitted to the exponential decay model in Figure 10 (solid lines). If the exponential decay model provides a better fit to the data than does the linear decay model, this analysis would result in larger r^2-values than when a regression to the linear decay model is performed. The response functions for this model are also presented in Table 4. The r^2-values for these functions indicate that the exponential decay model is a better fit to the data than is the linear decay model.

However, the decay constants of the exponential function calculated from this experiment do not appear to be systematic. Furthermore, the annoyance responses in the first three minutes of decay times seem to decay faster than the calculated exponential functions predict. The subjects were asked at the conclusion of this experiment to define their "threshold" value. The number of subjects whose annoyance response was at or below his threshold value is shown in Table 5. Notice that after three minutes, particularly in the lower amplitude conditions, the majority of subjects were at their threshold values. Subjects reported that, in many cases, when their response was below their threshold value, they would have said "zero" if it were permissable. They found it very difficult to define a response proportional to an annoying standard when they were not annoyed. If only the first three minutes of decay times are considered, a consistent pattern forms (Table 6). The functions for the amplitudes between 75 and 85 dB(A) have decay constants between -.180 and -.195. At 90 and 95 dB(A) the decay constants increase sharply to between -.325 and -.343. The parameters for the linear decay model are also included in Table 6 for the first three minutes of decay times. The exponential decay function predicts the results better than the linear decay function for three of the five amplitudes (75, 90 and 95 dB(A)). For 85 dB(A) both functions fit the data well and for 80 dB(A), neither function predicts the data well.

Before conclusions can be made regarding the form of the decay function from the values of the parameters and the least-squares estimates, it is suggested that an additional experiment be performed that focuses only on the first three minutes of decay. Afterall, the conclusions that can be made in this experiment are based on only four data points per amplitude. It is suggested that these four values be verified in addition to including other decay times within the first three minutes of decay in a supplementary

TABLE 5
NUMBER OF SUBJECTS WHOSE ANNOYANCE RESPONSE WAS AT OR BELOW HIS THRESHOLD VALUE FOR EXPERIMENT 2 (N = 9)

Decay Times:	0	1	2	3	4	5	6	7	8	minutes
75 dB(A)	2	5	4	4	6	6	6	7	7	
80 dB(A)	1	4	4	4	3	7	4	6	6	
85 dB(A)	2	2	3	4	5	5	5	4	4	
90 dB(A)	0	2	3	5	4	2	4	4	7	
95 dB(A)	0	0	2	2	3	3	4	3	4	

TABLE 6
COMPARISON BETWEEN A LINEAR DECAY MODEL AND AN EXPONENTIAL DECAY MODEL FOR EXPERIMENT 2 (DECAY RATES BETWEEN 0 AND 3 MINUTES ONLY)

Linear Decay Model: Annoyance Response = $A_0(1-at)$

	r^2	a	A_0
75 dB(A)	.688101	6.477567	45.479000
80 dB(A)	.462891	6.540437	57.824000
85 dB(A)	.958233	7.183345	68.249000
90 dB(A)	.780964	4.357328	100.602000
95 dB(A)	.969121	4.483495	143.712000

Exponential Decay Model: Annoyance Response = $A_0 e^{at}$

	r^2	a	A_0
75 dB(A)	.696934	-.195885	45.266659
80 dB(A)	.416390	-.133099	55.711937
85 dB(A)	.955532	-.180655	69.353365
90 dB(A)	.863263	-.325170	98.842602
95 dB(A)	.993283	-.343681	155.499704

experiment.

6.0 Method - Experiment 3

6.1 Subjects

Five of the nine Ss who participated in Experiment 2 continued in this experiment along with five additional Ss. Most of the Ss, 9, were male. Ss were paid $10.00 for participation-- $5.00 per hour for a total of 2 hours. All Ss reported normal hearing.

6.2 Stimuli

The time between presentation and response and the amplitude of a noise burst within a game session varied in a 7X5 factorial experiment with repeated measures on all factors. The time between the conclusion of a 10-second noise burst and the response is 0, .5, 1, 1.5, 2, 2.5 or 3 minutes; and the peak amplitude of the noise burst is 75, 80, 85, 90 or 95 dB(A). The noise burst occurs 50 seconds after the start of the game session. As in Experiment 2, the burst of noise is not ramped on and off. The standard is a 10-second burst of white noise with peak amplitude of 85 dB(A) which occurs 50 seconds after the start of a 1-minute game session.

6.3 Apparatus

The audio-control system is represented by the block diagram of Figure 1 and described in the Method section of Experiment 1.

6.4 Procedure

The procedure is the same as Experiment 2 and described in the Method section of Experiment 2.

An experimental session lasted approximately one hour. The 35 experimental conditions were randomly assigned. Since the game sessions were short, more sessions were run per experimental hour than in either Experiment 1 or 2. The first experimental session consisted of a practice game session followed by the standard and several game sessions totalling between 37.5 and 40 minutes of game time. The practice game session was eliminated for the second day. The experimental session consisted of the standard and game sessions totalling between 47.5 and 50 minutes of game time. S was paid a participation fee of $10.00 at the end of the second day.

7.0 Results and Discussion - Experiment 3

The results of Experiment 3 concur with the results of the prior

ORIGINAL PAGE IS
OF POOR QUALITY

experiments. The annoyance response function is a decreasing function (Figure 11). The annoyance response means were fitted to a linear decay function and an exponential decay function as was done in Experiment 2. Results indicate that for all amplitudes except 80 dB(A) the exponential decay function (solid lines) provides a better fit to the data than does the linear decay function (dotted lines) as an annoyance response function (Table 7). These results for 80 dB(A) seem to be an anomoly because of the last two data points in the function (2.5 and 3.0 minute decay times). Indeed when the regression analysis is performed on the response function for 80 dB(A) without the last two data points, the exponential decay function provides a better fit to the data than does the linear decay function. The results shown in Table 7 indicate that the three higher amplitudes have similar decay constants whereas the two lower amplitudes have a smaller decay constant. In fact, a test of the significance between the decay constants (actually the significance between the two slopes resulting from the linear regression of the log geometric means) indicates that, at the .10 level, the decay constants for the response functions of 75 and 80 dB(A) do not differ significantly from each other, nor do the decay constants for 85, 90 and 95 dB(A). However, the decay constants for 75 and 80 dB(A) differ significantly from the decay constants for 85, 90 and 95 dB(A). Therefore, it can be concluded that the rate of decay for the higher amplitude bursts is significantly greater than for the lower amplitudes.

The results of this experiment indicate that the function that provides a good description of the annoyance response function is:

$$\text{Annoyance Response} = A_0\, e^{at}.$$

Though the value of the decay constant alpha is dependent on the units of time used in its calculation, the decay constant has no immediately obvious meaning in that its value is unitless. Therefore the half-life is introduced to add meaning to the decay constant. This is the length of time, Δt, in which the annoyance response is reduced by 50%. To relate Δt with alpha, consider two time instances, t_1 and $t_2 = t_1 + \Delta t$ (see Figure 12) and let A_1 and A_2 denote the annoyance response at these time instances, respectively. From the formula for the exponential decay function:

$$A_1 = A_0\, e^{at_1}, \text{ and}$$
$$A_2 = A_0\, e^{at_2} = A_0\, e^{a(t_1+\Delta t)} = A_0\, e^{at_1} * e^{a\Delta t}, \text{ or}$$
$$A_2 = A_1\, e^{a\Delta t}.$$

Since the definition of half-life assumes that only half of the annoyance response remains after the time interval, $A_2 = 1/2\, A_1$. This yields:

$$e^{a\Delta t} = 1/2,$$
$$a\Delta t = \ln 1/2 = -0.69315.$$

**ORIGINAL PAGE IS
OF POOR QUALITY**

FIGURE 11A -- 75 DBA

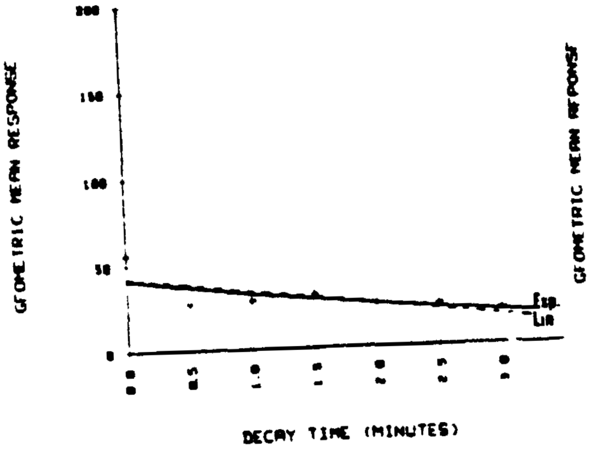


FIGURE 11B -- 80 DBA

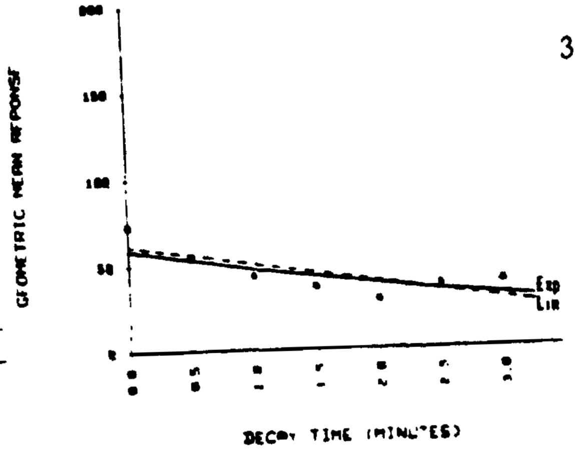


FIGURE 11C -- 85 DBA

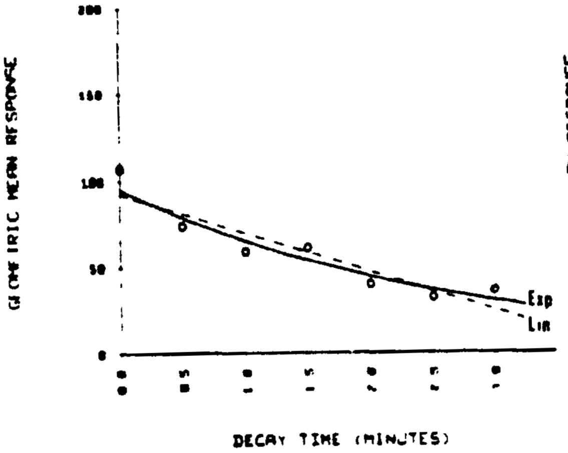


FIGURE 11D -- 90 DBA

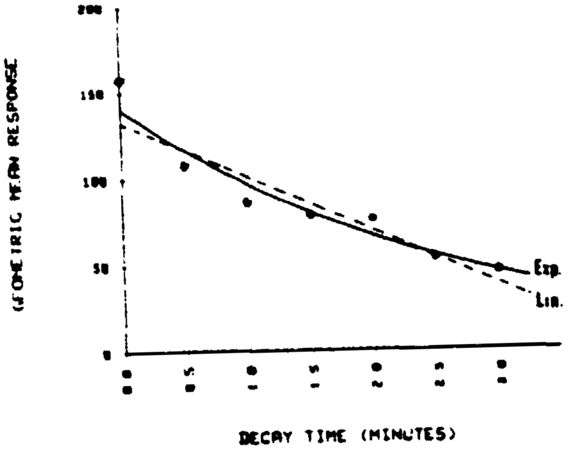


FIGURE 11E -- 95 DBA

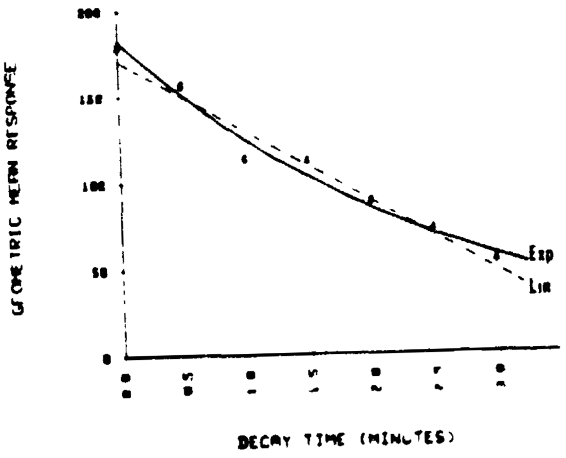


TABLE 7

COMPARISON BETWEEN A LINEAR DECAY MODEL AND AN EXPONENTIAL DECAY MODEL FOR EXPERIMENT 3

Linear Decay Model: Annoyance Response = $A_0(1-at)$

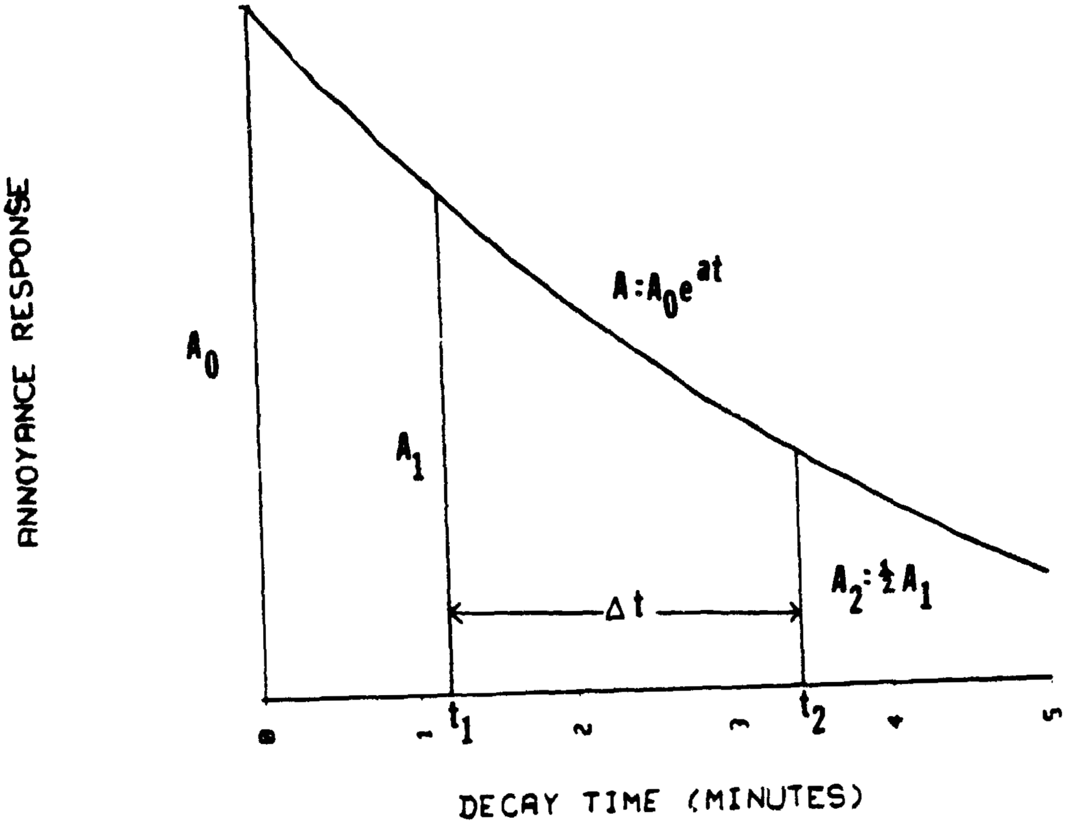
	r^2	a	A_0
75 dB(A)	.590825	5.037318	42.615714
80 dB(A)	.626572	5.617140	60.556786
85 dB(A)	.861182	4.100432	92.271429
90 dB(A)	.874364	4.166917	134.939643
95 dB(A)	.964670	4.228163	171.093214

Exponential Decay Model: Annoyance Response = $A_0 e^{at}$

	r^2	a	A_0
75 dB(A)	.683987	-.251139	41.396634
80 dB(A)	.603558	-.218984	59.118162
85 dB(A)	.912166	-.378814	95.301076
90 dB(A)	.949757	-.374055	140.420170
95 dB(A)	.977339	-.384411	182.667329

ORIGINAL PAGE IS
OF POOR QUALITY

FIGURE 12



Hence, the half-life is

$$\Delta t = 0.69315/-a.$$

Therefore whenever a time interval of length Δt elapses, the annoyance response is reduced by one-half. Since the half-life is a linear transform of the decay constant, conclusions made regarding the decay constants in Experiments 2 and 3 are also true for half-life estimates, i.e., the half-life for the higher amplitude bursts is significantly shorter than for the lower amplitude bursts (Table 8). The results for 0 to 3 minute decay times for Experiments 2 and 3 indicate that the higher amplitudes have a half-life between 1.8 and 2.1 minutes and the lower amplitudes have a half-life between 2.7 and 3.8 minutes.

