BOLT BERANEK AND NEWMAN INC

CONSULTING • DEVELOPMENT • RESEARCH

Report 3699

N-96-01
II-A-363

THE RELATIONSHIP BETWEEN ANNOYANCE AND DETECTABILITY OF LOW LEVEL SOUNDS

Sanford Fidell

Sherri Teffeteller

September 1978

Submitted to:

Office of Noise Abatement and Control
Environmental Protection Agency
Washington, D.C. 20460
Contract Number 68-01-4491

B O L T   B E R A N E K   A N D   N E W M A N   I N C

C O N S U L T I N G · D E V E L O P M E N T · R E S E A R C H

Report 3699

THE RELATIONSHIP BETWEEN ANNOYANCE AND DETECTABILITY OF LOW LEVEL SOUNDS

Sanford Fidell

Sherri Teffeteller

September 1978

Submitted to:

Office of Noise Abatement and Control
Environmental Protection Agency
Washington, D.C. 20460
Contract Number 68-01-4491

ABSTRACT

The relationship between the predicted detectability and judged annoyance of 25 low level sounds heard in three noise backgrounds was investigated by an adaptive paired comparison procedure under free field listening conditions. The predicted detectability of the set of sounds accounted for almost 90% of the variance in the annoyance judgments in a conventional (falling spectrum) background noise environment. This strong relationship between predicted detectability and annoyance appears capable of supporting objective scales of the intrusiveness of low level sounds heard under everyday circumstances.

ACKNOWLEDGEMENTS

The authors have been greatly assisted in the design, conduct, and analysis of this study by Dr. David M. Green, Mr. Richard Horonjeff, and Mr. Karl Pearsons.

TABLE OF CONTENTS

I. INTRODUCTION

Annoyance produced by noise sources in the community correlates tolerably well for most purposes with integrated measures of physical exposure such as the Day-Night Sound Level (L_{dn}) over a range of values at the high end of commonly observed exposure conditions. There is little doubt, for example, that a ten decibel change in aircraft noise exposure in a community would be associated with a correspondingly large change in public reaction.

Community noise sources must be of high absolute level and long duration to generate L_{dn} values within this range, since low level and/or infrequent noise sources contribute little to long term L_{dn} values. Thus, the predictive usefulness of a measure like L_{dn} is greatest for sources such as transportation noise.

Many low level environmental noise intrusions seem to be disproportionately annoying, however. Heel clicks in apartment buildings, indistinct conversations, distant garbage compactors, and many other noises that neither materially affect L_{dn} nor cause speech or sleep interference may nonetheless create considerable annoyance. In fact, for noise sources with A-weighted levels below about 65 dB, community annoyance reactions are quite variable and do not appear to be sufficiently strongly related to levels of exposure to support confident prediction of annoyance or activity interference.

Instead, it appears that the degree to which low level or infrequent high level noises annoy people may be more closely related to the degree to which they intrude upon awareness. Systematic efforts to quantify the "intrusiveness" of low level noises have not been notably successful as yet. A number of factors seem to be loosely related to intrusiveness, including the familiar list of "correction factors" often invoked in accounting for unusual community reaction to noise: novelty, tonality, impulsiveness, and so forth (Fidell, 1978).

Until the "intrusiveness" of low level noises can be quantified more rigorously, the concept is of little value for assessment purposes. The current study was undertaken as an initial attempt to explore the utility of the psychophysical Theory of Signal Detectability as a predictor of the annoyance of intrusive noises. It was reasoned that if audibility (that is, bandwidth-corrected signal to noise ratio), rather than absolute level per se were closely related to annoyance of intrusive noises, then a theoretical framework might be available for a formal definition of intrusiveness. This approach explicitly focuses attention upon the role of the background noise in which sounds are heard as a partial determinant of annoyance.

If annoyance could be successfully predicted on the basis of detectability, a number of substantial benefits might follow. For example, detectability may be mathematically predicted from physical properties of noise sources, and directly measured without subjective judgments. Annoyance may not be so measured. The advantages of an objective definition of intrusiveness for assessment purposes include simplicity and ease of application, directness of interpretation, and straightforward manipulation.

