NASA CONTRACTOR REPORT

NASA CR-2

NASA CR-2108

PERCEIVED NOISINESS UNDER ANECHOIC, SEMI-REVERBERANT AND EARPHONE LISTENING CONDITIONS

by Frank R. Clarke and Karl D. Kryter

Prepared by
STANFORD RESEARCH INSTITUTE
Menlo Park, Calif. 94025
for Langley Research Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • AUGUST 1972

1. Report No. NASA CR-2108	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Perceived Noisiness Under Anechoic, Semi-Reverberant and Earphone Listening Conditions		5. Report Date August 1972
		6. Performing Organization Code
7. Author(s) Frank R. Clarke and Karl D. Kryter		8. Performing Organization Report No.
		10. Work Unit No.
9. Performing Organization Name and Address Stanford Research Institute Menlo Park, California 94025		11. Contract or Grant No. NAS1-10017
		13. Type of Report and Period Covered Contractor Report
12. Sponsoring Agency Name and Address National Aeronautics & Space Administration Washington, DC 20546		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Magnitude estimates by each of 31 listeners were obtained for a variety of noise sources under three methods of stimuli presentation: loudspeaker presentation in an anechoic chamber, loudspeaker presentation in a normal semi-reverberant room, and earphone presentation. Comparability of ratings obtained in these environments were evaluated with respect to predictability of ratings from physical measures, reliability of ratings, and to the scale values assigned to various noise stimuli. Acoustic environment was found to have little effect upon physical predictive measures and ratings of perceived noisiness were little affected by the acoustic environment in which they were obtained. The need for further study of possible differing interactions between judged noisiness of "steady state" sound and the methods of magnitude estimation and paired comparisons is indicated by the finding that in these tests the subjects, though instructed otherwise, apparently judged the maximum rather than the "effective magnitude" of "steady-state" noises.

17. Key Words (Suggested by Author(s)) aircraft noise annoyance; subjective response to noise; noise	18. Distribution Statement unclassified - unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
unclassified	unclassified	29	$3.00

*For sale by the National Technical Information Service, Springfield, Virginia 22151

CONTENTS

	Page
LIST OF ILLUSTRATIONS	iv
LIST OF TABLES	v
INTRODUCTION	1
PROCEDURE	2
DATA ANALYSIS AND RESULTS	7
CONCLUSIONS	17
REFERENCES	19

LIST OF ILLUSTRATIONS

		Page
Figure 1	Power-Law Function for Standard Noise in Anechoic Test	20
Figure 2	Power-Law Function for Standard Noise in Anechoic Retest	20
Figure 3	Power-Law Function for Standard Noise in Reverberant Test	21
Figure 4	Power-Law Function for Standard Noise in Reverberant Retest	21
Figure 5	Power-Law Function for Standard Noise in Earphone Test	22
Figure 6	Power-Law Function for Standard Noise in Earphone Retest	22
Figure 7	Test-Retest Data in Anechoic Chamber	23
Figure 8	Test-Retest Data in Reverberant Room	23
Figure 9	Test-Retest Data with Earphones	23
Figure 10	Power-Law Function for Standard Noise in Anechoic Chamber (Test-Retest Combined)	24
Figure 11	Power-Law Function for Standard Noise in Reverberant Room (Test-Retest Combined)	24
Figure 12	Power-Law Function for Standard Noise with Earphones (Test-Retest Combined)	24
Figure 13	Correspondence between Results in Anechoic Chamber and Reverberant Room	25
Figure 14	Correspondence between Results in Anechoic Chamber and Earphones	25
Figure 15	Correspondence between Results in Reverberant Room and Earphones	25

LIST OF TABLES

		Page
Table I	Noises Employed in Experiment Comparing Psychophysical Results Obtained in Three Acoustic Environments	4
Table II	Order of Presentation of Tests and Experimental Conditions	5
Table III	Instructions	6
Table IV	Minimum Root-Mean-Square-Error Introduced by Playback Equipment and Acoustic Environment	9
Table V	Root-Mean-Square-Error of Prediction of Psychophysical Data Obtained in Anechoic Chamber Based on Physical Measures Taken From Original Test Tape	14
Table VI	Root-Mean-Square-Error of Prediction of Psychophysical Data Obtained in Reverberant Room Based on Physical Measures Taken From Original Test Tape	15
Table VII	Root-Mean-Square-Error of Prediction of Psychophysical Data Obtained with Earphones Based on Physical Measures Taken From Original Test Tape	15
Table VIII	Root-Mean-Square-Error of Prediction of Psychophysical Data Obtained in Anechoic Chamber Based on Physical Measures Taken *In Situ*	16
Table IX	Root-Mean-Square Error of Prediction of Psychophysical Data Obtained in Reverberant Room Based on Physical Measures Taken *In Situ*	16
Table X	Root-Mean-Square Error of Prediction of Psychophysical Data Obtained with Earphones Based on Physical Measures Taken From Recording Made Through Earphone and 6-cm^3 Coupler	17

PERCEIVED NOISINESS
UNDER ANECHOIC, SEMI-REVERBERANT AND EARPHONE LISTENING CONDITIONS

By Frank R. Clarke and Karl D. Kryter

Stanford Research Institute

INTRODUCTION

Different physical measures have provided varying degrees of predictive accuracy for assessing the effects of aircraft noises upon human judgments of annoyance. Data from the various experiments are remarkably consistent in providing a gross ordering of the effectiveness of the various physical measures. The best physical measures typically show root-mean-square-error of prediction on the order of 2 to 3 dB for predicting relative annoyance of various aircraft noises. For some purposes this degree of accuracy of prediction is more than sufficient. However, for standards that are used to determine the acceptance or lack of acceptance of a particular type of aircraft engine, 1 or 2 dB may be the determining factor between acceptance or rejection of extremely expensive aircraft and related research and development programs.

Any attempt to evaluate further or refine physical measures for predicting human annoyance judgments, must take into account potential sources of variability associated with various psychophysical methods as well as differences which may arise from various acoustic environments in which the judgments are obtained. Another report (ref. 1) from this project describes an experiment that compared the relative effectiveness of two psychophysical techniques: the method of paired comparisons and the magnitude estimation technique. The experiment reported herein was designed to evaluate different methods for presenting noise stimuli for obtaining human judgments of noise annoyance. Specifically, subjective judgments were obtained for a variety of noise sources in each of three environments: loudspeaker presentation in an anechoic chamber, loudspeaker presentation in a "normal" semi-reverberant room, and earphone presentation. Physical measures were obtained for four sets of tape recordings. The original test tapes which were played back over each of the three reproduction systems and recordings made at the output of each of the three reproduction systems, namely, in the anechoic chamber, in the reverberant room, and through a 6-cm^3 coupler driven by an earphone.

The experiments were designed to provide information relevant to four questions:

1. To what extent do the three playback systems with their

associated acoustic environments affect the physical spectra of the test noises, and in particular how does this affect the resultant predictive physical measures? Comparison of measures based on the original test tape with those made in the actual environments employed in the psychophysical tests will enable us to determine if the various acoustic environments contribute systematic errors to the noisiness ratings obtained therein.

2. Is test reliability affected by the acoustic environment in which the tests are conducted? For example, it is possible that changes in earphone placement during a test or between tests may affect the intensity and spectrum of the physical signal at the listener's ear. Small changes in head position in a reverberant room may dramatically affect the intensity of a "tonal" component of a noise. Such changes might be expected to differentially affect the reliability of test results as obtained in different acoustic environments. If the three acoustic environments differ in this respect, it should be reflected in different estimates of test-retest reliability and possibly by differential adequacy in prediction of the rating data from the physical measures made for that system.

3. To what extent are the various noises scaled the same by human observers in each of the three acoustic environments? The degree of variation among scaled values for the three environments will reflect both relative reliability and validity of test results obtained in these environments.