8.0 General Discussion

This series of experiments demonstrates that annoyance decays as time progresses from the occurrence of a noise burst and that the form of the decay is best approximated as an exponential. The purpose of this research was to define which parameters were appropriate for defining an annoyance response function. Once the important parameters of a model were defined, an examination of the value of those parameters for specific stimuli within a particular context were calculated. Past research indicated that number, duration and amplitude were three important parameters in the determination of community annoyance response to aircraft noise. Usually these parameters are examined by researchers in either a community setting in which residents are asked to categorize their annoyance produced by noise exposure or in a laboratory setting in which subjects are asked to categorize or magnitude estimate their annoyance. The design of these experiments fail to take into account the dynamic environment, that is, how each noise event contributes to the individual's annoyance level. The models most commonly proposed by researchers lack a temporal parameter. Before an examination of the annoyance response function for individual noise events could proceed, it was necessary to determine whether the inclusion of a temporal parameter in the annoyance response function would significantly reduce the amount of variance about this function. When other attempts had been made to examine the effect of presentation time on annoyance response, researchers failed to confirm any systematic time-effects (Powell, Note 6; Shepherd, 1981). Therefore, it was first necessary to determine why other researchers had not found what intuitively seemed to be an important parameter in the determination of annoyance response.

An examination of the techniques used by Powell and Shepherd in their research indicated that subjects in these experiments, as is the case in most experiments which test models of emotional response, were asked to attend to the emotion-provoking stimuli exclusively or while engaged in passive activities. It was hypothesized that this lack of activity did not

TABLE 8
HALF-LIFE ESTIMATES OF ANNOYANCE RESPONSE BASED ON THE DECAY CONSTANTS CALCULATED FOR EXPERIMENTS 2 AND 3
(0 - 3 MINUTE DECAY TIMES ONLY)

	Experiment 2 (Minutes)	Experiment 3 (Minutes)
75 dB(A)	3.53856	2.76002
80 dB(A)	3.78565	3.16531
85 dB(A)	3.83263	1.82979
90 dB(A)	2.13166	1.85307
95 dB(A)	2.01684	1.80315

encourage "reported" annoyance to decay. As we know in real-life as noise occurs around us, we are not attending to the noise exclusively. We are engaged in activities that range from being extremely passive to being extremely active. The amount of involvement in the activities that interest us should affect the amount of attention we allocate to the noise events. The request of subjects to attend to noise exclusively implies a sort of accounting procedure where subjects hear a noise, evaluate the noise and store this information for future reference. The ability to store this information and the ability to recall it can be affected by the presence of a secondary task. Therefore, if there exists a temporal parameter, the value of the parameter should be dependent on the context of the experiment. In the first experiment, it was necessary to establish the effect of attention level on the temporal parameter as well as to determine the general form of the annoyance response function.

Results of Experiment 1 indicate that the temporal parameter does improve the fit of the data to the annoyance response function in a dynamic environment and that the general form of the annoyance response function which includes a temporal parameter is a decreasing function, i.e., the longer the time to response, the more the annoyance response is reduced. When the annoyance responses made by subjects involved in a compensatory tracking task are compared to those made by subjects involved in reading a magazine, the rate of decay increases as the activity level of the context increases.

The parameters considered in Experiment 1 were the session length, the number of noise bursts, the amplitude of the noise bursts and the time of presentation of these noise bursts. The results of an analysis of variance indicate that all the parameters excluding session length are important parameters in defining the annoyance response. Session length may be an important parameter in the determination of annoyance response outside the range of values used in this experiment. However, within the range of three to nine minutes, it is not. Of the models considered (peak dB(A), energy averaging, energy summation and annoyance decay), the annoyance decay model is the only model that has a temporal parameter. Therefore, the results of Experiment 1 support a decay model of annoyance response.

Once the response function was determined, estimates of the temporal parameter could proceed. Annoyance judgments made for single noise bursts with various times to response could define the annoyance response function. The annoyance response function for Experiment 2 indicates that annoyance appears to decay exponentially and then level off or decrease more slowly after about three minutes. This decrease in the rate of decay after three minutes is attributable to the annoyance level falling below each subject's threshold for annoyance. At this point, subjects drift more slowly downward in their responses, in that once they are no longer annoyed, it is difficult for them to respond with an appropriate annoyance response.

Experiment 3 proceeded in the same manner as Experiment 2 but considered only decay times between zero and three minutes with more divisions between those two decay times considered. The results of this experiment concur with the results of Experiment 2 for the same presentation time interval.

Since the determination of decay constants for Experiments 2 and 3 provide no obviously intuitive information regarding the rate of decay, the half-life for these functions is calculated. This is the time necessary for the annoyance response to decrease by one-half. Results indicate that within 1.8 to 3.8 minutes, the annoyance response for all amplitudes decreases by one-half. The higher amplitudes decay more rapidly (half-life between 1.8 and 2.1 minutes) than the lower amplitudes (half-life between 2.7 and 3.8 minutes).

This experiment provides both viable results and methodological tools for continued research on the effect of noise environments upon the emotional environment of individuals. First, if one is interested in short time periods of noise, the results of these experiments may generalize to other environments provided the attention level of individuals is approximated by the attention level of the subjects in this experiment to the noise bursts. Secondly, if one is interested in other noise sources or other contexts, the procedures established here are certainly reasonable for many other stimuli and contexts. Also, if one is interested in multiple noise sources, these experiments establish a foundation for this research effort.

Further investigations might attempt to focus more intensely on the cause of the decay in annoyance response. The cause of the decay whether the decay of annoyance response is attributable to a decrease in the memory strength for the source of annoyance and therefore the associated annoyance response is reduced as time passes, or whether the decay is attributable to the cognitive interaction on the decaying physiological response is not addressed in this research.

It is established in these experiments that the annoyance response function for individual noise bursts as a function of time is a decreasing function which can be best described as an exponential decay function of the form:

$$\text{Annoyance Response} = A_0 \, e^{at},$$

where A_0 is the annoyance response at $t=0$ and $\underline{a}$ is the decay constant. The value of the decay constant appears to be dependent on the level of activity within the context of the experiment (the more active the task, the greater the rate of decay) and on the amplitude of the noise burst (the greater the amplitude the greater the rate of decay).

Reference Notes

1. Galanter, E. & Popper, R. *The annoyance of ensembles of multiple simulated aircraft overflights* (Report No. DOT-FATOWA-3871). Columbia University, New York: Psychophysics Laboratory, June 1977.

2. Balin, L., Haber, N.L. & Popper, R. *Energy summation versus group size experiment*. Unpublished manuscript, Columbia University, 1977.

3. Haber, N.L. & Karsten, G. *Energy averaging versus energy summation*. Unpublished manuscript, Columbia University, 1978.

4. Perera, T.B., Popper, R. & Galanter, E. *The relation between annoyance judgments and aircraft overflight number*. Paper presented at the meeting of the Eastern Psychological Association, Boston, April 1977.

5. Haber, N.L. *Arrivals versus departures: do they differ?* Unpublished manuscript, Columbia University, 1977.

6. Powell, C.A. *Multiple-event airplane noise annoyance* (Tech. Paper NASA TP-2101). Hampton, Va.: NASA Langley Research Center, January 1983.

References

Berglund, B., Berglund, K. & Lindvall, T. Scaling loudness, noisiness, and annoyance of aircraft sounds. *Journal of the Acoustical Society of America*, 1975, *57*, 930-934.

Blatz, W.E. The cardiac, respiratory, and electrical phenomena involved in the emotion of fear. *Journal of Experimental Psychology*, 1925, *8*, 109-132.

Bloomfield, P. *Fourier Analysis of Time Series: An Introduction*. New York: Wiley & Sons, 1976.

Borsky, P.N. *A comparison of a laboratory and field study of annoyance and acceptability of aircraft noise exposures*. Columbia University, New York: Noise Research Unit, February 1977. (NTIS No. N77-17608)

Borsky, P.N. & Leonard, H.S. *A new field survey-laboratory methodology for studying human response to noise*. *Proceedings of the International Congress on Noise as a Public Health Problem* Held at Dubrovnik (Yugoslavia) on May 13-18 1973. Environmental Protection Agency, Arlington, Va.: Office of Noise Abatement Control, May 1973. (NTIS No. PB-241-060)

Buck, R. *Human Motivation and Emotion*. New York: Wiley & Sons, 1976.

Coates, G.D., Adkins, C.J. & Alluisi, E.A. Human performance and aircraft-type noise interactions. *Journal of Auditory Research*, 1975, *15*, 197-207.

Coates, G.D., Alluisi, E.A. & Adkins, C.J. *A review of methodological factors in performance assessments of time-varying aircraft noise effects*. Norfolk, Va.: Old Dominion University, March 1977. (NTIS No. N77-19752)

Cochran, W.G. & Cox, G.M. *Experimental Designs*. New York: Wiley & Sons, 1950.

Connor, W.K. & Patterson, H.P. *Analysis of the effect of numbers of aircraft operations on community annoyance*. Austin, Tex.: Tracor, Inc., August 1976. (NTIS No. N76-30181)

Ekman, P. Differential communication of affect by head and body cues. *Journal of Personality and Social Psychology*, 1965, *2*, 726-735.

Eldred, K.M. Assessment of community noise. *Journal of Sound and Vibration*, 1975, *43*, 137-146.

Frijda, N. & Van der Ceer, J. Codability and recognition: an experiment with facial expressions. *Acta Psychologica*, 1961, *18*, 360-368.

Hilgard, E.R. Pain as a puzzle for psychology and physiology. *American Psychologist*, 1969, *24*, 103-113.

Jeffries, R. & Fisher, G. *Pet™ Fun and Games*. Berkeley, California: Osborne/McGraw-Hill, 1981.

Kryter, K.D. *The Effects of Noise on Man*. New York: Academic Press, 1970.

Kryter, K.D. & Pearsons, K.S. Some effects of spectral content and duration on perceived noisiness level. *Journal of the Acoustical Society of America*, 1963, *35*, 866-883.

Loeb, M. Expected effects of aircraft-type noises. *Journal of Auditory Research*, 1975, *15*, 191-196.

Mackay, C.J. The measurement of mood and psychophysiological activity using self-report techniques. In I. Martin & P.H. Venables (Eds.), *Techniques in Psychophysiology*. New York: Wiley & Sons, 1980.

Molino, J.A. Annoyance and noise. In C.M. Harris (Ed.), *Handbook of Noise Control* (2nd ed.). New York: McGraw-Hill, 1979.

Nisbett, R.E. & Schachter, S. Cognitive manipulation of pain. *Journal of Experimental Social Psychology*, 1966, *2*, 227-236.

Nixon, C.W., von Gierke, H.E. & Rosinger, G. Comparative annoyances of "approaching" versus "receding" sound sources. *Journal of the Acoustical Society of America*, 1969, *45*, 330.

Pearsons, K.S. & Bennett, R. *Handbook of Noise Ratings*. Canoga Park, Calif.: Bolt Beranek & Newman, Inc., April 1974. (NTIS No. N74-23275)

Poulton, E.C. The new psychophysics: six models for magnitude estimation. *Psychological Bulletin*, 1968, *69*, 1-19.

Powell, C.A. *Annoyance due to multiple airplane noise exposure*. Hampton, Va.: Langley Research Center, August 1980. (NTIS No. N80-28946)

Rice, C.G. Investigation of the trade-off effects of aircraft noise and number. *Journal of Sound and Vibration*, 1977a, *52*, 325-344.

Rice, C.G. Development of cumulative noise measures for the prediction of general annoyance in an average population. *Journal of Sound and Vibration*, 1977b, *52*, 345-364.

Rylander, R. & Sorenson, S. *Aircraft noise determinants for the extent of annoyance reactions*. Proceedings of the International Congress on Noise as a Public Health Problem Held at Dubrovnik (Yugoslavia) on May 13-18 1973. Environmental Protection Agency, Arlington, Va.: Office of Noise Abatement Control, May 1973. (NTIS No. PB-241-060)

Rylander, R., Sorenson, S., Alexandre, A. & Gilbert, Ph. Determinants for aircraft noise annoyance-- a comparison between French and Scandinavian data. Journal of Sound and Vibration, 1973, 28, 15-21.

Rylander, R., Sorenson, S. & Berglund, K. Re-analysis of aircraft noise annoyance data against dB(A) peak concept. Journal of Sound and Vibration, 1974, 36, 399-406.

Rylander, R., Sorenson, S. & Kajland, A. Annoyance reactions from aircraft noise exposure. Journal of Sound and Vibration, 1972, 24, 419-444.