Furthermore, detectability affords an absolute zero point on which a ratio scale of annoyance could in principle be built; it is irrefutable that people are not annoyed by noises they do not hear. The intrusiveness of disparate noise sources could also be defined on a common scale of detectability, rather than in limited empirical comparisons made in subjective tests.

The current experimentation was undertaken to investigate preliminary hypotheses about the relationship between detectability and annoyance of low level signals. The overall goal of the study was to determine whether the relationship was sufficiently strong and orderly to support justifiable inferences about the intrusiveness of low level sounds. It was hoped that if such strong relationships were observed, it would be possible to offer a definition of "intrusiveness" in terms of the detectability index d' (Green and Swets, 1966).

II. TECHNICAL DISCUSSION

This section is intended to acquaint the reader in a general way with the quantities involved in predictions of acoustic detectability. The reader is referred to a basic text on the psychophysical Theory of Signal Detectability (e.g., Green and Swets, 1966) for detailed discussion.

The most important physical parameter for purposes of predicting detectability is the signal-to-noise ratio measured over a band of frequencies encompassing the signal energy. Most existing research concerns how signal-to-noise ratio influences masking when the noise is steady state and the signal is a brief sinusoidal pulse. How masking varies as a function of signal duration, frequency, and multiple component signals is also well understood and readily predicted, however.

For signals of finite duration observed in specified intervals of time, the detectability of the signal (or the masking effectiveness of the noise) is governed by the ratio of signal energy (E) to the noise power density (N_o), i.e., the noise power per cycle, often called the spectral level of the noise. For a single sinusoid in noise of short duration (e.g., about 1/10 sec.), the detectability index d' is approximately

$$d' \approx g(f)\, E/N_o \qquad \text{Eq. 1}$$

where g(f) is a constant that depends on frequency and d' is the detectability of the signal. A d' = 1 (sometimes called a threshold value) implies correct selection of the interval that contains a signal 76% of the time in a two-interval forced choice test. The function g(f) is about 1/10 when f = 1000 Hz, and is monotonic with frequency: g(250 Hz) = .15, g(2000 Hz) = .063, g(4000 Hz) = .025. Thus, the higher the signal frequency, the less noise power is needed to achieve a given level of masking.

A major difference between this body of research and the current problem of predicting the detectability of complex sounds is that such sounds are not of short duration, but are more or less continuous, or at least of prolonged duration. This difference has been explored and there are experimental studies (e.g., Fidell et al., 1974), indicating that a useful approach is to treat the signal as incoherent and of effective duration about 1/3 sec. Detectability may then be predicted as in Equation 2*:

$$d' = \eta\ (W)^{1/2}\ S/N \qquad \text{Eq. 2}$$

where d' is again the detectability index, η is an efficiency term (a constant for any given situation), W is the 1/3 octave bandwidth centered at the signal frequency, and S/N is the signal-to-noise ratio (ratio of powers) measured in the same 1/3 octave band. No empirical check of Eq. 2 was made in this study.

For a complex signal spectrum, there are separate detectability indices for each spectral region. The combination of these different detectabilities is still an unsettled issue. Two rules have been suggested.

The simpler is the peak or d'_{max} rule:

$$d'_c = \max\ (d_1,\ d_2,\ d_i\ \ldots\ d_n) \qquad \text{Eq. 3}$$

*Note that Eq. 2 is not intended for very short duration or impulsive signals, nor for signals composed primarily of pure tones.

where d'_c is the combined detectability and d_i is the detectability measured in each spectral region with a significant signal to noise ratio.

The parallel rule is an incoherent combination of the various bands and produces the following formula:

$$d'_c = (d^2_1 + d^2_2 + d^2_i + \dots . d^2_n)^{1/2} \qquad \text{Eq. 4}$$

The latter is also called a vector combination rule since it is like computing the magnitude of a vector composed of the sum of different vectors. Note that if one of the detectabilities, d'_{max}, is much greater than any of the others both rules will predict nearly the same value of d'_c.