4. Although not a central point of the study, it was desired to obtain further information about the growth of perceived noisiness as a function of stimulus intensity level.

PROCEDURE

Experimental Subjects. Paid volunteers drawn from the community served as subjects in this experiment. Ages ranged from 14 to 67 with a median age of 22. Twenty-six of the subjects were female, five male. Fourteen of the thirty-one subjects had previous experience with magnitude estimation technique in an experiment conducted approximately three months earlier.

Acoustic Environments. Three acoustic environments were employed in this experiment: (1) an anechoic chamber; (2) a semi-reverberant or "typical" room; and (3) earphones.

The anechoic chamber had 21-inch long fiberglass wedges on all six surfaces. Measured from the tips of the wedges, the internal dimensions of the anechoic chamber were 8.5 by 17.75 by 8 ft. The noises to be judged were presented via two Altec-Lansing A7-500 speaker systems, each driven by an 80-watt McIntosh power amplifier. Conventional playback circuitry was employed with the exception of artificial quieting of the system noise between stimulus presentations and an equalization network designed to provide frequency responses as flat as possible at the eight listener positions within the room. Each speaker system was directed at four subjects seated in an arc of radius 8-1/2 feet. The chord of each arc was approximately 5 feet. The range of sound pressure levels of octave bands of noise with center frequencies ranging from 63 cycles to 8000 cycles did not exceed 5 dB at any listener position. A low pass filter with 3 dB downpoint at 8000 Hz was used to minimize tape hiss. As all signals were recorded on the test tapes at a constant value of dBD_2*, signal intensity was varied for the test by means of an attenuator in the circuit.

The relatively reverberant room was a "normal" room with plasterboard walls, tile floor, and plaster ceiling. The room was approximately 14 by 20 by 8 ft, and approximately one-half of the wall surface was covered by draperies. The listeners were seated from 9 to 12 ft from a Bozak B 302A speaker system in staggered rows of four each. The speaker system was driven by a McIntosh 50-watt amplifier and conventional circuitry. No attempt was made to compensate for speaker difficulties or variations in spectra at various locations in the room. As in the previous case, signal intensity was varied by means of an attenuator in the playback circuit.

Earphone presentation of the stimuli was accomplished in the reverberant room. The output of the 50-watt amplifier appropriately attenuated was delivered to eight sets of TDH-39 earphones with MX-41/AR cushions.

Noise Stimuli. Ten different noises were employed in this experiment: a standard noise and nine comparison noises. The standard noise was 12 seconds in duration with intensity rising at the rate of 2-1/2 dB

---

*References for the definitions of dBD_2 and other physical measures of noises used are given in a later section of this report.

per second for four seconds, unchanging for four seconds, and then falling at the rate of 2-1/2 dB per second for the remaining four seconds. The noises were obtained by shaping the output of a white noise generator such that it had a low frequency roll-off of 3 dB per octave and a high frequency roll-off of 6 dB per octave. Three dB downpoints were at 63 Hz and 500 Hz respectively. Comparison noises were selected to encompass a variety of spectra. They included three household appliances, a motorcycle, two jet aircraft, and three helicopter recordings. In the first four instances, intensity was relatively constant over duration of the stimulus. Aircraft noises and helicopter noises were begun and terminated at 15 dBD_2 downpoints from maximum intensity. The noise stimuli, their durations, and actual levels of presentation during test sessions are given in Table I.

Table I

NOISES EMPLOYED IN EXPERIMENT COMPARING PSYCHOPHYSICAL RESULTS OBTAINED IN THREE ACOUSTIC ENVIRONMENTS

Source	Duration	Levels dBD_2*
Vacuum Cleaner	20	83,93
Blender	23	81,91
Air-conditioner	19	84,94
Motorcycle (idling)	20	85,95
DC-9 (takeoff)	14	90,100
747 (takeoff)	16	86,96
Hughes C (flyby)	16	78,88
Hughes C (landing)	31	83,93
CH 47 (flyby)	9	89,99
Standard Noise	12	74-99**

*Levels in the table apply to the anechoic chamber and to earphone presentation. Levels in the reverberant room were approximately 3 dB lower.

**Six levels were employed, ranging from 74 to 99 dB in 5-dB steps.

Test Tapes and Order of Presentation. Six test tapes of 30 items each were constructed. The first half of Test Tape #1 contained each of the nine comparison stimuli and six renditions of the standard noise. The order of the 15 items was determined randomly. Pad settings were randomly assigned to each item such that the standard stimulus was presented at six levels spaced 5 dB apart and each of the comparison noises was designated either high or low. The second half of this tape had stimuli in the same order but a new random assignment of pad settings to the standard stimuli and comparison stimuli--high if they had been low in the first half of the test and vice versa. (High and low are relative terms specifying reproduction differing by 10 dB.) All test stimuli were recorded on the test tape at a constant dBD_2 level--pad settings were used to vary intensity during playback to the listeners. Actual test intensities are given in Table I. Test Tape #2 contained items from Test Tape #1 but in the order No. 6-15, 1-5, 21-30, 16-20, and Tape #3 had the same items in the order No. 11-15, 1-10, 26-30, 16-25. Each of the three test tapes had associated with it a retest tape with the item order reversed. The interval between items was five seconds, during which time the listener made his response and the item number for the subsequent stimulus was announced.

It was impractical to achieve complete counter-balancing of experimental conditions but partial counter-balancing was used as shown in Table II. Instructions were given to the subjects; they were taken to the appropriate test chamber; the test tape was presented; and after a few moments rest, the retest for that condition was administered.

Table II

ORDER OF PRESENTATION OF TESTS AND EXPERIMENTAL CONDITIONS

	Test Order					
Group	1	2	3	4	5	6
1	A(T1)	A(R1)	R(T2)	R(R2)	E(T3)	E(R3)
2	R(T3)	R(R3)	E(T1)	E(R1)	A(T2)	A(R2)
3	E(T2)	E(R2)	R(T1)	R(R1)	A(T3)	A(R3)
4	A(T2)	A(R2)	E(T3)	E(R3)	R(T1)	R(R1)

A = Anechoic chamber; R = Reverberant room; E = Earphone; (Ti) = ith. form of test; (Rj) = jth. form of retest.

Following a 15-minute break (and a change of locale if required) test and retest for the second experimental condition were administered and following a second 15-minute break, test and retest for the third experimental condition were administered.

The magnitude estimation technique was employed to elicit responses from the listeners. Table III shows the instructions which give a description of the method and an answer sheet. The standard presented at the beginning of each test (and retest) was the standard noise described above at an intensity level of 89 dBD_2.

Table III

Name ____________________________________ Date __________

Sex ______ Age _____ Session ______ Listening Position ________

INSTRUCTIONS

We are asking you to help us solve a problem concerned with noise: How noisy, annoying, or unwanted are various kinds of sound when heard in your home? You will be asked to give a score to each sound.

First, we will produce a sound whose noisiness score is 10. Use that sound as a standard, and judge each succeeding sound in relation to that standard. For example, if a sound seems twice as noisy as the standard, you will write 20 in the appropriate box on the answer sheet. If it seems only one-quarter as noisy, write 2.5. If it seems three times as noisy, write 30, and so on.

Please try to judge each sound carefully, and give it a score that tells how strong the annoyance seems to you. There are no right or wrong answers. The important thing is to say how you rate each of the sounds.

Be sure to judge each noise in its entirety. Listen to the full duration of the noise, not just to its peak level. Judge the relative annoyance of the noises in terms of how you would react to these noises were they to occasionally occur in or near your home from actual aircraft or other sources.