Schachter, S. & Singer, J.E. Cognitive, social, and physiological determinants of emotional state. Psychological Review, 1962, 69, 379-399.

Shepherd, K.P. Cumulative annoyance due to multiple aircraft flyovers with differing peak noise levels. Hampton, Va.: Langley Research Center, May 1981. (NTIS No. N81-23712)

Sorenson, S., Berglund, K. & Rylander, R. Reaction patterns in annoyance response to aircraft noise. Proceedings of the International Congress on Noise as a Public Health Problem Held at Dubrovnik (Yugoslavia) on May 13-18 1973. Environmental Protection Agency, Arlington, Va.: Office of Noise Abatement Control, May 1973. (NTIS No. PB-241-060)

Stevens, S.S. & Galanter, E.H. Ratio scales and category scales for a dozen perceptual continua. Journal of Experimental Psychology, 1957, 54, 377-411.

Winer, B.J. Statistical Principles in Experimental Design (2nd ed.). New York: McGraw-Hill, 1971. (Originally published, 1962)

Young, R.W. & Peterson, A. On estimating noisiness of aircraft sounds. Journal of the Acoustical Society of America, 1969, 45, 834-838.

Appendix II

N83 36675 D3

SCALING LOUDNESS DIFFERENCES BY INTRA-MODAL MATCHING: EVIDENCE FOR A SINGLE SCALE UNDERLYING LOUDNESS AND LOUDNESS DIFFERENCE

Richard Dorian Popper

The perception of loudness was studied in four experiments. In Experiments 1 and 2 (difference matching), subjects judged the magnitude of loudness differences by adjusting the loudness of a single tone to equal the loudness difference between a pair of tones. Seventy-two tone pairs were used, constructed from nine sound pressure levels of a 1000 Hz tone ranging from 46-94 dB (Experiment 1) or 55-95 dB (Experiment 2). In Experiment 3 (difference estimation), subjects magnitude-estimated the loudness differences of the same tone pairs as in Experiment 2, and in Experiment 4 they magnitude-estimated the loudness of the individual tones used to construct the tone pairs. The same subjects served in Experiments 2-4, and binaural listening through earphones was used throughout.

Difference matches and difference estimates were analyzed nonmetrically to derive underlying scales of loudness. For both kinds of judgments, loudness grew as a power function of sound pressure, although in Experiment 1 some departure from the power function was evident at low intensities.

A comparison of the exponents for Experiments 2 and 3 showed that for the majority of subjects, difference matching produced larger exponents than difference estimation. The difference matching exponents were in the vicinity of the sone scale exponent for loudness, whereas the difference estimation exponents were smaller.

Further analysis showed that the scale underlying the difference matches was similar to the loudness scale for the tones used to do the matching and for the single tones magnitude-estimated in Experiment 4. This suggests that the same scale can underlie both judgments of loudness and loudness difference, a hitherto unobserved phenomenon. Possible sources for the disagreement between difference matching and estimation are discussed, as are the individual differences observed in the performance of each task.

INTRODUCTION

Demonstrations of the simultaneous and independent operation of two scales of loudness (Marks, 1979a) argue in favor of an heirarchical model of judgment, in which separate scales reflect separate processing stages. Consider therefore an experiment in which two tones are played and the subject is instructed to adjust the loudness of a third tone to equal the size of the loudness difference. This task involves the matching of a loudness difference and might therefore involve the use of both sensory and difference scales, as Marks calls them. Formally, the matching task can be represented as follows:

$$I_j^{\alpha} - I_i^{\alpha} = I_m^{\beta} \quad (1)$$

I_j is the louder tone in the pair, and I_m is the single tone whose loudness is used by the subject to match the size of the loudness difference. The exponent alpha governs the perception of loudness differences, the exponent beta the perception of loudness.

The possibility of matching the loudness of a single tone to a loudness difference (difference matching) was already conceived by Fechner, who wrote, "given two tones of differing physical intensity, it is possible to imagine a third tone, whose intensity equals the difference in intensity between them" (1860. I, 48). Stumpf, on the other hand, claimed that it was impossible to subtract one sensation from another and "to feel the remainder by itself" (quoted in James, 1890, I, p. 547). Russell quipped: "A change of length is itself a length, but a change of temparature or illumination is not itself hot or bright ... With intensive quantities ... these differences in quantity are not themselves quantities. The difference between two intensive quantities, in fact, differs from each as much as the difference between two horses differs from a horse" (quoted in Titchener, 1905, p. lxxxiii). However, this a priori argument is nothing but a restatement of the "quantity objection." Whether a measurement scale can be formed on the basis of such a task is subject to empirical investigation.

The procedure of difference matching can be compared to the task of numerically estimating the size of a loudness difference, represented as

$$I_j^{\gamma} - I_i^{\gamma} = N^{\delta} \qquad (2)$$

In Equation 2, N represents the number reported by the subject, the exponent gamma governs the perception of loudness differences, and the exponent delta represents any nonlinearity in the subject's use of numbers.

Several predictions can be made concerning the relationships between the exponents in the above equations. First, note that the exponents alpha and gamma govern how loudness grows when loudness differences are being judged. The most obvious prediction concerning these two exponents is that they will be the same. That is, the expectation is that the same power function is recovered, regardless of whether the method of judgment is difference matching or estimation. Furthermore, since loudness differences are being judged, these exponents should fall in the neighborhood of previous estimates of the interval scale exponent for loudness, namely around 0.30. On the other hand, differences between alpha and gamma can result from either perceptual differences associated with the two procedures or response bias factors.

A second prediction concerns the exponent beta in Equation 1, which reflects the growth of the loudness of the single tone used to match the loudness difference. One prediction for this exponent is that it should approximate the exponent for sensation magnitude, since loudness, not loudness difference is involved. Thus, beta is expected to be in the neighborhood of 0.60, the exponent for the sone scale of loudness. However, it is possible to interpret the difference matching task as involving two loudness differences, one between the two tones in the pair, the other between the loudness of the single tone and "zero" loudness. According to this interpretation, subjects are engaged in matching one kind of loudness difference to another, instead of matching a loudness magnitude to a loudness difference. The second interpretation predicts that beta will equal the difference exponent for loudness, not the exponent for sensation magnitude.

One reason for doubting this outcome derives from the results on the bisection of loudness (Stevens, 1975). Loudness bisection generally leads to scales similar to those obtained with other interval scaling procedures. However, when a single tone is presented for bisection (with subjective zero as the lower endpoint), subjects switch from judging intervals to judging ratios and produce a tone that is half as loud as the single tone. Such "bisections" are in fact fractionations and imply that the single tone is judged on the scale of sensory magnitude, not sensory difference. On the other hand, in the context of magnitude estimating a series of loudness differences, Marks (1979a) found that subjects judged the difference between a tone and a blank stimulus on a scale approximating the difference scale. Thus, it seems possible that the single tone could be judged on the difference scale, instead of the magnitude scale, in which case beta would equal 0.30.

The present research will address several questions. Is the exponent for loudness difference the same for difference matching and difference estimation? Do the data from difference matching support the dualistic position that there are two loudness scales, one for differences and the other for magnitudes? If a single scale can be found that underlies both kinds of judgment, how does this scale compare to that obtained from the magnitude estimation of loudness ratios?

In the practical realm, the utility of both difference and magnitude scales has been demonstrated. The sone scale for loudness underlies several schemes for the calculation of the loudness of complex sounds including noise (Kryter, 1970), whereas the Munsell scale of light reflectance, a category scale, has been applied to printing and photography (Newhall, Nickerson & Judd, 1943). A dualistic viewpoint allows for the peaceful coexistence of these two types of scales. A monistic theory, on the other hand, is troubled by the need to account for the utility of the "biased" scale.

From the point of view of theory, the persistent discrepancies between scales of sensory magnitude and sensory difference have proved to be a continued source of concern. With such divergent results, one is less confident that any psychophysical scale succeeds in measuring sensation at all. Marks has responded by making a fundamental distinction between the perception of magnitude and the perception of difference. The present research will entail a closer examination of the "psychological reality" of this distinction. The investigation of these issues in the psychophysical realm will make possible more informed applications of the psychophysical methods of magnitude and difference scaling to the measruement of nonsensory attributes.

The proposed experiments entail studying the perception of loudness differences using the novel task of difference matching. The operation of differencing, as the early history of psychophysics demonstrates, has often been viewed as the central mechanism for establishing psychological relationships. The new method for studying this task promises to provide further insight into the nature of difference perception.

Four experiments are reported here. Experiments 1 and 2 involve the task of difference matching (see Eq. 1) and differ only in the stimulus set. Experiment 3 involves the magnitude estimation of loudness differences (see Eq. 2), and Experiment 4 the magnitude estimation of the loudness ratios of single tones. The same subjects participated in Experiments 2-4. Experiment 4 yields an estimate of the ratio scaling exponent for these subjects and serves as a "control" to aid in the interpretation of the other exponents.

Several analytic strategies are available for fitting Equations 1 and 2 to the data. One procedure is to estimate the two parameters in each equation simultaneously. However, with this approach it is impossible to evaluate separately the appropriateness of a difference model and the form of the psychophysical function. By using nonmetric scaling, it becomes feasible first to assess the extent to which data from both difference matching and estimation can be described by difference structures. Furthermore, since nonmetric scaling makes weak assumptions about the functional relationship between the responses and the "true" differences, the comparison between the exponents alpha and gamma will be free of the influence of monotonic biases that may be present in either difference estimation or matching.

Loudness differences have been scaled nonmetrically before (see Schneider, 1982, for review). The scaling technique originated with the work of Shepard (1962a,b), who showed that with a sufficient number of stimuli the ordinal ranks of the interstimulus differences are sufficient for a recovery of interval scales of measurement (i.e. a scale unique up to multiplication and addition by arbitrary constants). Following Kruskal (1964), the observed differences are interpreted as monotonic with interstimulus distance (Euclidean or otherwise) in an n-dimensional stimulus space. On the assumption that loudness is unidimensional, the scaling problem reduces to one of finding points on a line, whose interpoint distances best match, in an ordinal sense, the loudness differences. This is accomplished on a computer using an iterative algorithm.

The nonmetric scaling of loudness differences and the estimation of the power function exponents alpha and gamma completes only the first stage of the analysis. In the second stage, the estimated loudness differences are used to derive estimates of beta (Equation 1) and delta (Equation 2). For this purpose, the full metric information in the responses is needed. It may seem inconsistent to neglect this information in the first stage, but subsequently to introduce it in the second. However, a nonmetric analysis is preferred in the first stage for the reasons cited above. The present approach is consistent with that employed by others (Birnbaum, 1982; Marks 1978, 1979a; Rule & Curtis, 1982) who have sought to differentiate between the "input" function, such as the scale for loudness difference, and the "output" function, such as the scale governing the use of numbers, or, in the case of difference matching, the scale for the single tone.

METHOD

Experiments 1 and 2

Subjects:

Three subjects participated in Experiment 1; subsequently, two of these (CJ and NH) and an additional five subjects participated in Experiment 2. Among the five additional subjects, two (MR and MP) had previously taken part in Experiment 3. Of these eight subjects, four were paid for their participation, either as work-study students or as undergraduate recruits. The remaining subjects, including three graduate students in psychology, were not paid. Subjects' ages ranged from 21 to 34. All subjects reported having normal hearing.

Apparatus and Procedure:

The apparatus was configured as shown in Figure 1. A 1000 Hz tone, generated by a Hewlett-Packard oscillator (Model 200 CD), was fed to Relay 1 (R1). The normally-open side was connected to an electronic switch (Grason-Stadler Model 829E). The normally-closed side was passed through a subject-controlled attenuator and from there, through the normally-closed side of Relay 2 (R2), to the electronic switch. The switch was triggered externally and gated the signal with a rise-decay time of

ORIGINAL PAGE IS
OF POOR QUALITY

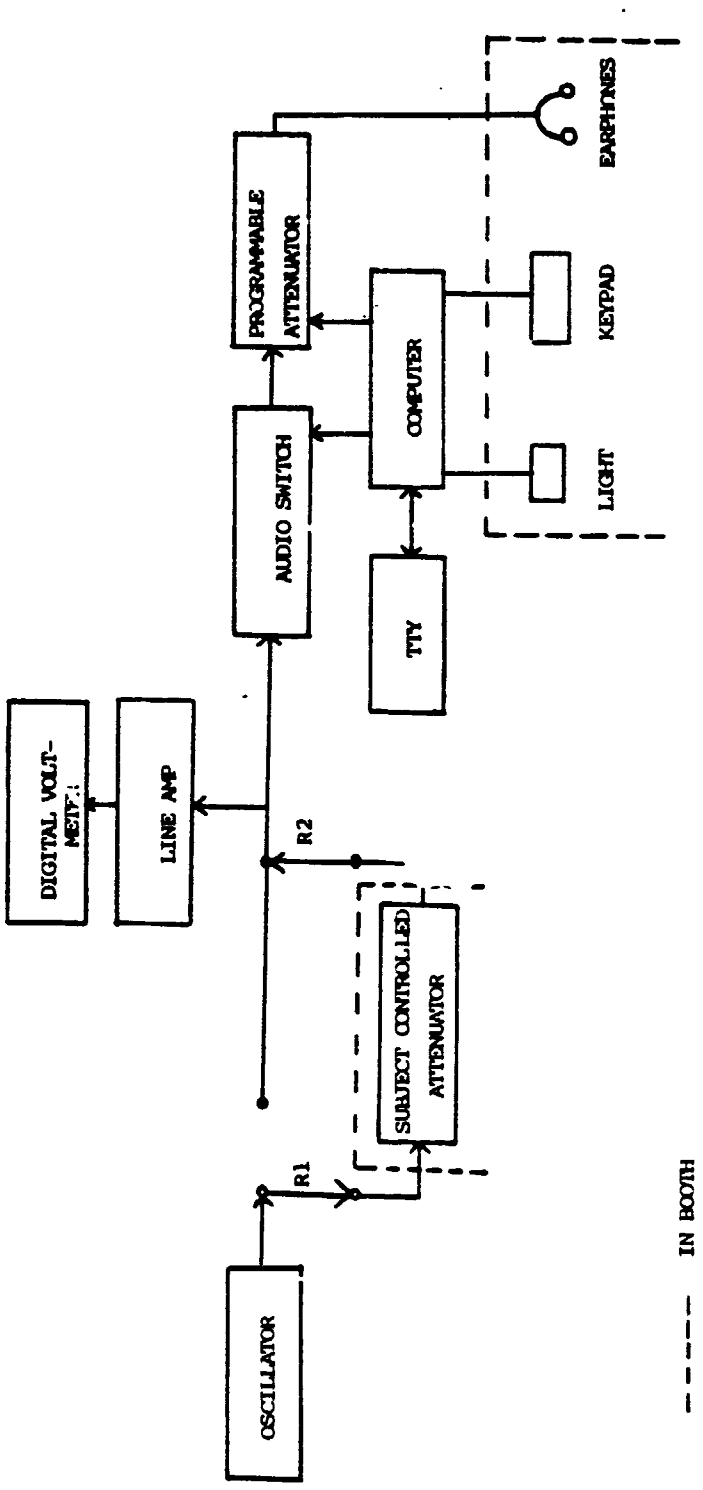

PROGRAMMABLE ATTENUATOR
EARPHONES
COMPUTER
KEYPAD
LIGHT
AUDIO SWITCH
TTY
DIGITAL VOLT-METER
LINE AMP
R2
SUBJECT CONTROLLED ATTENUATOR
R1
OSCILLATOR
IN BOOTH

FIGURE 1

10 msec. The output of the switch was passed to a programmable attenuator, and from there to the subject's earphones (TDH 39, 300 ohms, mounted in MX 41/AR cushions). Listening was binaural throughout the experiments.