1. _____	11. _____	21. _____
2. _____	12. _____	22. _____
.	.	.
.	.	.
10. _____	20. _____	30. _____

DATA ANALYSIS AND RESULTS

Physical Analysis of the Stimuli. As previously noted, there were ten noises employed in this experiment: nine stimuli each presented at two intensity levels and a standard noise presented at six levels of intensity. Each of the six levels of the standard stimulus appeared twice on each test tape resulting in a total of thirty items. The noises as they appear on the test tape (without concern at this point with how these noises were obtained and recorded) are regarded as the stimuli for which psychophysical ratings are desired. We wished to determine the degree to which the three acoustic environments employed in this experiment systematically introduced deviations from the original spectra as recorded on the test tape. More specifically we were interested in the degree to which the physical predictive measures were affected, if at all, by the various reproduction systems involved.

Physical measures were obtained for the stimulus items for each of four recordings:

1. An original test tape.
2. A tape recording of the test stimuli as actually presented to the listeners in the anechoic chamber. This recording was obtained by placing a high-quality condenser microphone in the anechoic chamber at one of the listener's positions and recording the results of reproducing the test stimuli in this environment.
3. A tape recording obtained in a similar manner after reproduction of the test tape in the reverberant room.
4. A recording obtained for earphone presentation of the stimuli in the experiments as transduced by an earphone coupled to a 6-cm^3 coupler.

Physical measures of the noises on each of the four tapes were computed from one-third octave band sound pressure levels sampled and averaged over 1/2-second time intervals. A General Radio Type 1921 real-time analyzer was used to produce each 1/2-second sound pressure level measurement in 24 one-third octave bands covering a frequency range of 50 to 10,000 Hz. These data were recorded and processed in digital form. The end results of the analysis include the time histories of sound pressure levels in each of the 24 bands and the so-called Max and effective values of various weighted measures, dBA, dBC, dBD_2, PNdB, PNdBM, and PNdB- and PNdBM-corrected for tonal content by two procedures, and designated by the subscripts t1 and t2. These units and related frequency weightings and calculation procedures are given in refs. 2 and 3.

Matrices of 18 rows by 18 columns were constructed for each of the four analyses. The 18 rows corresponded to the nine test stimuli each as presented at two levels. The 18 columns corresponded to the 18 physical predictive measures employed in this experiment. Three difference matrices of the same dimensions were then computed, one for each of the three acoustic environments. These difference matrices were obtained by subtracting each entry in the matrix associated with the direct electrical playback of the test tape from corresponding entries in the matrices resulting from analysis of stimuli as reproduced in the three different acoustic environments. Thus each entry may be regarded as showing the change introduced in the physical spectrum by reproduction in a particular acoustic environment as reflected by each physical measure for each of the test stimuli. Minimum root-mean-square errors for each of these three difference matrices are shown in Table IV. These are not random perturbations introduced by the analysis process. Such random perturbations associated with independent recordings of the "same stimulus" are on the order of r.m.s. error of 0.3 dB for max measures, and 0.1 dB for effective measures. These data will be treated further in a later section of this paper.

Analysis of the Magnitude Estimation Data to Obtain Reliability Evaluation. For any given test item the raw data consisted of a magnitude estimate by each of the 31 subjects employed in this experiment. The geometric mean ($GM = [X_1 \cdot X_2 \ldots X_n]^{1/n}$) of these 31 responses constituted the basis for all further data analysis. Test and retest data for each of the three acoustic environments were analyzed separately. For each test and retest the function relating the geometric mean of the magnitude estimates to the intensity level of the standard stimulus was plotted. These functions are shown in figures 1 through 6 with the least square error fit of the power law function to these data. It can be seen generally that these fits are quite adequate. The exponents associated with these functions (ranging from .29 to .37) are in the range typically found in such experiments.

We define an Equivalent Standard Score for a given test item as the level in dB at which the standard noise must be presented to the subjects in order to result in a magnitude estimate equal to that obtained for the test item. For these conversions the fitted power functions depicted in figures 1 through 6 were used. For example, if the magnitude estimates assigned to a given stimulus item on the original test in the anechoic chamber had a geometric mean of 20, it is seen from figure 1 that the equivalent standard score assigned to the stimulus item would be approximately 21 dB (re an arbitrary zero in this case). Geometric means were converted to Equivalent Standard Scores for each of the nine test stimuli at each of two levels of presentation by using the appropriate

Table IV

MINIMUM ROOT-MEAN-SQUARE-ERROR
INTRODUCED BY PLAYBACK EQUIPMENT AND ACOUSTIC ENVIRONMENT

	Anechoic	Reverberant	Earphone
Max dBA	0.8	1.4	0.7
Max dBC	1.3	1.5	0.8
Max dBD_2	1.4	1.6	0.6
Max PNdB	1.3	1.4	0.7
Max $PNdB_{t1}$	3.1	1.8	1.2
Max $PNdB_{t2}$	1.8	1.4	1.3
Max PNdBM	1.5	1.4	0.8
Max $PNdBM_{t1}$	3.2	1.7	1.2
Max $PNdBM_{t2}$	2.0	1.4	1.3
Eff dBA	0.9	1.1	0.5
Eff dBC	1.4	1.1	0.8
Eff dBD_2	1.4	1.4	0.6
Eff PNdB	1.2	1.2	0.7
Eff $PNdB_{t1}$	1.6	1.1	0.6
Eff $PNdB_{t2}$	1.4	1.3	1.0
Eff PNdBM	1.3	1.2	0.6
Eff $PNdBM_{t1}$	1.7	1.2	0.7
Eff $PNdBM_{t2}$	1.4	1.2	1.0

power function (figures 1-6) for that test condition. Thus, magnitude estimates have in this manner been converted into a response variable having the decibel as a unit of measurement. It becomes meaningful to make comparisons of this Equivalent Standard Score with predictions based on various physical measures whose units are also the decibel. All geometric means for test items were converted to Equivalent Test Scores in this manner, and further discussion of psychophysical results will be limited to data expressed as Equivalent Standard Scores.

For the type of data collected in these experiments it is not reasonable to evaluate test-retest reliability by use of correlation techniques. There is no meaningful parent population of noises and intensity levels for which the test stimuli employed in these experiments might be considered a random sample. The noises employed in these experiments might be considered a random sample. The noises employed in these experiments and the levels at which they were presented were selected to provide reasonable variation in the spectra and a range of intensity levels of practical interest. The test-retest correlations that one might obtain in such experiments would depend heavily upon the range of chosen intensities and perhaps upon the diversity of noises employed in the experiments. Thus, if the noises were relatively homogeneous both with respect to spectra and intensity levels, the test-retest correlation would be close to zero even though test results might be quite precise. If the range of intensities were large, then the test-retest correlation might be quite high even in the case where the measures were quite imprecise.

A meaningful way to investigate the question of test-retest reliability in such a situation is to compute the r.m.s. error resulting from the attempt to predict the retest scores from the test scores. Scatter plots showing test scores plotted against retest scores for each of the 18 test stimuli in each of the three acoustic environments are shown in figures 7-9. The r.m.s. error for predicting retest scores from test scores is 1.65 dB in the anechoic chamber, 1.56 dB in the reverberant room, and 2.09 dB with earphone presentation. These results give an estimate of the order of magnitude of error to expect with the magnitude estimation technique as employed in this experiment.

There is reason to expect that the error observed from test to retest is not due entirely to chance factors, but in part to systematic error arising from differences in the sequence of items in the test tape and retest tape, and/or perhaps to the fact that listeners had relatively greater experience for the retest than for the test. If fluctuations from test to retest were purely attributable to chance factors, then the test score minus the retest score for each item should be a random variable, and the difference score obtained in one acoustic environment should be unrelated to the difference score obtained in a different environment. If, on the other hand, differences between test and retest scores were due in part to systematic context effects or to temporal effects, then difference scores in one environment should be related to those in a different environment. A Friedman two-way analysis of variance (ref. 4) for ranked scores rejects the null hypothesis that there is no difference among difference scores for the test items over the three environments at the 1% level of confidence. Thus, it would appear that the r.m.s. errors reported above are only in part due to chance error.

When test scores and retest scores are combined for comparing the three acoustic environments the r.m.s. error due to non-chance factors will drop out since all tests were used in all environments, and r.m.s. error arising from chance factors should drop by a factor of 0.7 (i.e., $\sqrt{1/2}$). Therefore, if two acoustic environments were for all practical purposes identical, test reliability is such that r.m.s. error in predicting results in one environment from those in the second environment should be something less than one to 1-1/2 dB.

Comparability of Data from Three Acoustic Environments. Figures 10 through 12 show the power law function obtained for the combined test and retest data in each of the three acoustic environments. For these figures, the data were combined and treated as described above. Once again the fits are quite reasonable, and the theoretical power function was used for converting geometric means for the combined test and retest data to Equivalent Standard Scores.

Comparisons of the Equivalent Standard Scores for test items as obtained in the three acoustic environments are shown in figures 13 through 15. In general, results are very similar regardless of acoustic environment. Root-mean-square error in predicting test results in the reverberant room given test results in the anechoic chamber is 1.67 dB; for predicting results with earphones on the basis of data obtained in the anechoic chamber, r.m.s. error is 1.7. In both of the preceding cases the playback systems employed were different in detail. The system used in the anechoic chamber was generally superior to that used in the reverberant room and for earphone presentation. In comparing results obtained in the reverberant room to those obtained with earphones, we find an r.m.s. error of 0.99. In this case the playback system used was identical with the exception of the final electroacoustic transducer. An r.m.s. error of 0.99 is approximately equal to our previous error estimate due to the inherent reliability of the combined test-retest data. Root-mean-square error of magnitude 1.7 dB is probably greater than might be expected on this basis alone. In both cases where an r.m.s. error of this size was obtained, a great part of the contribution to this error camefrom a single noise-stimulus--the household blender which had particularly intense tonal elements and might be strongly affected by microphone placement. Certainly, the differences between the results obtained in the three acoustic environments employed in this study are relatively small compared to the r.m.s. error typically reported in predicting noisiness ratings from physical measures of the stimuli. As will be seen shortly, however, they are not negligible when compared to the results obtained in this study.

Prediction of Noisiness Ratings from Physical Measures. Here we shall consider four sets of physical data. The data arising from analysis of noises directly from the test tapes and the data arising from analyses of the three different recordings made in the different acoustic environments. We wish to compare these physical predictive measures to the obtained rating data of which there are three sets--those obtained in the anechoic chamber, those obtained in the reverberant room, and those utilizing earphones for electroacoustic transduction. To make these comparisons the Equivalent Standard Scores cannot be based on an arbitrary zero. The arbitrary zero in the previous figures corresponded to the intensity levels of the least intense of the standard stimuli. Thus, for any given physical measure the Equivalent Standard Scores re arbitrary zero may be converted to an absolute level by obtaining the value of the physical measure associated with the least intense of the standard stimuli (in the appropriate environment) and adding that value to obtain the Equivalent Standard Scores re the same zero point as used for the physical measure. This procedure is appropriate for the relatively high levels of presentation of the stimuli used in this experiment as all physical measures tracked dB for dB with changes in pad settings associated with reproduction of the standard noise.

The initial analysis of the data for all nine noise stimuli (at two levels each) gave results surprising in three respects: (1) measures based on maximum values of the physical predictors appeared generally more accurate as predictors than those based on integrated values of the physical predictors; (2) there was little difference between prediction of psychophysical data from measures based on in situ recordings and recordings and measures based on the original test tape; and (3) there was less differentiation than commonly observed among the physical measures in terms of their effectiveness in predicting the rating data. These unexpected results were partially resolved by an item-by-item analysis of the data. Data based on the aircraft noises as a class differed from data based on the "steady-state" noises. In general, for the aircraft noises (which started at 15 dB below their maximum intensity level and then dropped 15 dB following maximum level) effective predictive measures were superior to maximum predictive measures. The opposite was true for the steady-state stimuli.

We recognize two possible bases for this result: (1) the standard noise with its varying intensity contour led to comparable judgments for aircraft noises with similar varying intensity contour but was not comparable to the steady-state noises, and (2) aircraft noises (takeoffs, landings and flyovers) have a natural duration and properly instructed listeners may take this into account when judging resulting annoyance, whereas the steady-state noises employed in this study have a duration that is completely at the discretion of the user. This may have been recognized by the listeners so that they may have merely regarded the test stimuli as samples of noises having arbitrary duration but meaningful intensity, and in consequence they may have judged the annoyance due to the intensity of the noises without regard to the duration. However, there have been conflicting findings, both in our laboratory and in others, on the relative effectiveness of maximum and effective measures as predictors of perceived noisiness of aircraft noises. The experiment reported herein does not provide a basis for resolving this issue.

Tables V, VI, and VII illustrate the above dichotomy. For aircraft noises considered as a class the effective (integrated) physical measure is superior to the maximum measure in 24 of 27 cases. For the steady-state noises the maximum measure is superior in all 27 cases.

Tables V through VII show r.m.s. error of predictions resulting from physical measures based on the original test tape. Similar data is shown in Tables VIII through X for predictions resulting from physical measures based on recordings made in situ. Both sets of data are very similar though on the average, predictions based on the original test tape are slightly better than those based on recordings made in situ.

Perhaps the original test tapes are a better indication of the average noise spectra occurring at listener positions than are the in situ recordings made at a single listening position. In any event, the differences are slight and there is no indication that the measured differences between original and reproduced noises reported in Table III have any effect upon accuracy of prediction of psychophysical results.

Restricted Range of Spectra of Aircraft Noises. There is a remarkable lack of variability in predictive power among the various physical measures. This cannot be attributed to the hypothesis that listeners are not responding to spectral features of the noises, for r.m.s. error is quite small for the more sophisticated measures and particularly low for data obtained through earphone presentation. Indeed the smallness of the r.m.s. errors for all the physical units, including that of dBC, appears to make the present results somewhat anomalous. Most likely, the constricted range can be attributed to the particular sample of aircraft noises employed, there being only five in all--three helicopter and two jet aircraft.

Growth of Noisiness as a Function of Intensity Level. Each of the nine test stimuli employed in this experiment were presented at two intensity levels. It is possible to obtain an estimate of the change in intensity level required for a doubling of perceived noisiness for each stimulus. Averaging over the results obtained in the three acoustic environments the following results were obtained: (1) the steady-state noises required on the average an increase in intensity level of 8.6 dB for a doubling of rated noisiness with a range from 7.2 dB to 10.8 dB; (2) the aircraft noises required on the average an increase of intensity level of 9.9 dB for a doubling of rated noisiness with a range from 8.4 dB to 11.7 dB. These data are referenced and discussed in a companion report (ref. 1).

Table V

ROOT-MEAN-SQUARE-ERROR OF PREDICTION OF PSYCHOPHYSICAL DATA OBTAINED IN ANECHOIC CHAMBER BASED ON PHYSICAL MEASURES TAKEN FROM ORIGINAL TEST TAPE

	Aircraft Noises		"Steady-State" Noises	
	Max	Eff	Max	Eff
dBA	2.3	1.8	1.7	3.9
dBC	4.2	3.5	3.3	5.0
dBD_2	2.0	1.5	1.6	4.2
PNdB	2.6	1.6	1.2	3.7
$PNdB_{t1}$	4.4	2.6	2.4	4.0
$PNdB_{t2}$	4.3	3.1	2.3	5.6
PNdBM	2.3	1.5	1.2	3.8
$PNdBM_{t1}$	4.1	2.5	2.7	4.1
$PNdBM_{t2}$	4.0	2.7	2.5	5.7

Table VI

ROOT-MEAN-SQUARE-ERROR OF PREDICTION OF PSYCHOPHYSICAL DATA
OBTAINED IN REVERBERANT ROOM
BASED ON PHYSICAL MEASURES TAKEN FROM ORIGINAL TEST TAPE