All process control operations were performed by an ADS 1800E computer. These included the timing of stimulus durations and interstimulus intervals, opening and closing Relays 1 and 2, triggering the electronic switch, and controlling the programmable attenuator.

The subject-controlled attenuator was a "sone potentiometer" consisting of two 2000-ohm variable resistors ganged and cascaded (Stevens & Guirao, 1964). By rotating a knob through approximately 270 degress, the subject could vary the attenuation continuously over a range of about 77 dB. The change in attenuation with rotation is plotted in Figure 2. The knob was unmarked and no numeric scale was indicated.

The subject was provided with three keys. Pressing and releasing Key 1 resulted in the opening of Relays 1 and 2 and the presentation of two 850 msec tone bursts separated by a silent interval of 500 msec. With Relays 1 and 2 open, the subject-controlled attenuator was completely decoupled from the signal path, and the level of each burst at the subject's earphones was controlled by the setting of the programmable attenuator. Relay

ORIGINAL PAGE IS
OF POOR QUALITY

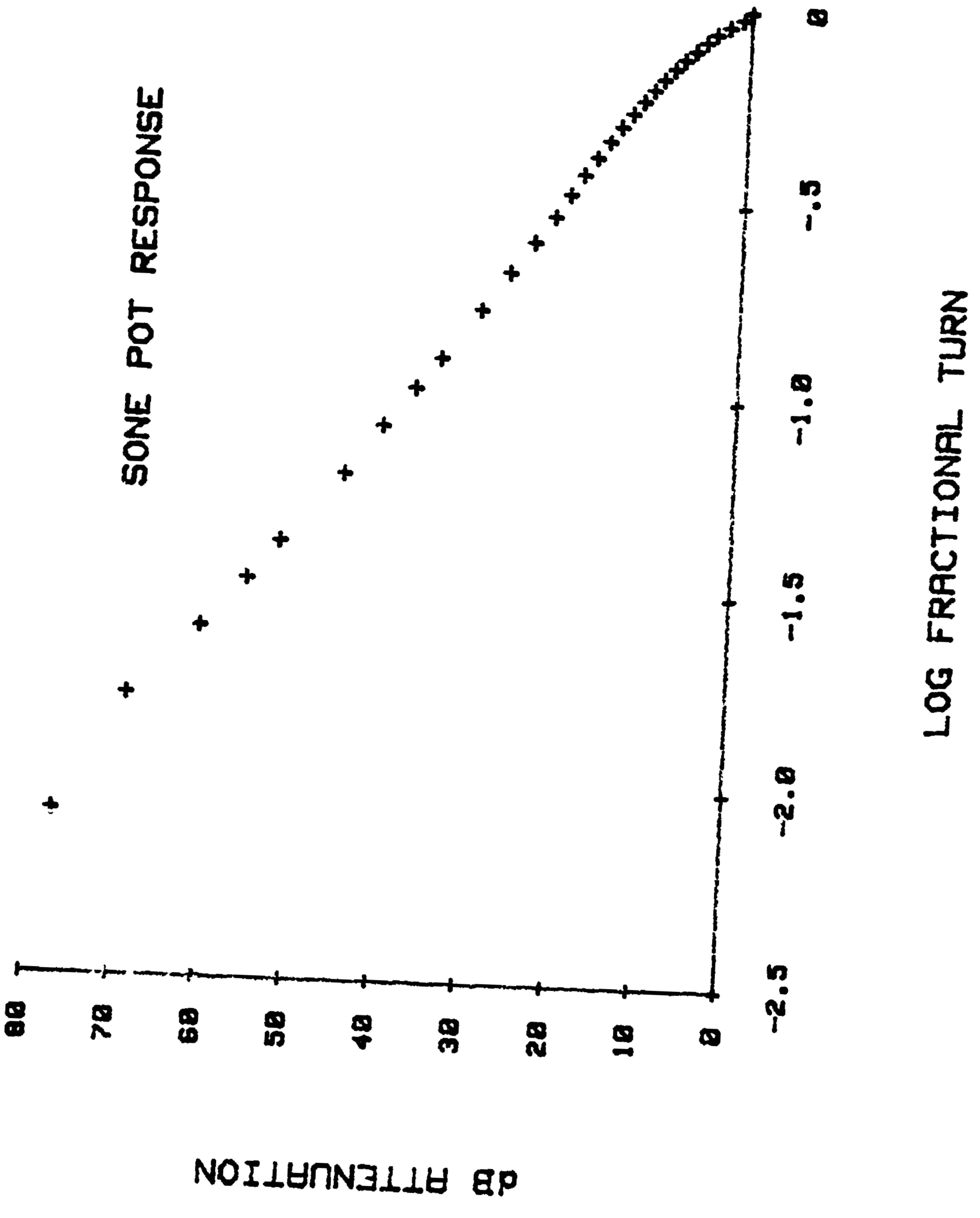


FIGURE 2

switching and changes in attenuation were accomplished without audible transients in the silent intervals surrounding the tone bursts. After the second tone burst, both relays were closed.

Pressing Key 2 resulted in the presentation of a single continuous tone that remained on for as long as the key was depressed. During its presentation, Relays 1 and 2 were in their normally closed position, passing the signal through the subject-controlled attenuator. Therefore, the level of the continuous tone was under the subject's control.

For the duration of the single tone, the programmable attenuator was set at one of two values, separated by 10 dB. Alternate values were chosen on successive trials. This served to reduce the association between the position of the knob on the subject's attenuator and the signal level at the earphones, and also prevented the subject's setting during the previous trial from being carried over to the present one.

The subject was allowed to sample between Keys 1 and 2 at will. A dead time following the presentation of either the single tone or the tone pair resulted in a minimum temporal separation of 500 msec between them. By pressing Key 3, the subject signalled the experimenter that he had completed the trial. To mark the intertrial interval, a light was turned on at the subject's station and left on until extinguished by the experimenter at the beginning of the next trial.

During the intertrial interval, the subject's setting of the single tone was determined by measuring the signal voltage at the output of the subject-controlled attenuator with a Hewlett-Packard digital voltmeter (Model 3476 B). At low levels, these measurements were taken after passing the voltage through a Hewlett-Packard line amplifier (Model 450 AR) with a maximum gain of 40 dB.

Tone frequency was set using a Hewlett-Packard electronic counter (Model 521 C) and confirmed occasionally between experimental sessions. The earphones were calibrated prior to the first experiment using a 6 cc NBS earphone coupler and a Bruel & Kjaer sound level meter (Model 2203) with octave band filter set (Model 1613). The oscillographic traces of the signals transmitted by the earphones were found to be free of clipping.

In addition, a calibration was performed that related the signal voltage at the output of the subject's attenuator to the sound pressure level at the earphones. A log unit change in attenuation produced a log unit change in sound pressure over the entire range of measurable sound levels (down to approximately 30 dB). Therefore, the voltages measured during the experiments were later converted to dB SPL by reference to a fixed voltage and its corresponding sound pressure level.

The maximum adjustable level of the single tone was 107 dB or 97 dB SPL, depending on the setting of the programmable attenuator on a given trial. The corresponding minimum levels were 30 and 20 dB SPL. By increasing the setting of the programmable attenuator, the experimenter -- upon the subject's request -- could shift the dynamic range of the sone potentiometer downward. This enabled the subject to adjust the level of the single tone as low as desired.

At the beginning and end of each experimental session, the voltages at the input to the earphones were checked at the highest and lowest levels used to construct the tone pairs. In addition, the sound pressure level produced by a middle setting of the sone potentiometer was determined. These level checks revealed only minimal drift over the course of an experimental session.

The subjects were seated in an Industrial Acoustics sound-attenuating chamber (Model 1204). In order to familiarize them with the apparatus, a warm-up task was presented at the beginning of the first session. (Subjects in Experiment 1 had performed this task during a pilot version of the experiment and did not repeat it). This task consisted of making the loudness of the tone on Key 2 match the loudness of a single tone presented on Key 1. Four such matches were performed, one of which required subjects to request a lowering of the range on the potentiometer. Communication during the experiment was conducted via intercom.

The subjects were then instructed in the experiment proper. (A complete set of instructions for all experiments is contained in Appendix 1). They were informed that Key 1 would deliver a pair of tones differing in loudness, whereas Key 2 would deliver a single tone whose loudness they could control with the potentiometer. The instructions continued as follows:

> "Your task in this experiment is to tell me how different in loudness the two tones in the pair are. You do this by adjusting the level of the single tone. In particular, adjust the level of the single tone to equal the loudness difference between the tones in the pair. That is, make the loudness of the single tone equal the difference that results from subtracting the softer tone in the pair from the louder one. If the difference is small, make the single tone soft. If the difference is large, make it loud. In other words, make the loudness of the single tone equal the size of the loudness difference."

Subjects were encouraged to arrive at their settings by bracketing. In addition, subjects were told they were free to sample the keys in any sequence and that they could request a lowering of the potentiometer range on any given trial.

Nine levels of the 1000 Hz tone were used to generate 36 pairs of unequal tones for presentation on Key 1. The levels were different for Experiments 1 and 2 and are listed in Table 1. Subjects in each experiment participated in four sessions. Within a session, each of

TABLE 1

Levels of the 1000 Hz tone used in Experiments 1-4

Experiment 1	Experiments 2,3,4
(dB SPL re 20 μPa)	
46	55
52	60
58	65
64	70
70	75
76	80
82	85
88	90
94	95

the 36 pairs occurred twice, with the order of the tones within the pair counterbalanced. The 72 pairs were presented in random order during a session. Sessions lasted 70-90 minutes, including at least one 10 minute break after the first half hour.

Experiment 3

Subjects:

The same 7 subjects who participated in Experiment 2 participated in Experiment 3. Of these, subjects MP and MR performed Experiment 3 before performing Experiment 2.

Apparatus and Procedure:

The apparatus was the same as that for Experiment 2, except that the sone potentiometer was not used. The two tones in the pair were presented in the same way as in Experiment 2, but the computer program was modified to disable all but one of the three response keys.

Subjects were instructed as follows:

> "In this experiment you will be listening to pairs of tones. The tones in each pair will differ in loudness. Your task is to decide how different in loudness the two tones are and to assign a number to that difference. You will first hear a pair of tones whose difference we will assign the number '60', to give us a starting point. For any subsequent pair, if the tones in the pair sound twice as different as did those in the first pair, assign it the number '120'. If in some pair the tones sound half as different as did those in the first pair, assign it

> the number '30'. You may use any positive number you like -- integer, fraction or decimal. You may not use negative numbers or zero."

Subjects were told that a key press would deliver the tone pair and that they were free to listen to a pair repeatedly. Each subject spoke his response over the intercom and then proceeded to the next trial.

The sound levels used to construct the tone pairs were the same as in Experiment 2. Each subject participated in four sessions, designed as in Experiments 1 and 2. Preceding the presentation of the 72 pairs was the standard pair, which was identified as the pair whose difference was equal to '60'. This standard pair was made up of tones with sound pressure levels of 65 and 80 dB. The order of these tones in the pair was alternated across sessions for each subject, and their order in the first session counterbalanced across subjects. Although the standard pair reoccurred in the series of 72 pairs, it was identified by the experimenter as the standard only at the beginning of the session. Sessions lasted 45-60 minutes, including a short break at the half-way point.

Experiment 4

Subjects:

The same subjects as in Experiments 2 and 3 participated in this experiment. Experiment 4 was performed last by all subjects.

Apparatus and Procedure:

The apparatus was the same as in Experiment 3. However, the computer program was modified for the presentation of a series of single tones.

Subjects were instructed as follows:

> "In this experiment you will be listening single tones of different loudness. Your task is to tell me how loud each tone seems to you by assigning a number to its loudness. The first tone you hear we will assign the number '100', to give us a starting point. For any subsequent tone, if the tone seems twice as loud as the first tone, assign it the number '200'. If a tone seems half as loud as the first tone, assign it the number '50'. You may use any positive number you like -- integer, fraction or decimal. You may not use negative numbers or zero."

Subjects were informed that a key press would deliver the tone, but that they could listen to each tone only once. After speaking their response, subjects went on to the next trial.

The nine sound levels used to construct the tone pairs in Experiments 2 and 3 were presented in 10 successive sets. Within a set, the presentation order of the nine tones was random, with the constraint that the first tone in a set could not be identical to the last tone of the previous set. Fach subject participated in only one session, which lasted 20-30 minutes. The standard was a tone of 75 dB SPL and was played at the beginning of the session. Subsequent occurrences of the standard in the series of 90 tones were not identified.

RESULTS

Experiment 1

Each subject's adjustments of the level of the single tone were converted to decibels of sound pressure and averaged arithmetically across the tone pairs (i,j) and (j,i) and the last three sessions. Data from the first session were discarded. The 36 average difference matches per subject are listed in Appendix 2. In order to assess between-session reliability, Kendall's coefficient of concordance, W, was computed from the ranks of the within-session averages (of pairs i,j and j,i). The values of W were 0.94, 0.94, and 0.97 for subjects BL, CJ and NH respectively, indicating good agreement across sessions.

Each subject's average difference matches were submitted to TORSCA-9 (Young, 1968) for nonmetric scaling. This scaling routine consists of two separate algorithms. The first prepares an initial configuration using factor-analytic methods and is "semi-metric" in that multiplications and additions are performed on the input data. The second algorithm is fully nonmetric and starts with the configuration derived by the first part. On a given iteration, the points corresponding to each stimulus are moved in search for the best monotonic fit between the interpoint distances and the average decibels. The degree of fit is indexed by a measure called stress (Kruskal, 1964), usually expressed as a percentage. A perfect fit has a stress of zero and exists when the interpoint distances are perfectly monotonic with the raw differences. However, with fallible data there will usually be some violation of monotonicity. Stress reflects the extent to which the interpoint distances deviate from this perfect monotonic relationship and therefore resembles the residual sums of squares in regression analysis. As an index of fit, stress also provides a criterion for determining when the best possible solution has been obtained. The iterative process terminates when further changes in the interpoint distances produce only insignificant reductions in stress.