	Aircraft Noises		"Steady-State" Noises	
	Max	Eff	Max	Eff
dBA	2.3	2.6	2.2	3.4
dBC	3.4	2.4	3.1	4.2
dBD_2	2.1	2.3	2.3	3.8
PNdB	2.0	1.6	2.0	3.3
$PNdB_{t1}$	3.2	1.7	2.9	3.6
$PNdB_{t2}$	3.1	2.0	2.4	5.1
PNdBM	1.7	1.8	2.1	3.4
$PNdBM_{t1}$	2.9	1.8	3.2	3.7
$PNdBM_{t2}$	2.8	1.7	2.7	5.3

Table VII

ROOT-MEAN-SQUARE-ERROR OF PREDICTION OF PSYCHOPHYSICAL DATA
OBTAINED WITH EARPHONES
BASED ON PHYSICAL MEASURES TAKEN FROM ORIGINAL TEST TAPE

	Aircraft Noises		"Steady-State" Noises	
	Max	Eff	Max	Eff
dBA	2.3	2.0	2.4	3.9
dBC	3.8	2.9	2.6	4.2
dBD_2	1.9	1.7	2.9	4.4
PNdB	2.1	1.3	2.4	3.8
$PNdB_{t1}$	3.6	1.9	3.8	4.1
$PNdB_{t2}$	3.5	2.3	3.1	5.6
PNdBM	1.9	1.4	2.6	4.0
$PNdBM_{t1}$	3.2	1.9	4.1	4.3
$PNdBM_{t2}$	3.2	2.0	3.4	5.8

Table VIII

ROOT-MEAN-SQUARE-ERROR OF PREDICTION OF PSYCHOPHYSICAL DATA OBTAINED IN ANECHOIC CHAMBER BASED ON PHYSICAL MEASURES TAKEN *IN SITU*

	Aircraft Noises		"Steady-State" Noises	
	Max	Eff	Max	Eff
dBA	2.9	1.7	1.7	4.5
dBC	5.1	3.9	4.7	6.0
dBD_2	3.1	1.9	1.6	4.7
PNdB	3.1	2.4	2.2	4.4
$PNdB_{t1}$	3.4	3.2	2.4	4.6
$PNdB_{t2}$	3.6	2.8	2.1	5.3
PNdBM	3.0	2.3	2.0	4.5
$PNdBM_{t1}$	3.2	2.8	2.1	4.7
$PNdBM_{t2}$	3.4	2.6	2.0	5.4

Table IX

ROOT-MEAN-SQUARE-ERROR OF PREDICTION OF PSYCHOPHYSICAL DATA OBTAINED IN REVERBERANT ROOM BASED ON PHYSICAL MEASURES TAKEN *IN SITU*

	Aircraft Noises		"Steady-State" Noises	
	Max	Eff	Max	Eff
dBA	2.7	2.4	2.0	3.0
dBC	3.5	2.1	4.4	4.2
dBD_2	2.8	2.1	1.3	3.4
PNdB	2.4	1.8	2.1	3.3
$PNdB_{t1}$	2.9	2.0	2.0	3.9
$PNdB_{t2}$	2.8	1.8	1.6	4.3
PNdBM	2.6	2.1	2.2	2.9
$PNdBM_{t1}$	2.4	1.7	2.0	3.6
$PNdBM_{t2}$	2.2	1.7	1.5	4.0

Table X

ROOT-MEAN-SQUARE ERROR OF PREDICTION OF PSYCHOPHYSICAL DATA OBTAINED WITH EARPHONES BASED ON PHYSICAL MEASURES TAKEN FROM RECORDING MADE THROUGH EARPHONE AND 6-cm^3 COUPLER

	Aircraft Noises		"Steady-State" Noises	
	Max	Eff	Max	Eff
dBA	2.5	2.0	2.1	3.4
dBC	3.9	3.1	3.3	4.2
dBD_2	2.2	1.7	2.3	3.9
PNdB	2.3	1.6	2.0	3.5
$PNdB_{t1}$	3.6	2.2	2.7	3.9
$PNdB_{t2}$	2.2	1.5	2.2	3.1
PNdBM	2.2	1.5	2.1	3.8
$PNdBM_{t1}$	3.4	2.2	3.1	4.2
$PNdBM_{t2}$	1.9	1.6	2.3	3.4

CONCLUSIONS

1. Compared to the original signal as recorded on magnetic tape, the earphone presentation of the stimuli resulted in significantly less spectral distortion than that resulting from loudspeaker presentation in the anechoic chamber or in the semi-reverberant room. The anechoic chamber and the semi-reverberant room did not differ significantly in spectral distortion. Root-mean-square-error for the physical measures as introduced by acoustic environment were typically less than 1 dB for earphone presentation and between 1 and 2 dB for the other environments. However, the reliability of earphone presentation can be adversely affected in their use by the listeners, i.e., variability in placement of the earphones on their heads. These differences did not affect the predictive accuracy of the physical measures--predictions were equally good whether based on the original test tape or on recordings made *in situ*.

2. Test-retest reliability of the magnitude estimation technique as employed in this study is sufficiently good to provide rating data having an estimated r.m.s. error attributable to chance factors of

less than 1.0 dB for loudspeaker presentation of stimuli and less than 1.5 dB for earphone presentation of stimuli.

3. Test results obtained in the three acoustic environments are equivalent. Data obtained with earphones predicted those obtained in the semi-reverberant room with an r.m.s. error of 1 dB. Data obtained in the anechoic chamber were predicted from that obtained in either of the other acoustic environments with an r.m.s. error of 1.7 dB.

4. *Although the finding was somewhat unexpected and the experiment was* not designed to explore the phenomenon, it appears from this study that subjects tend to judge the magnitude of maximum noisiness rather than the total effective noisiness of "steady state." Further investigation of this finding is needed.

REFERENCES

1. F. R. Clarke, and K. D. Kryter, "The Methods of Paired Comparisons and Magnitude Estimation in Judging the Noisiness of Aircraft." NASA CR-2107, 1972, prepared by Stanford Research Institute under Contract No. NAS1-10017 for NASA Langley Research Center, Hampton, Virginia.

2. K. D. Kryter, "Possible Modifications to the Calculation of Perceived Noisiness." NASA CR-1636, 1970, prepared by Stanford Research Institute under Contract No. NAS1-6885 for NASA, Washington, D.C.

3. K. D. Kryter, "The Effects of Noise on Man." Academic Press, New York, New York (1970).

4. S. Siegel, "Nonparametric Statistics." McGraw-Hill Book Company, New York, New York (1956).

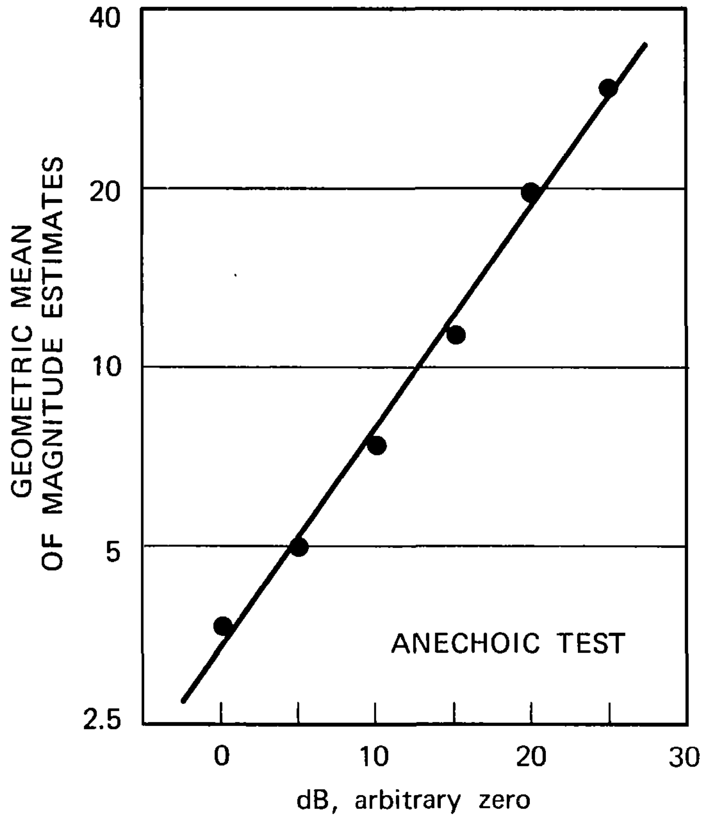

FIGURE 1 POWER-LAW FUNCTION FOR STANDARD NOISE IN ANECHOIC CHAMBER

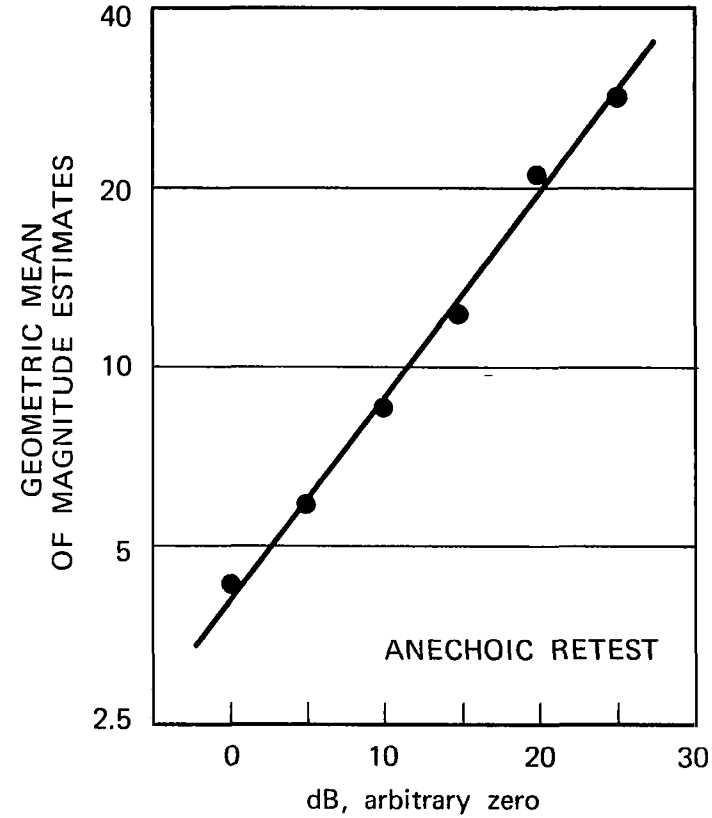

FIGURE 2 POWER-LAW FUNCTION FOR STANDARD NOISE IN ANECHOIC RETEST

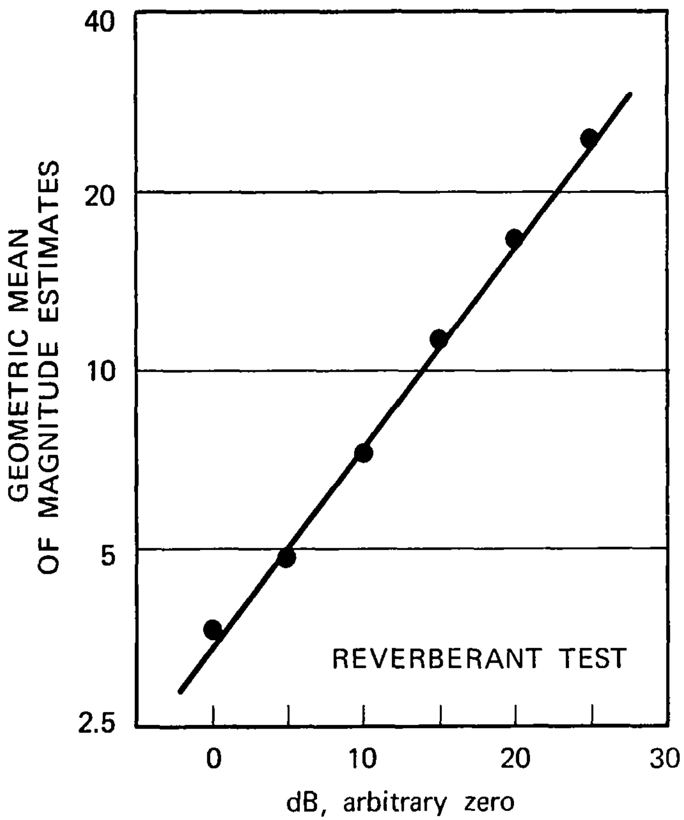

FIGURE 3 POWER-LAW FUNCTION FOR STANDARD NOISE IN REVERBERANT TEST

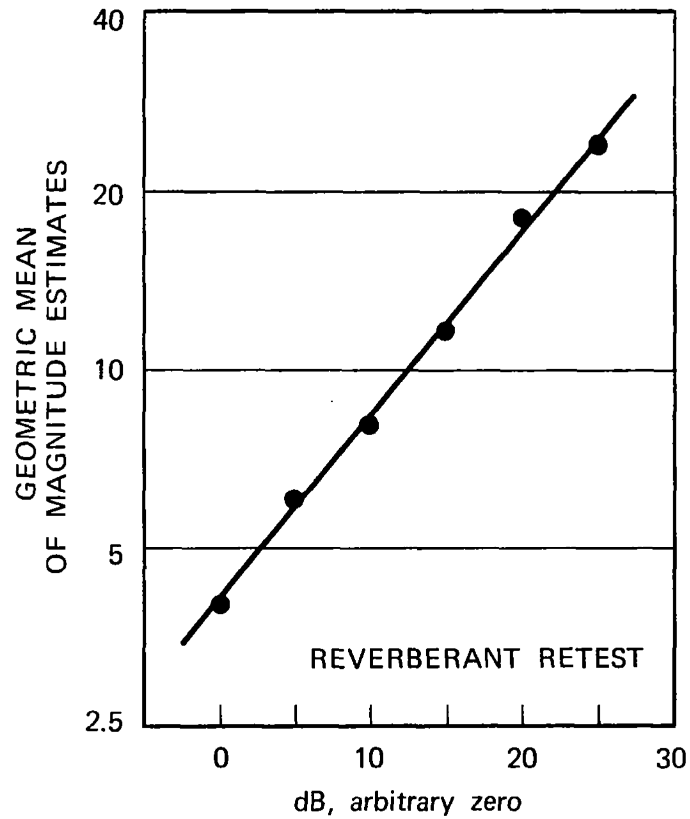

FIGURE 4 POWER-LAW FUNCTION FOR STANDARD NOISE IN REVERBERANT RETEST