One difficulty with iterative scaling algorithms is the possibility of terminating the iterative search at a local minimum. The risk of such nonoptimal solutions is greater in the one-dimensional than the multidimensional case, because points cannot easily move past each other during iterative computation if they lie on a line (Kruskal, Young & Seery, 1973). TORSCA-9 was chosen as a scaling routine, since it is known to be relatively robust against local minimum problems (Spence, 1970, 1972).

The final stress values for subjects BL, CJ, and NH were 5.8%, 8.0% and 2.2%. A means of evaluating the quality of the obtained solution is provided by Young's

(1970) index of metric determinacy, M, which is defined as the squared correlation between the true interpoint distances and the distances recovered by the nonmetric analysis. The index can be interpreted as the proportion of variance in the true distances which is accounted for by the recovered distances. Since the true distances are unknown, M can only be estimated. Young provides a nomogram for estimating M on the basis of the number of stimuli, the number of dimensions, and stress. For BL, CJ and NH the estimates of M were 0.98, 0.97 and 0.99, respectively.

The principal output of TORSCA consists of scale values for the nine stimulus intensities, which are contained in Appendix 3. Insofar as the nonmetric scaling results are characterized by low stress and high metric recovery, these scale values represent interval-scale measures of loudness. That is, $L_i=aT_i+b$, where L_i is the loudness of stimulus i, T_i is its TORSCA value, and a and b are arbitrary constants. A power function relationship between loudness and stimulus intensity implies that $L=aT+b=kI^n$. Equivalently, $T+(b/a)=(k/a)I^n$, or $T+u=vI^n$. Taking logarithms on both sides yields $\log(T+u)= n\log I+\log v$, which says that for some value u, the logarithm of (T+u) is a linear function of stimulus intensity in decibels. The value u is a free parameter and was chosen to maximize the squared correlation (r^2)

between log(T+u) and decibels. This was accomplished by computing r2 while varying u in small steps over a range of values. The value of u for which r2 is maximum is designated u^*, and P^* [$=T+u^*$] multiplied by 100 for convenience in plotting is called a loudness projection.

In Figures 3-5, the loudness projections are plotted on a logarithmic axis as a function of sound pressure level in decibels for each subject. In these coordinates, a power function relationship appears as a straight line. The exponent of the power function is equal to the slope of the line and can be estimated by least-squares methods. For subjects BL, CJ and NH the exponents (slopes) were 0.66, 0.46 and 0.55, and the values of r^2 0.966, 0.996, 0.961. These exponents estimate alpha in Equation 1 [$I_j^{\alpha} - I_i^{\alpha} = I_m^{\beta}$]. Both subjects BL and NH show a systematic deviation from a power function at the lowest stimulus levels, which is considered in more detail below.

The objective of this experiment was to determine two loudness function exponents: the exponent alpha (estimated above) underlying the perception of loudness difference and the exponent beta governing the loudness growth of the single tone adjusted by the subjects. Taking logarithms on both sides of Equation 1 yields $\log(I_j^{\alpha} - I_i^{\alpha}) = \beta(\log I_m)$, which states that the logarithm of the (estimated) loudness differences is linearly related to the decibel settings produced by the subject, with a slope equal to

ORIGINAL PAGE IS
OF POOR QUALITY

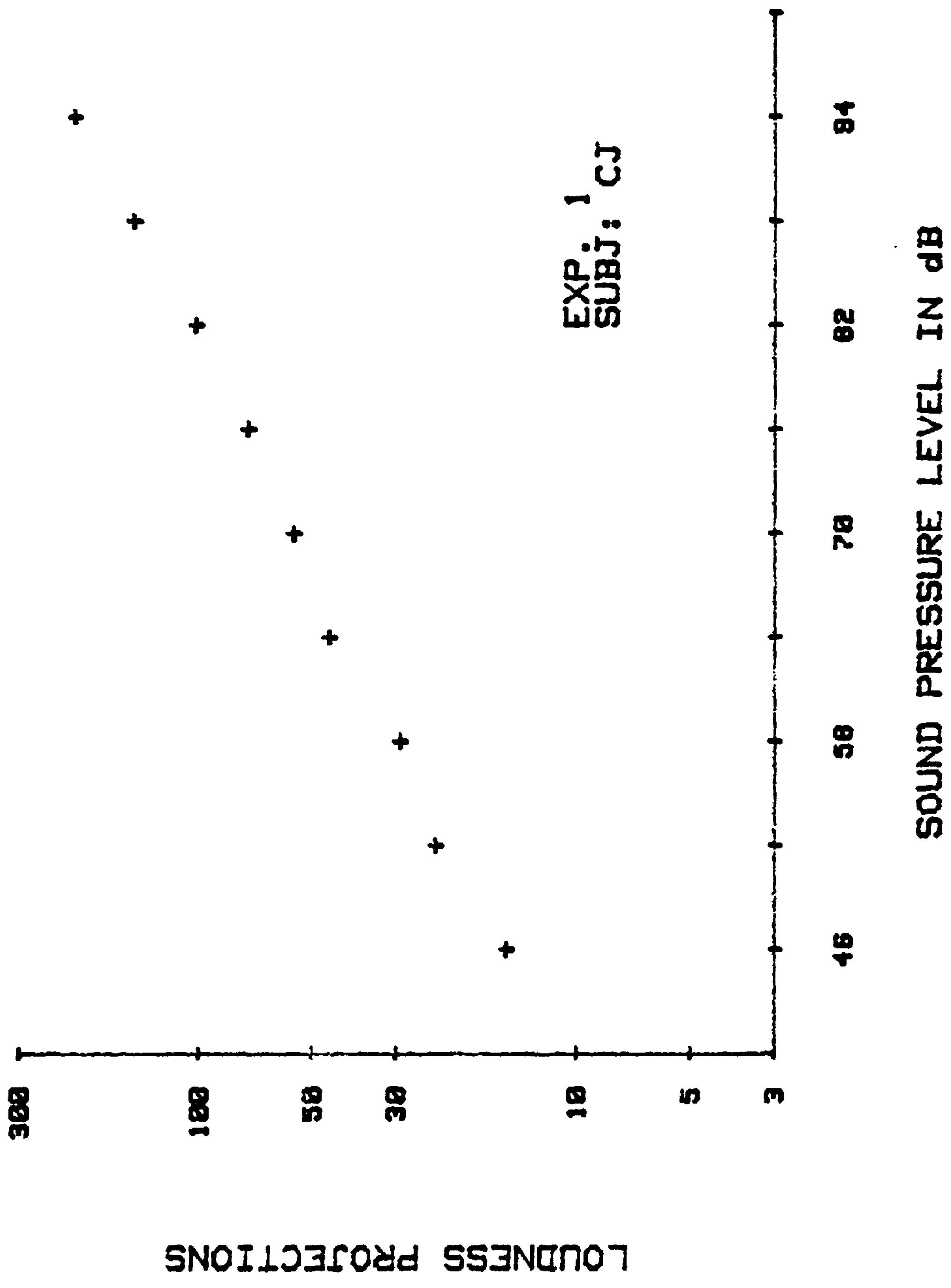


FIGURE 4

beta. The values of alpha from above can be used to estimate $(I_j^{\alpha} - I_i^{\alpha})$, and a linear regression of log I_m on $\log(I_j^{\alpha} - I_i^{\alpha})$ provides a least-squares estimate of 1/beta, from which beta can be determined by taking the inverse. The values of beta for subjects BL, CJ and NH were 0.82, 0.53 and 0.47, the corresponding r^2's 0.931, 0.893 and 0.956. The functions are shown in Figures 6-8.

In these Figures, deviations from the best fitting power function are again visible in the form of a flattening of the function towards the lower end. This suggests that subjects are responding similarly (producing similar decibel settings) to a range of small loudness differences. The curvature is consistent with evidence from Figures 3 and 5, which showed that, for subjects BL and NH, the adjusted loudness projections of the bottom three stimuli were virtually the same. Clearly, these subjects did not assign a distinct loudness to the three lowest intensities. Indeed, all three subjects reported difficulty in telling the louder from the softer tone when the tone pair was comprised of the two bottom stimuli.

These departures from a simple power function make the estimation of exponents in Eq. 1 somewhat problematic. For example, the curvature of the plot in Figure 6 reduces the slope $(1/\beta)$ of the best-fitting function, thereby leading to an overestimate of beta for subject BL. Also, for subjects BL and NH, the exponent alpha is clearly not

ORIGINAL PAGE IS
OF POOR QUALITY

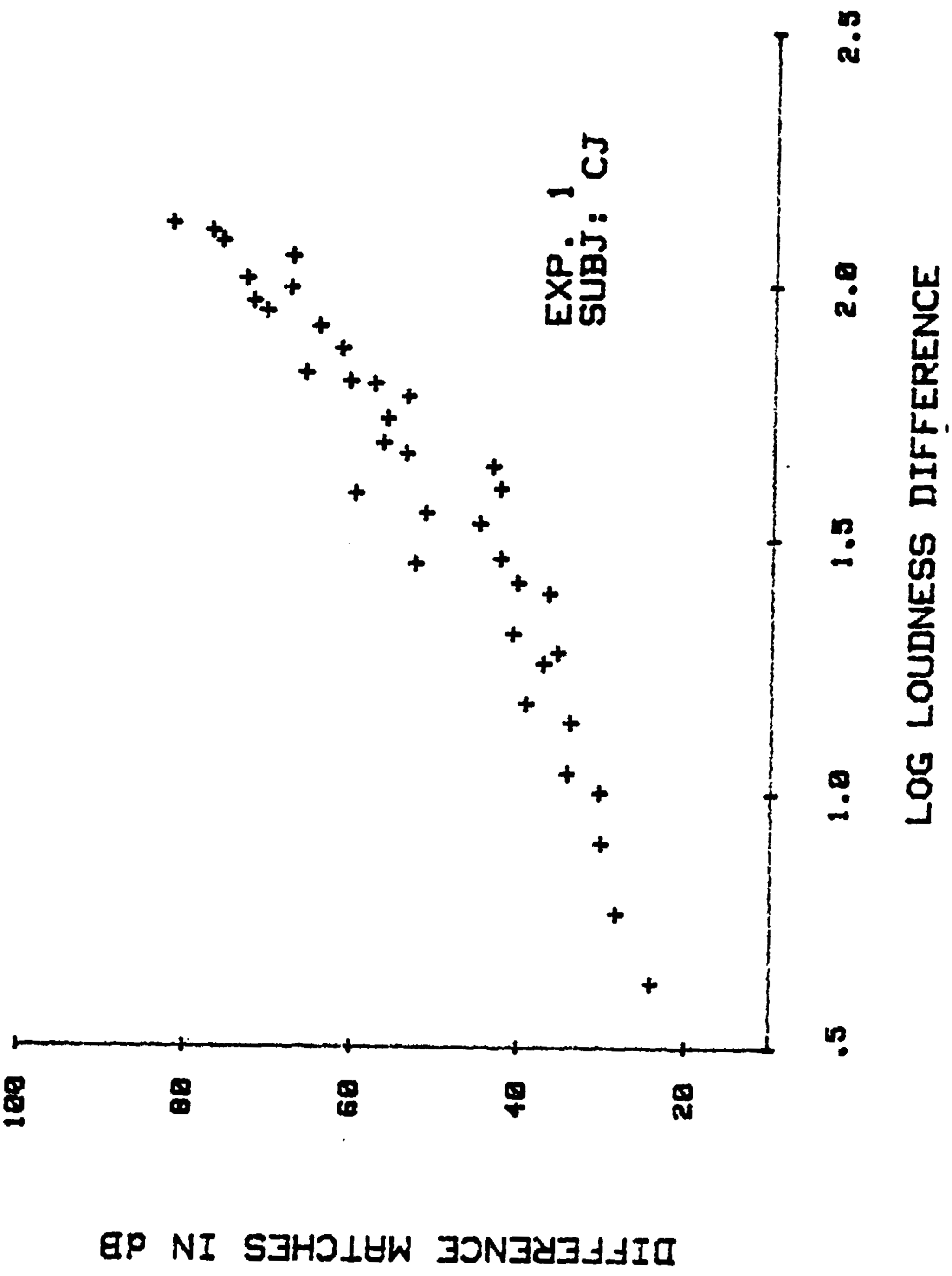


FIGURE 7

representative of the growth of loudness over the entire stimulus range. Furthermore, confusions among a subset of stimuli can influence the outcome of the nonmetric scaling in a less obvious way. Thus, while subject CJ's plot of loudness projections versus decibels accords well with a power function, confusions among the low level stimuli could well have produced a reduction in interstimulus spacing, since there would be a tendency for these stimuli to be clustered together. This would result in depressing the estimate of alpha.

Thus, Experiment 1 represents only a partial success. Subjects appear able to perform the task of difference matching and produce functions that are nearly power functions, for both loudness difference and loudness. In this respect the results are encouraging. However, further evaluations of the specific models of psychophysical judgment are hampered by the observed deviations from a simple power function. Experiment 2 was conducted in order to improve on these results by a change in the stimulus set. In this experiment, the lowest stimulus was 9 dB above the lowest stimulus in Experiment 1. This change was predicated on the assumption that the problems in Experiment 1 were due primarily to the absolute level of the stimuli. By eliminating some of the low level sounds, it was hoped that better fitting power functions could be obtained.

Experiment 2

As in Experiment 1, the difference matches from the first of four sessions were discarded and the remaining data averaged as before. (Results for subject MP are based on the last three of five sessions. Following the second session, subject MP reported having been influenced by factors other than loudness, such as the annoyance, of the tones in the pair. Therefore, data from this subject's second session were also discarded, the subject was asked to judge only loudness differences, and a fifth session was added as a replacement.) The average difference matches for the seven subjects are contained in Appendix 2.

Kendall's coefficient of concordance was again computed as a measure of across-session reliability (see Table 2). With the exception of CJ, all values are in excess of 0.91, indicating good agreement across sessions. Subject CJ shows somewhat greater variability from session to session.

The scale values produced by TORSCA-9 for each stimulus are contained in Appendix 3. Table 2 shows the final stress values. According to Young's nomogram, between 97-99% of the variance in the true distances is accounted for by the recovered distances. As a precaution against local minimum problems, the difference matches

TABLE 2

Values of the coefficients of concordance (W) and Kruskal's stress for Experiments 2 and 3.