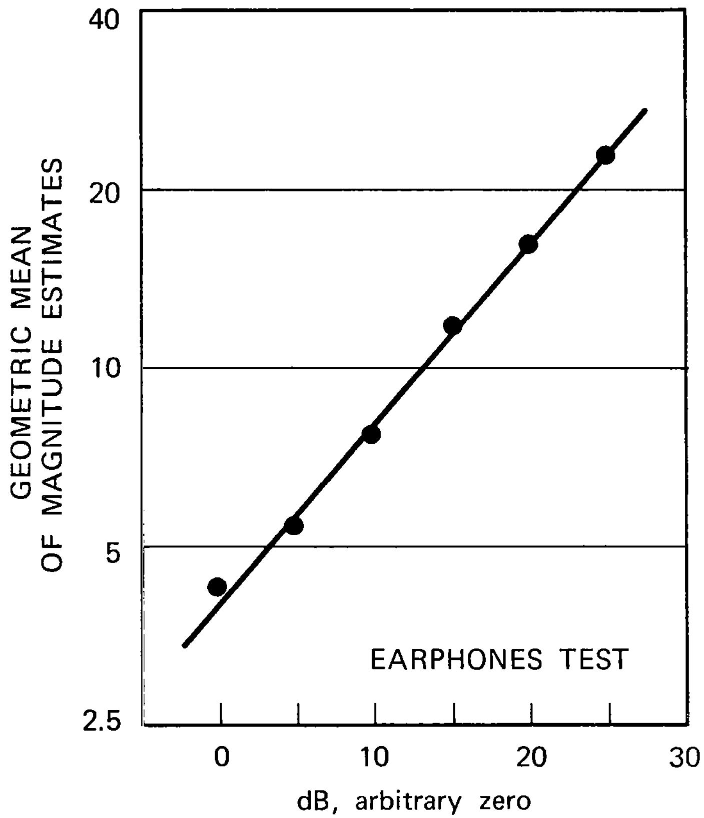

FIGURE 5 POWER-LAW FUNCTION FOR STANDARD NOISE IN EARPHONE TEST

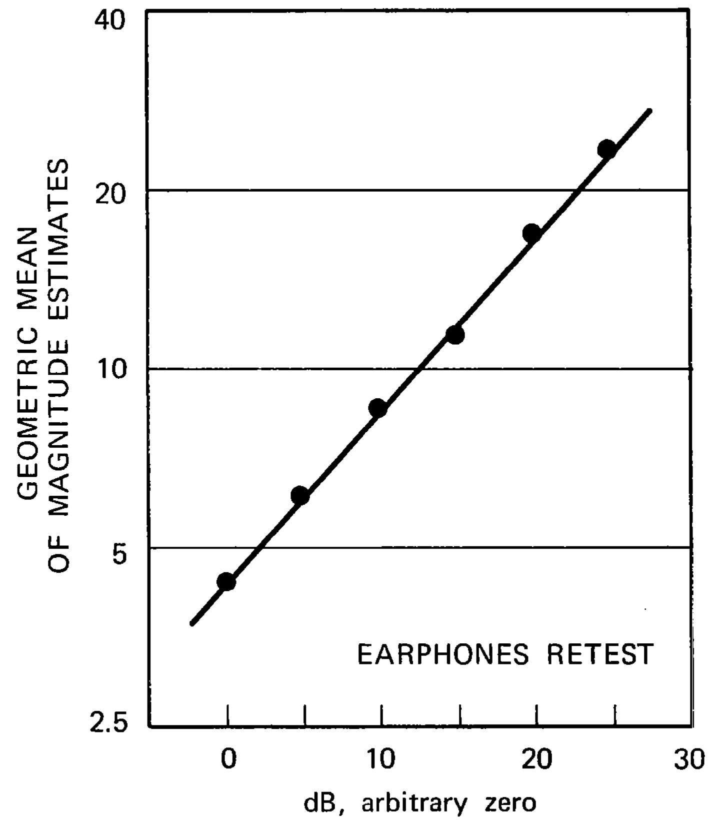

FIGURE 6 POWER-LAW FUNCTION FOR STANDARD NOISE IN EARPHONE RETEST

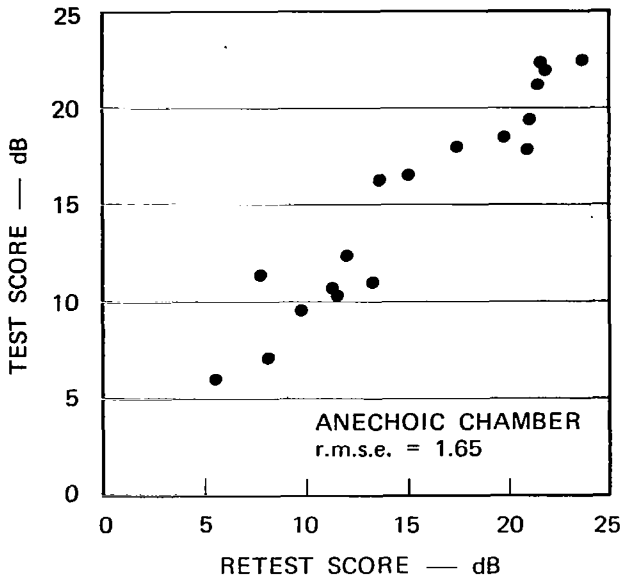

FIGURE 7 TEST-RETEST DATA IN ANECHOIC CHAMBER

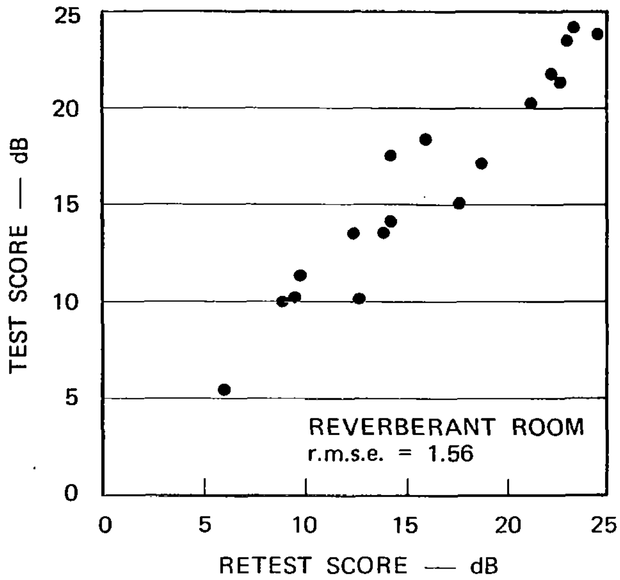

FIGURE 8 TEST-RETEST DATA IN REVERBERANT ROOM

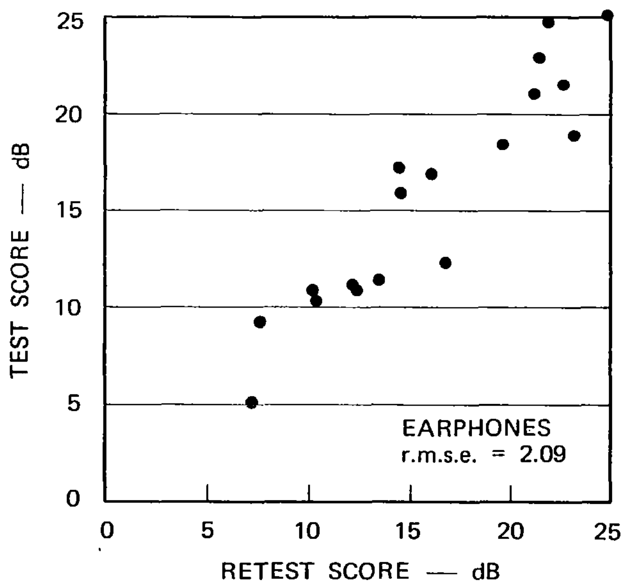

FIGURE 9 TEST-RETEST DATA WITH EARPHONES

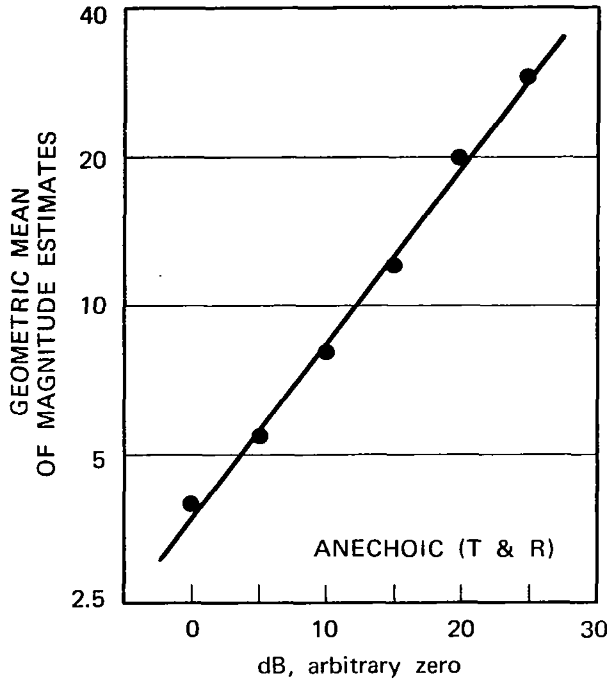

FIGURE 10 POWER-LAW FUNCTION FOR STANDARD NOISE IN ANECHOIC CHAMBER (Test-Retest Combined)

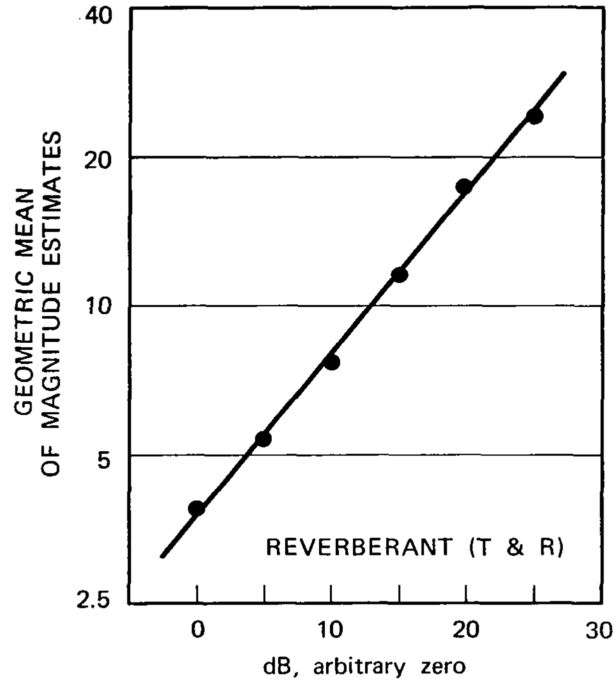

FIGURE 11 POWER-LAW FUNCTION FOR STANDARD NOISE IN REVERBERANT ROOM (Test-Retest Combined)

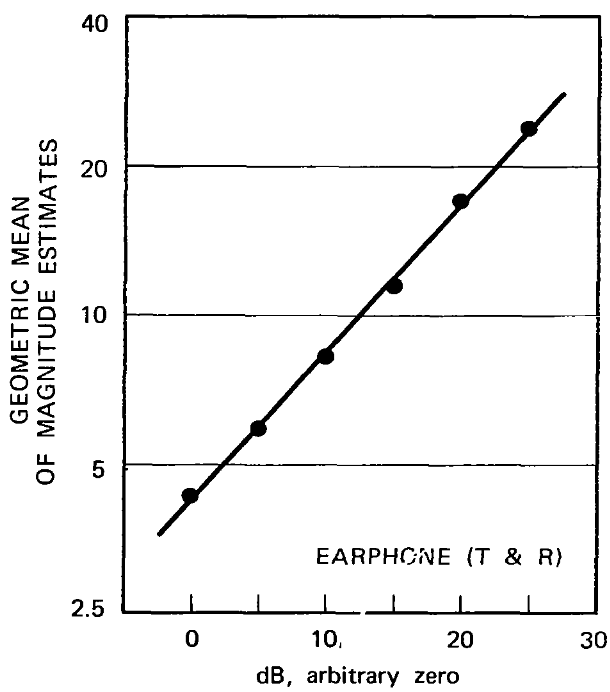

FIGURE 12 POWER-LAW FUNCTION FOR STANDARD NOISE WITH EARPHONES (Test-Retest Combined)

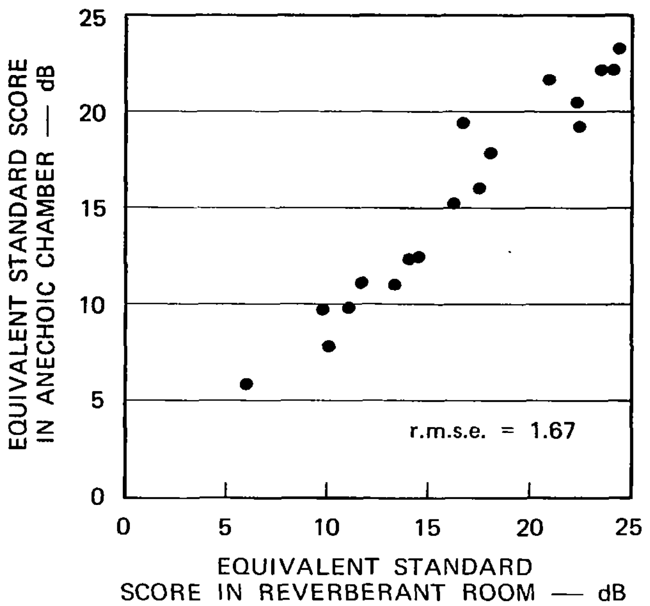

FIGURE 13 CORRESPONDENCE BETWEEN RESULTS IN ANECHOIC CHAMBER AND REVERBERANT ROOM

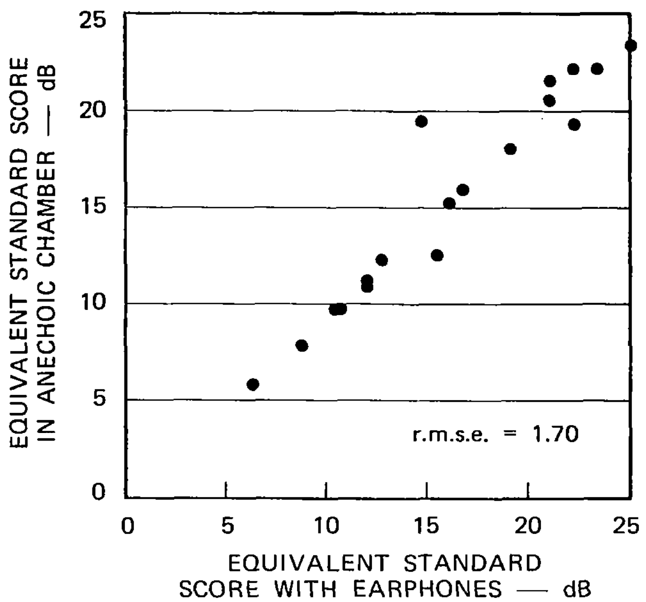

FIGURE 14 CORRESPONDENCE BETWEEN RESULTS IN ANECHOIC CHAMBER AND EARPHONES

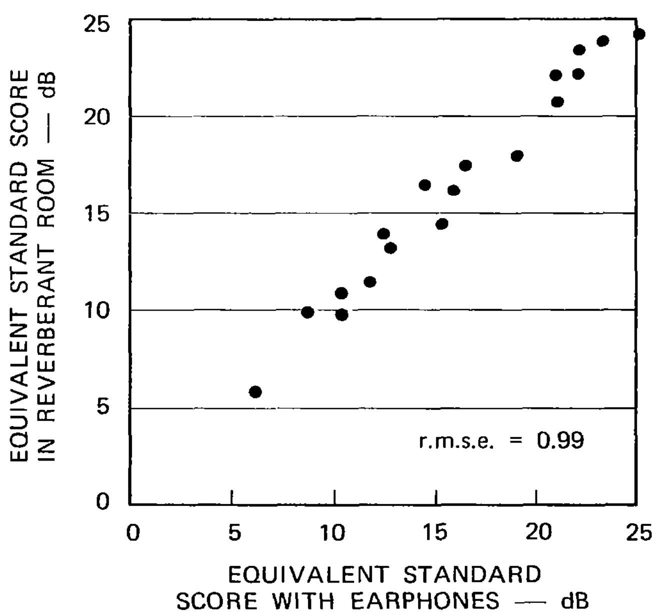

FIGURE 15 CORRESPONDENCE BETWEEN RESULTS IN REVERBERANT ROOM AND EARPHONES