	Experiment 2		Experiment 3	
Subject	W	Stress (%)	W	Stress (%)
CJ	.87	8.5	.85	7.9
GA	.98	2.7	.97	2.1
			(.97)	(2.8)
JF	.96	4.3	.92	4.4
JP	.94	6.4	.95	4.9
MP	.96	6.1	.92	5.3
MR	.94	7.1	.95	5.3
NH	.96	3.7	.94	3.2

were reanalyzed using UNICON (Roskam, 1977). This scaling routine is entirely nonmetric and quite different in its algorithmic structure from TORSCA-9. Despite these differences, the squared correlations for each subject between the two sets of scale values were equal to or exceeded 0.9994, indicating excellent agreement between the two scaling routines. Two such different scaling algorithms are unlikely to arrive at the same suboptimal solution, indicating that the obtained solutions represent global rather than local minima.

In Figures 9-15, the loudness projections are plotted on a logarithmic axis against sound pressure level in decibels. The procedure for optimizing the power function fits was the same as the one used in the analysis of Experiment 1. Only two subjects (MP and NH) exhibit any pronounced flattening at the bottom end of the function. Table 3 shows the exponents (slopes) alpha of the power functions (estimated by the method of least squares); the corresponding squared correlations are contained in Table 4. Four of the seven exponents fall between 0.56 and 0.64. Subject JP's exponent of 0.32 is considerably smaller, and subjects GA and NH have exponents of intermediate size. Visual inspection and the squared correlations show a good fit to the power function model.

Figures 16-22 contain the plots of produced decibels versus log estimated loudness differences. Visual inspec-

TABLE 4

Squared correlations for the power function fits to Experiments 2 and 3

Subject	r^2_α	r^2_β	r^2_γ	r^2_δ
CJ	.975	.950	.991	.940
GA	.998	.973	.997 (.982)	.951 (.892)
JF	996	.927	.980	.937
JP	[illegible]	.913	.980	.937
MP	.974	.939	.996	.938
MR	.972	.910	.994	.925
NH	.980	.962	.991	.941

tion of these graphs also shows an increase in linearity over the results from Experiment 1 (compare, for example, subject CJ's data in Figures 7 and 16). The least squares estimates of the exponents (slopes) beta and the correpsonding r^2's are listed in Tables 3 and 4. Four of the seven exponents fall in the range 0.55 and 0.66. The remaining exponents are 0.42, 0.51 and 0.52. Although these plots and the plots of the loudness projections show instances in which the data depart somewhat from a simple power function model, the departures do not appear to be severe nor do they recur with every subject.

Recall that the objective of this experiment was to determine the relation between two exponents: the exponent underlying loudness difference (alpha) and the exponent underlying loudness (beta). The prediction was that beta would exceed alpha. In particular, based on past results it was predicted that beta would be approximately twice the value of alpha. Table 3, however, shows that alpha and beta are very close to each other. Indeed, the average difference between alpha and beta is not significant (paired $\underline{t}(6)=-1.00$, $\underline{p}>0.1$, one-tailed). This suggests that subjects judged loudness and loudness difference on the same scale, a quite unexpected result. Furthermore, alpha and beta fall in the vicinity of 0.60, the exponent of the sone scale of loudness, which is based on the judgments of single tone loudness ratios.

ORIGINAL PAGE IS
OF POOR QUALITY

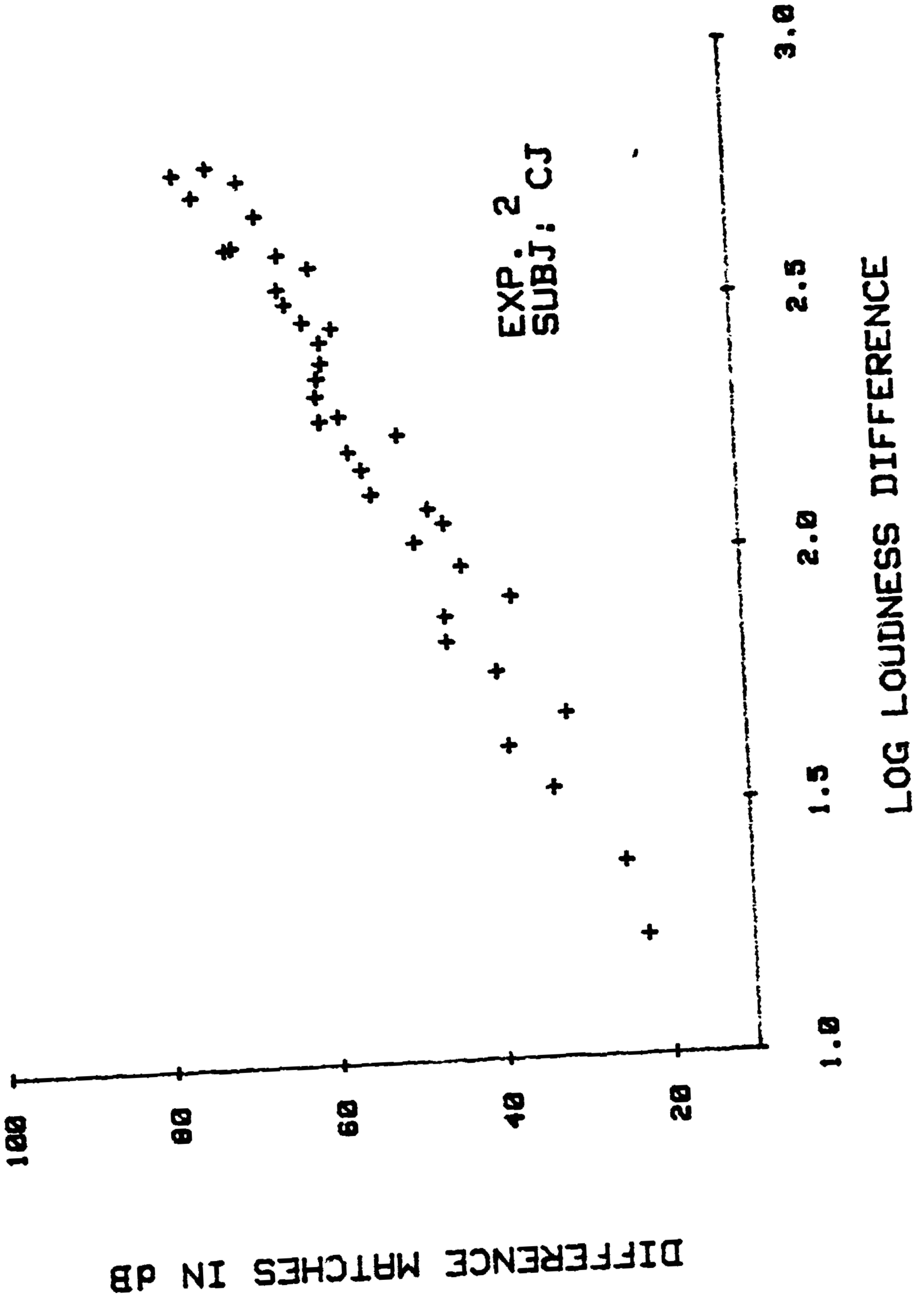


FIGURE 16

Past studies have found judgments of loudness and loudness difference to be distinct, but have primarily involved numeric estimates of loudness differences rather than difference matching. The interpretation of the foregoing results will heavily depend on the results of the remaining experiments, in which the same subjects are asked to magnitude estimate loudness differences and the loudness of single tones.

Experiment 3

In this experiment, subjects estimated numerically the ratio of loudness differences relative to a standard loudness difference. Data from the first session were ignored and the remaining data averaged by taking the geometric mean of the estimates for pairs (i,j) and (j,i) across the last three sessions. In the case of subject CJ, the nonmetric analysis of these averages resulted in a fairly large stress of 11.3%. For this reason, three additional sessions were run with this subject and included in the analysis, which resulted in a reduction in stress (see below). The estimates of the power function parameters, however, were only minimally affected by the inclusion of these additional data.

A further comment is required concerning the data from subject GA. During the debriefing following the fourth session, subject GA reported having judged loudness ratios instead of differences throughout this experiment (but not the previous one). This, according to GA, seemed "natural" when the task was numeric estimation. Since GA indicated that he could have judged differences if he had chosen to, he was asked to repeat the experiment, this time judging loudness differences. Both sets of results are reported below, with the ratio estimation data enclosed in parentheses.

Each subject's 36 geometric means are contained in Appendix 2; the TORSCA values are listed in Appendix 3. Table 2 shows the values of Kendall's coefficient of concordance and stress. Young's index of metric determinacy, indicating the amount of explained variance, was estimated at 97-99%. The data were reanalyzed using UNICON, in order to guard against suboptimal solutions. The squared correlations between the two sets of scale values were all equal to or in excess of 0.9999.

Plots of loudness projections (on a logarithmic axis) versus sound pressure level in decibels are shown in Figures 23-29. In this experiment, the exponents (slopes) of these functions estimate gamma (see Equation 2). The least-squares estimates of gamma are contained in Table 3; the associated r^2's are listed in Table 4. The squared correlations and a visual inspection of the Figures show an excellent fit to the power function model. The

ORIGINAL PAGE IS
OF POOR QUALITY

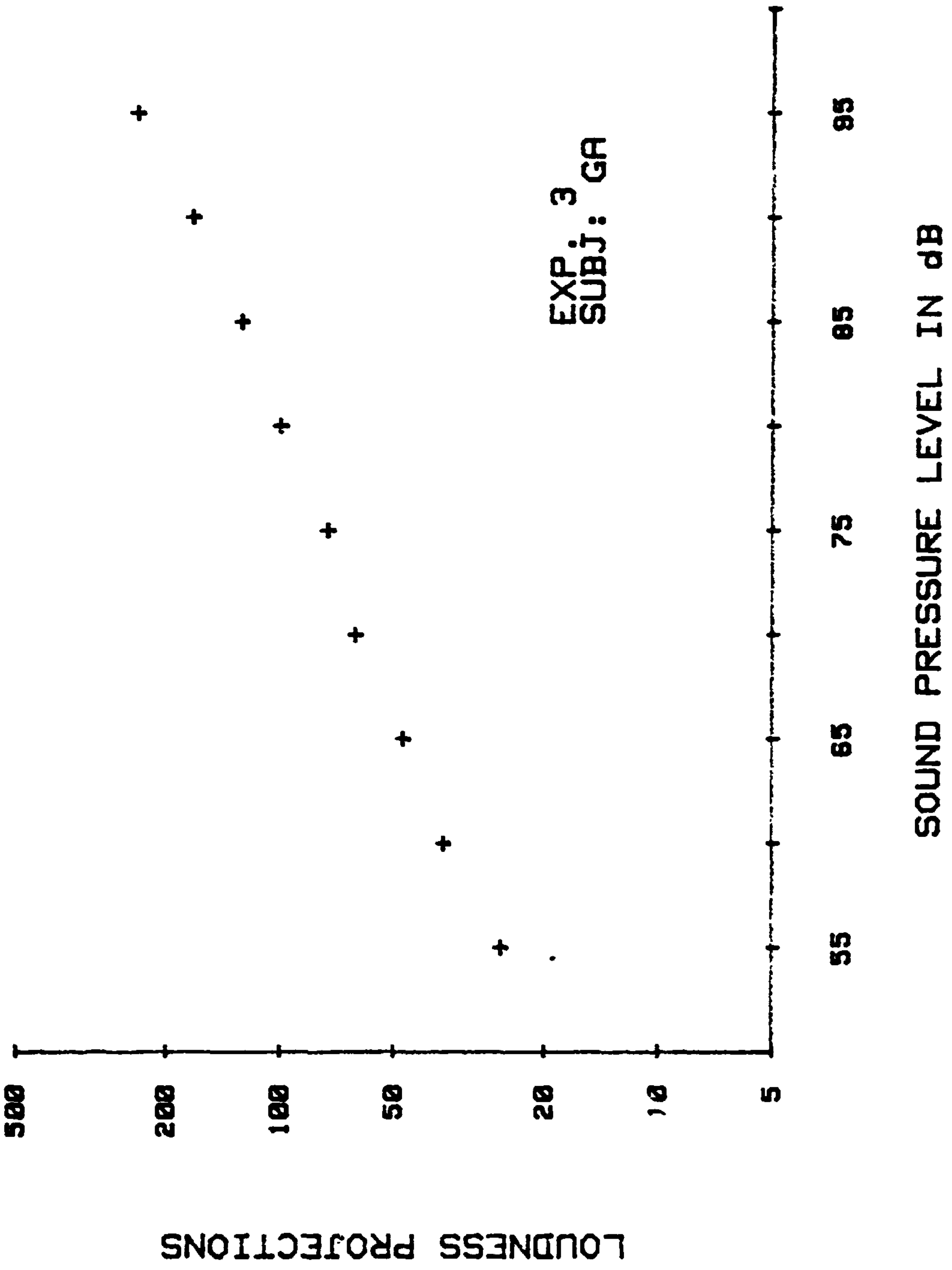


FIGURE 24

exponents range from 0.27 (excluding GA's exponent of 0.06) to 0.46.

Notice that the slope for GA's ratio estimation data (enclosed in parentheses) is considerably smaller than that based on his (and others') difference estimation results. This is consistent with the hypothesis that GA was judging loudness ratios instead of differences, for the following reason. If the first results represent judgments of loudness ratios L_i/L_j, then the application of a difference model to these data requires that the stimulus points be interpreted as projected onto a line scaled in log loudness (a log interval scale), since $\log(L_i/L_j)=\log L_i-\log L_j$. Hence, $T=a\log L+b$, and on the assumption that loudness is a power function of intensity (i.e. log loudness is linear with log I), $T=a'\log I+b'$. This says that the TORSCA values are a logarithmic function, not a power function of intensity. When power functions are fit to functions that closely approximate logarithmic functions, the exponent will aproach zero (see Stevens, 1975). Thus, subject GA does indeed appear to have judged loudness ratios in the first case, but not the second.

In order to estimate (1/delta), power functions were fit (by the method of least squares) to the relation between the log of the numeric response and the log of the loudness difference (see Figures 30-36). The values of delta and the squared correlations are listed in Tables 3 and 4, respectively. Delta ranges from 0.48 to 1.35, but five out of seven estimates are clustered between 0.73 to 0.85. The results indicate that the numerical responses in this Experiment were nonlinearly related to the estimated loudness differences, since linearity would have produced a delta of 1.0.

Comparison between the exponents alpha and gamma

Table 3 shows that alpha exceeds gamma for six out of seven subjects. The average difference is significantly greater than zero (paired $t(6)=2.78$, $p<0.05$, two-tailed). A subject-by-subject test for the difference between slopes is possible, based on the assumption that the relationship between the logarithm of the loudness projections and decibels follows a linear model (Kleinbaum & Kupfer, 1978). In this case, the test for the difference between two slopes is similar to a t-test for the difference between two means. Table 5 contains the results of these tests, together with the 95% confidence intervals for each individual slope. The t-tests indicate that for all but one subject (GA) the values of alpha and gamma differ significantly.

Among the seven subjects, four (CJ, JF, MP and MR) show alpha considerably greater than gamma (by

ORIGINAL PAGE IS
OF POOR QUALITY

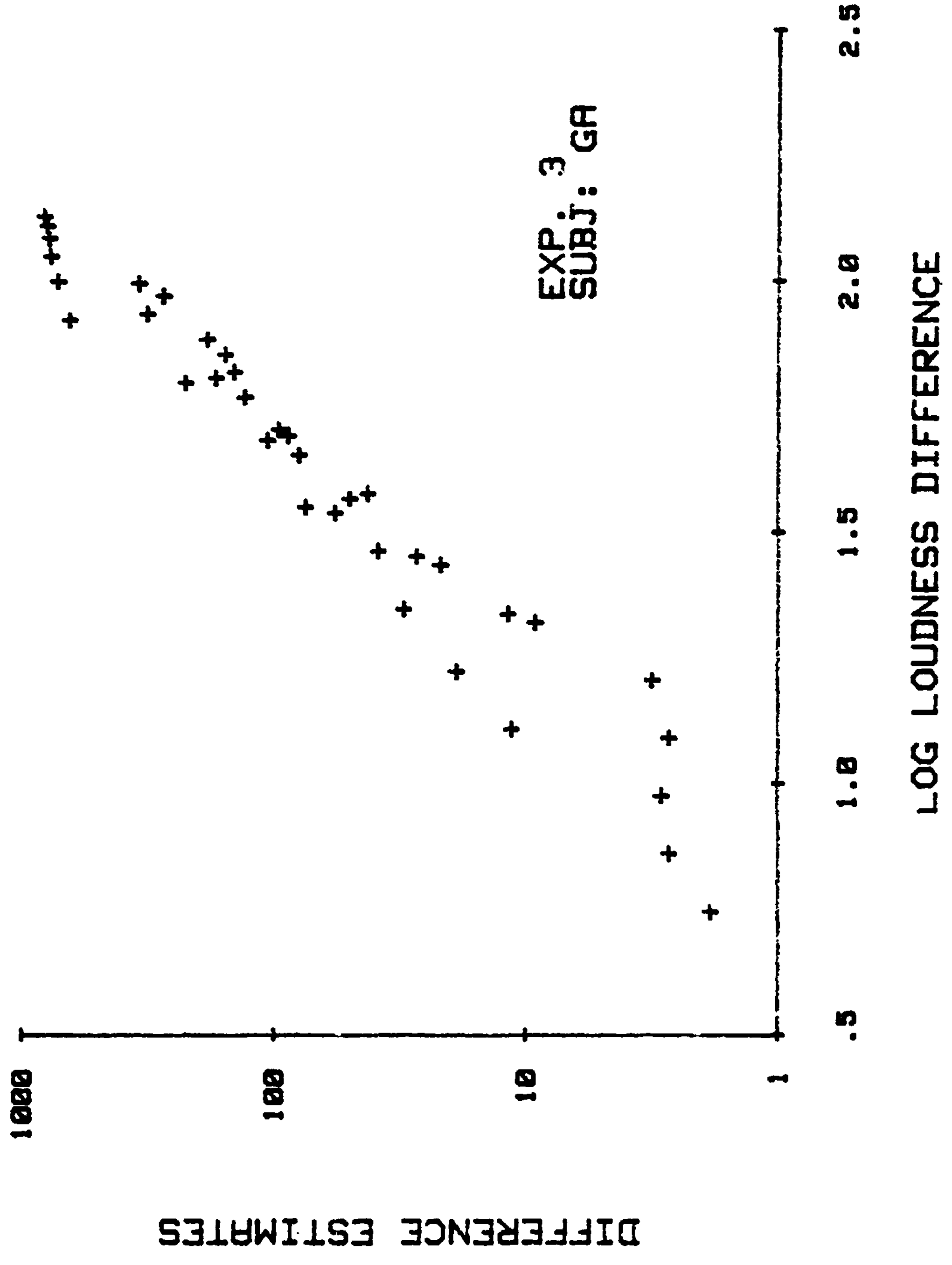


FIGURE 31

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 5

95% confidence limits and test for the difference between α and γ

Subject	α 95% CL	γ 95% CL	t(14)	
CJ	.51 - .68	.26 - .30	8.4	***
GA	.46 - .49	.44 - .48	1.2	ns
JF	.52 - .59	.32 - .42	7.6	***
JP	.31 - .34	.35 - .40	-4.9	***
MP	.51 - .68	.26 - .29	8.8	***
MR	.54 - .73	.39 - .46	4.9	***
NH	.46 - .59	.40 - .47	2.7	*

* p < .05, two-tailed

*** p < .001, two-tailed

approximately a factor of 1.5-2), JP and NH show smaller differences in exponent, with JP's gamma greater than alpha, and GA shows an insignificant difference. Notice, however, that the one reversal in the sign of the difference occurs with subject JP, whose exponent for Experiment 2 is noticeably below the other exponents obtained in that experiment. Clearly this subject is behaving quite differently from others in the difference matching task, and consequently might be expected to show a different relation between alpha and gamma.

One consideration in assessing the significance of the differences between alpha and gamma is the sensitivity of these exponents to the value of the additive constant used to optimize the power function fit. Recall that the loudness projections were adjusted additively so as to maximize the squared correlation between the logarithms of the loudness projections and decibels. Examination of this optimization process reveals that small deviations from the optimum r^2 can produce sizeable changes in the exponent as subsequently estimated by the method of least squares. In Table 6 are listed the upper and lower bounds for each exponent, as r^2 is allowed to deviate from its optimum by 0.001 and 0.005. It can be seen that for subjects CJ, JF, MP and MR the boundaries around the exponents show no or only marginal overlap between Experiments 2 and 3. The other subjects have mor

TABLE 6

Bounds on α and γ for changes of .001 and .005 in the maximum r^2 of the power function fit

Subject		α		γ	
		lower	upper	lower	upper
CJ	.001	.55	.63	.24	.32
	.005	.50	.68	.19	.36
GA	.001	.43	.52	.42	.50
	.005	.37	.58	.36	.55
JF	.001	.50	.61	.36	.41
	.005	.44	.67	.29	.45
JP	.001	.27	.37	.32	.43
	.005	.21	.43	.24	.50
MP	.001	.54	.66	.22	.32
	.005	.47	.73	.16	.38
MR	.001	.60	.67	.39	.47
	.005	.56	.71	.33	.51
NH	.001	.45	.59	.38	.49
	.005	.37	.69	.31	.56

extensive areas of overlap. These results, of course, do not carry the force of a statistical test. However, insofar as a change in the total explained variance of 0.001 or 0.005 is interpreted as defining a confidence region, the above analysis suggests that for four subjects the differences between alpha and gamma are real.

A similar pattern of results emerges if one considers the correlations between the TORSCA values of Experiments 2 and 3. For subjects GA and JP the r^2's are in excess of 0.997, whereas for subjects CJ, MP and NH the r^2's do not exceed 0.965. Subjects JF and NH fall in the middle with r-squares of 0.987 and 0.989, respectively. All these squared correlations are high, since in each experiment the scale values are highly correlated with stimulus level. However, within their narrow range, these correlations mirror the differences in exponents, which provides some reassurance that the application of power functions has not seriously distorted the relations between the two sets of scale values.

Since the analysis of the raw data (the decibels produced and the numbers uttered) is nonmetric, the differences between the data from Experiments 2 and 3 should be evident in a comparison between the ranks assigned to the stimulus pairs under the two procedures. An example of such a comparison in presented in Figure 37 (other comparisons lead to similar conclusions). The

ORIGINAL PAGE IS
OF POOR QUALITY

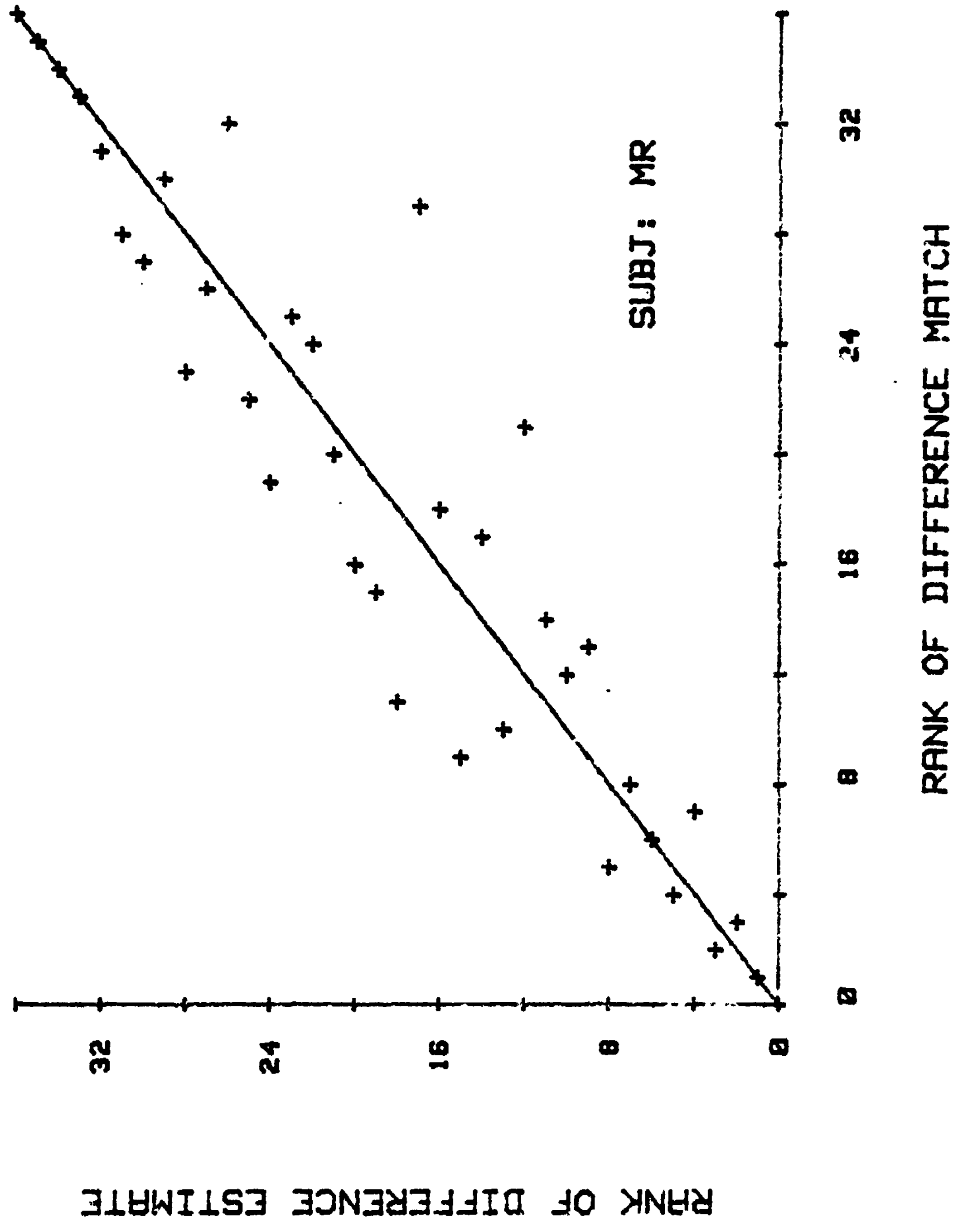


FIGURE 37

picture is consistent with what would be expected on the basis of differences in the exponents of the underlying power functions. The identity of the extreme pairs (that is, the pairs ranked lowest and highest) does not change, but there is movement in the middle. Examination of the raw averages (see Appendix 2) reveals that a pair of intense tones tends to receive a higher rank (representing a larger difference) based on difference matches than numeric estimates. For example, in Experiment 2, subject CJ judged the loudness difference between 85 and 90 dB to equal the loudness of a tone of 53.9 dB, thereby ranking it fifteenth among 36 loudness differences (the smallest loudness difference was ranked one). However, in Experiment 3, subject CJ gave the same pair a magnitude estimate of 11.69, thereby ranking it only seventh. The larger ranks for the pairs of intense tones in Experiment 2 result in a larger spacing among the intense stimuli and a steeper exponent in the recovered power function.

Experiment 4

In this experiment, subjects provided magnitude estimations of the loudness of single tones, relative to a standard tone of 75 dB SPL, whose loudness was designated 100. The responses were averaged within subjects by taking the geometric mean of the 10 estimates of each

C-2

stimulus. Appendix 2 contains the geometric means for each subject.

In Figures 38-44, the geometric means are plotted on a logarithmic axis against sound pressure level in decibels. A linear relationship in these coordinates indicates a power function relation between loudness and sound pressure. The exponents (slopes) epsilon in $N=kI^{\varepsilon}$ were estimated by the method of least squares and are listed in Table 7, along with the squared correlations. Six out of seven values of epsilon fall in the range 0.52 and 0.68. The exponent for subject CJ (0.26) is quite low by comparison. Figure 38 shows that CJ's function is indeed unusual and has a slope that would be even lower in the absence of the endpoints. Aside from this one subject, the other subjects provide exponents near the usual value of 0.60 for the magnitude estimation of single tone loudness.

ORIGINAL PAGE IS
OF POOR QUALITY

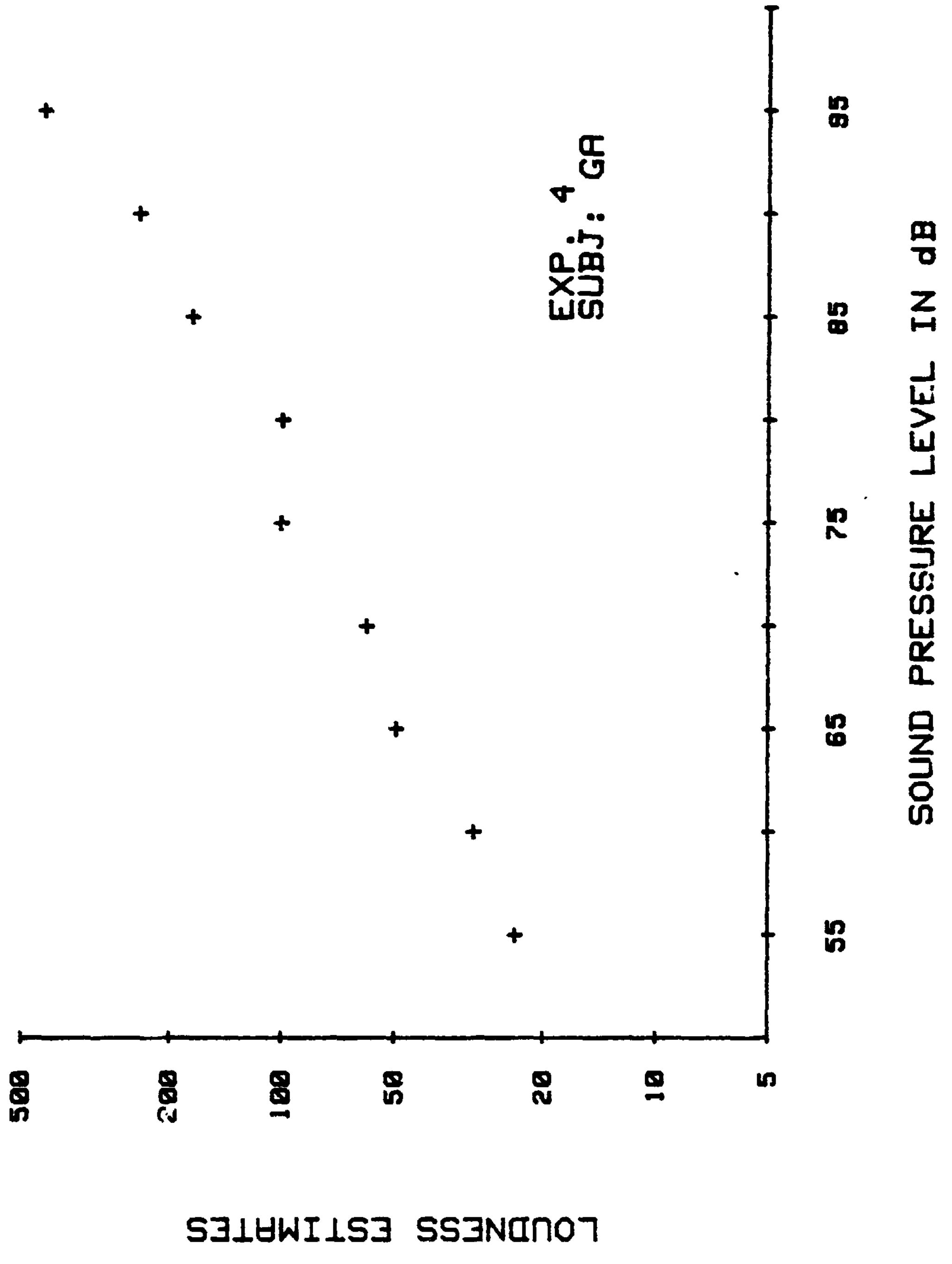


FIGURE 39

TABLE 7

Power function exponents and squared correlations for Experiment 4

Experiment 4: $N = I^{\varepsilon}$

Subject	ε	r^2
CJ	.26	.882
GA	.60	.982
JF	.52	.970
JP	.68	.988
MP	.62	.974
MR	.58	.980
NH	.52	.978

REFERENCE NOTES

1. Marks, L.E. Personal communication, 13 July 1982.
2. Marks, L.E. Personal communication, 28 January 1983.

REFERENCES

Attneave, F. Perception and related areas. In S. Koch (Ed.), Psychology: A study of a science. Vol. 4. New York: McGraw Hill, 1962.

Banks, W.P. & Coleman, M.J. Two subjective scales of number. Perception & Psychophysics, 1981, 29, 95-105.

Beck, J. & Shaw, W.A. Ratio estimations of loudness intervals. American Journal of Psychology, 1967, 80, 59-65.

Birnbaum, M.H. Controversies in psychological measurement. In B. Wegener (Ed.), Social attitudes and psychophysical measurement. Hillsdale, NJ: Lawrence Erlbaum Associates, 1982.

Birnbaum, M.H. & Elmasian, R. Loudness "ratios" and "differences" involve the same psychophysical operation. Perception & Psychophysics, 1977, 22, 383-391.

Dawson, W.E. Magnitude estimation of apparent sums and differences. Perception & Psychophysics, 1971, 9, 368-374.

Eisler, H. On the ability to estimate differences: A note on Birnbaum's subtractive model. Perception & Psychophysics, 1978, 24, 185-189.

Ekman, G. Methodological note on scales of gustatory intensity. Scandinavian Journal of Psychology, 1961, 2, 185-190.

Fechner, G.T. Elemente der Psychophysik. Th. 1 & 2. Leipzig: Breitkopf und Hartel, 1860.

Fletcher, H. & Munson, W.A. Loudness, its definition, measurement and calculation. Journal of the Acoustical Society of America, 1933, 5, 82-108.

Galanter, E. & Messick, S. The relation between category and magnitude scales of loudness. Psychological Review, 1961, 68, 363-372.

Graham, C.H. & Ratoosh, P. Notes on some interrelations of sensory psychology, perception, and behavior. In S. Koch (Ed.), Psychology: A study of a science. Vol. 4. New York: McGraw Hill, 1962.

Ham, L.B. & Parkinson, J.S. Loudness and intensity relations. Journal of the Acoustical Society, 1932, 3, 511-534.

Harris, J.D. Loudness discrimination. Journal of Speech and Hearing Disorders, 1963, Monograph Supplement II.

Hellman, R. Stability of individual loudness functions obtained by magnitude estimation and production. Perception & Psychophysics, 1981, 29, 63-70.

Hellman, R.P. & Zwislocki, J.J. Monaural loudness function at 1000 cps and interaural summation. Journal of the Acoustical Society of America, 1963, 35, 856-865.

Helson, H. Adaptation-level theory. New York: Harper & Row, 1964.

International Standards Organization (ISO) Recommendation R131 Expression of the physical and subjective magnitude of sound or noise, 1959.

James, W. The principles of psychology. Vols. I & II. New York: Holt, 1890.

Jones, F.N. & Marcus, M.J. The subject effect in judgments of subjective magnitude. Journal of Experimental Psychology, 1961, 61, 40-44.

Kleinbaum, D.G. & Kupfer, L.L. Applied Regression Analysis and Other Multivariable Methods. North Scituate, MA: Duxbury Press, 1978.

Kruskal, J.B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 1964, 29, 1-27.

Kruskal, J.B., Young, F.W. & Seery, J.B. "How to use KYST, a very flexible program to do multidimensional scaling." Bell Telephone Laboratories Manuscript, 1973.

Kryter, K. The effects of noise on man. New York: Academic Press, 1970.

Luce, R.D. What sort of measurement is psychophysical measurement? American Psychologist, 1972, 27, 96-106.

Luce, R.D. & Galanter, E. Psychophysical scaling. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology*, Vol.1. New York: Wiley & Sons, 1963.

Luce, R.D. & Green, D.M. Two tests of a neural attention hypothesis for auditory psychophysics. *Perception & Psychophysics*, 1978, *23*, 363-371

Marks, L.E. Stimulus range, number of categories, and form of the category scale. *American Journal of Psychology*, 1968, *81*, 467-479

Marks, L.E. *Sensory processes: The new psychophysics*. New York: Academic Press, 1974. (a)

Marks, L.E. On scales of sensation: Prolegomena to any future psychophysics that will be able to come forth as science. *Perception & Psychophysics*, 1974, *16*, 358-376. (b)

Marks, L.E. PHONION: Translation and Annotations concerning loudness scales and the processing of auditory intensity. In N.J. Castellan, Jr. & F. Restle (Eds.), *Cognitive Theory*, Vol.3. Hillsdale: Lawrence Erlbaum Associates, 1978.

Marks, L.E. A theory of loudness and loudness judgments. *Psychological Review*, 1979, *86*, 256-285. (a)

Marks, L.E. Sensory and cognitive factors in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 426-443. (b)

Marks, L.E. Psychophysical judgment: Wundt's theory revisited. *Psychological Research*, 1980, *42*, 157-164.

Marks, L.E. & Cain, W.S. Perception of intervals and ratios for three prothetic continua. *Journal of Experimental Psychology*, 1972, *94*, 6-17.

McGill, W.J. The slope of the loudness function: A puzzle. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications*. New York: Wiley, 1960, pp. 67-81

McGill, W.J. & Goldberg, J.P. Pure-tone intensity discrimination and energy detection. *Journal of the Acoustical Society of America*, 1968, *44*, 576-581.

Neisser, U. Cognitive Psychology. New York: Appleton Century Crofts, 1967.

Newhall, S.M., Nickerson, D. & Judd, D.B. Final report of the OSA Subcommittee on the spacing of the Munsell colors. Journal of the Optical Society of America, 1943, 33, 385-418.

Parker, S. Non-metric scaling of three perceptual continua using magnitude estimates of similarity and difference. Unpublished doctoral dissertation, Columbia University, 1974.

Parker, S. & Schneider, B. Non-metric scaling of loudness and pitch using similarity and difference estimates. Percpetion & Psychophysics, 1974, 15, 238-242.

Parker, S., Schneider, B. & Kanow, G. Ratio scale measurement of the perceived length of lines. Journal of Experimental Psychology: Human perception and Performance, 1975, 104, 195-204.

Poulton, E.C. The new psychophysics: Six models for magnitude estimation. Psychological Bulletin, 1968, 69, 1-19.

Reynolds, G.S. & Stevens, S.S. The binaural summation of loudness. Journal of the Acoustical Society of America, 1960, 32, 1337-1344.

Richardson, L.F. & Ross, J.S. Loudness and telephone current. Journal of General Psychology, 1930, 3, 288-306.

Riesz, R.R. Differential sensitivity of the ear for pure tones. Physics Review, 1928, 31, 867-875.

Robinson, G.H. Biasing power law exponents by magnitude estimation instructions. Perception & Psychophysics, 1976, 19, 80-84.

Roskam, E. A survey of the Michigan-Israel-Netherlands integrated series. In J.C. Lingoes (Ed.), Geometric representations of relational data. Ann Arbor, Mich.: Mathesis Press, 1977.

Rule, S.J. & Curtis, D.W. Levels of sensory and judgmental processing: strategies for the evaluation of a model. In B. Wegener (Ed.), Social attitudes and psychophysical measurement. Hillsdale, NJ: Lawrence Erlbaum Associates, 1982.

Rule, S.J., Curtis, D.W., & Mullin, L.C. Subjective ratios and differences in perceived heaviness. Journal of Experimental Psychology: Human Perception and Performance. 1981, 7, 459-466.

Scharf, B. & Fishken, D. Binaural summation of loudness: Reconsidered. Journal of Experimental Psychology, 1970, 86, 374-379.

Schneider, B. Individual loudness functions determined from direct comparisons of loudness intervals. Perception & Psychophysics, 1980, 28, 493-503.

Schneider, B. The nonmetric analysis of difference judgments in social psychophysics: Scale validity and dimensionality. In B. Wegener (Ed.), Social attitudes and psychophysical measurement. Hillsdale, NJ: Lawrence Erlbaum Associates, 1982.

Schneider, B., Parker, S., Farrell, G. & Kanow, G. The perceptual basis of loudness ratio judgments. Perception & Psychophysics, 1976, 19, 309-320.

Schneider, B., Parker, S., Ostrovsky, D., Stein, D. & Kanow, G. A scale for the psychological magnitude of number. Perception & Psychophysics, 1974, 16, 43-46.

Schneider, B., Parker, S. & Stein, D. The measurement of loudness using direct comparisons of sensory intervals. Journal of Mathematical Psychology, 1974, 11, 259-273.

Shepard, R.N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika, 1962, 27, 125-140. (a)

Shepard, R.N. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. Psychometrika, 1962, 27, 219-246. (b)

Shepard, R.N. Metric structures in ordinal data. Journal of Mathematical Psychology, 1966, 3, 287-315.

Spence, I. Local minimum solutions in nonmetric multidimensional scaling. Proceedings of the Social Statistics Section of The American Statistical Association, 1970, 13, 365-367.

Spence, I. A Monte-Carlo evaluation of three nonmetric multidimensional scaling algorithms. Psychometrika, 1972, 37, 461-486.

Stevens, J.C. & Guirao, M. Individual loudness functions. Journal of the Acoustical Society of America, 1964, 36, 2210-2213.

Stevens, S.S. A scale for the measurement of psychological magnitude: Loudness. Psychological Review, 1936, 43, 405-416.

Stevens, S.S. The measurement of loudness. Journal of the Acoustical Society of America, 1955, 27, 815-829. (a)

Stevens, S.S. On the averaging of data. Science, 1955, 121, 113-116. (b)

Stevens, S.S. On the psychophysical law. Psychological Review, 1957, 64, 153-181.

Stevens, S.S. To honor Fechner and repeal his law. Science, 1961, 133, 80-86.

Stevens, S.S. Perceptual magnitude and its measurement. In E.C. Carterette & M.P. Friedman (Eds.), Handbook of Perception, Vol.2, chap.11. New York: Academic Press, 1974.

Stevens, S.S. Psychophysics: Introduction to its perceptual, neural and social prospects. New York: Wiley & Sons, 1975.

Stevens, S.S. & Galanter, E. Ratio scales and category scales for a dozen perceptual continua. Journal of Experimental Psychology, 1957, 54, 377-411.

Stevens, S.S. & Greenbaum, H.B. Regression effect in psychophysical judgment. Perception & Psychophysics, 1966, 1, 439-446.

Teghtsoonian, R. Range effects in psychophysical scaling and a revision of Stevens' law. American Journal of Psychology, 1973, 86, 3-27.

Titchener, E.B. Experimental psychology: A manual for laboratory practice, Vol. II, Quantitative, Part 2, Instructor's Manual. New York: MacMillan, 1905.

Torgerson, W.S. Distances and ratios in psychological scaling. Acta Psychologica, 1961, 19, 201-205.

Treisman, M. & Irwin, R.J. Auditory intensity discriminal scale I. Evidence derived from binaural intensity summation. Journal of the Acoustical Society of America, 1967, 42, 586-592.

Young, F.W. TORSCA-9, a FORTRAN IV program for nonmetric multidimensional scaling. Behavioral Sciences, 1968, 13, 343-344.

Young, F.W. Nonmetric multidimensional scaling: Recovery of metric information. Psychometrika, 1970, 35, 455-473.

Zwislocki, J.J., Ketkar,I., Cannon, M.W. & Nodar, R.H. Loudness enhancement and summation in pairs or short sound bursts. Perception & Psychophysics, 1974, 16, 91-95.