Naive (untrained) observers, unfamiliar with a particular signal, may tend to focus their attention exclusively on the portion of its spectrum that is least masked by background noise. Their behavior may be best modelled by the d'_{max} rule. More experienced observers, who are very familiar with the signal to be detected, may be able to improve their detection performance by taking advantage of information in other spectral regions as well. Although artificial signals can be constructed for which there are large differences in predicted performance between the d'_{max} and $d'_{parallel}$ rules, such differences are rare in real world signals. In the current signal set, the mean difference in detectability as predicted by the peak and parallel rules was on the order of 2 dB.

III. METHOD

A. Signal Selection

The goal of signal selection was to maximize both the range of detectabilities of a set of signals and the differences between their A-levels and their predicted detectabilities. This goal was adopted to facilitate discrimination of traditional measures of annoyance from detection - theoretical measures. Complete differentiation of the two types of measures can never be accomplished, however, since in a constant background, a measure such as A-level increases to some degree as detectability increases.

Maximization of the *range of detectabilities* of the signals was achieved in two ways. First, the absolute levels at which the signals were heard by test subjects varied by about 30 dB, as may be seen in Table I. Second, all test signals were presented in three different background noise environments which had been spectrally shaped to mask the various signals differentially. Since the detectability of complex sounds is governed by signal to noise ratios in different spectral regions, presentation of the same sound in differently shaped backgrounds changes detectability without changing absolute signal levels. Three spectral shapes for background noise environments were selected to maximize differences in detectability of a set of signals of constant level: PNC-40 (a falling spectrum resembling everyday ambient noise environments), a flat spectrum, and a rising spectrum.

Maximization of the *relative difference* between A-level measurements and predicted detectabilities of the signals was accomplished

TABLE I. LIST OF SIGNALS AND PRESENTATION LEVELS

SIGNAL #	DESCRIPTION	A-WEIGHTED PRESENTATION LEVEL (dB)
1	Transformer	48.4
2	1 kHz Octave of Noise	variable
3	Blender	48.4
4	Egg Beater	52.9
5	Carving Knife	57.6
6	Jig Saw	58.4
7	Hair Dryer	62.9
8	Lathe	57.3
9	Router	60.7
10	Belt Sander	56.8
11	Hand Drill	55.7
12	Radial Arm Saw	66.6
13	Air Compressor	63.2
14	Model Airplane	59.5
15	Typewriter	58.5
16	Toy Car	53.4
17	Toy Dog	50.7
18	Vacuum Cleaner	60.0
19	Air Conditioner	51.9
20	Garbage Compactor	57.9
21	Train	76.3
22	Motorcycle	64.5
23	Automobile	63.3
24	Bus	62.4
25	Lawnmower	63.9

by judicious selection of 24 signals from a much larger set.* Approximately eighty common noise sources (home appliances, office equipment, toys, construction equipment, etc.) were recorded on magnetic tape and played at the subject's ear position in an anechoic chamber. Real time spectral analyses of the entire signal set were made in one-third octave bands and submitted to several computerized analyses.

The signal analysis software first adjusted the spectra of the signals mathematically to a constant A-level. It then computed predicted detectabilities (according to Eq. 2) in one-third octave bands from 50 Hz to 10 kHz for each signal in each background. Peak and parallel summations of each signal were calculated from the one-third octave band detectabilities according to Eqs. 3 and 4.

Signal selection was accomplished by analysis of differences in predicted detectabilities of each signal in each background. Rank orders of these differences were assigned to each signal in pairs of backgrounds. Signals were selected if their predicted detectabilities differed greatly in different pairs of backgrounds. Figure 1 shows the relative predicted detectabilities of the final set of twenty four signals in the three background noise environments.

---

*An attempt to maximize the differences between A-level measurements and relative detectabilities of a large set of signals by analytic techniques was abandoned for a variety of reasons. These included the small number of maximally different signals, the difficulty of synthesizing them, and the need for annoyance judgments of complex real world signals.

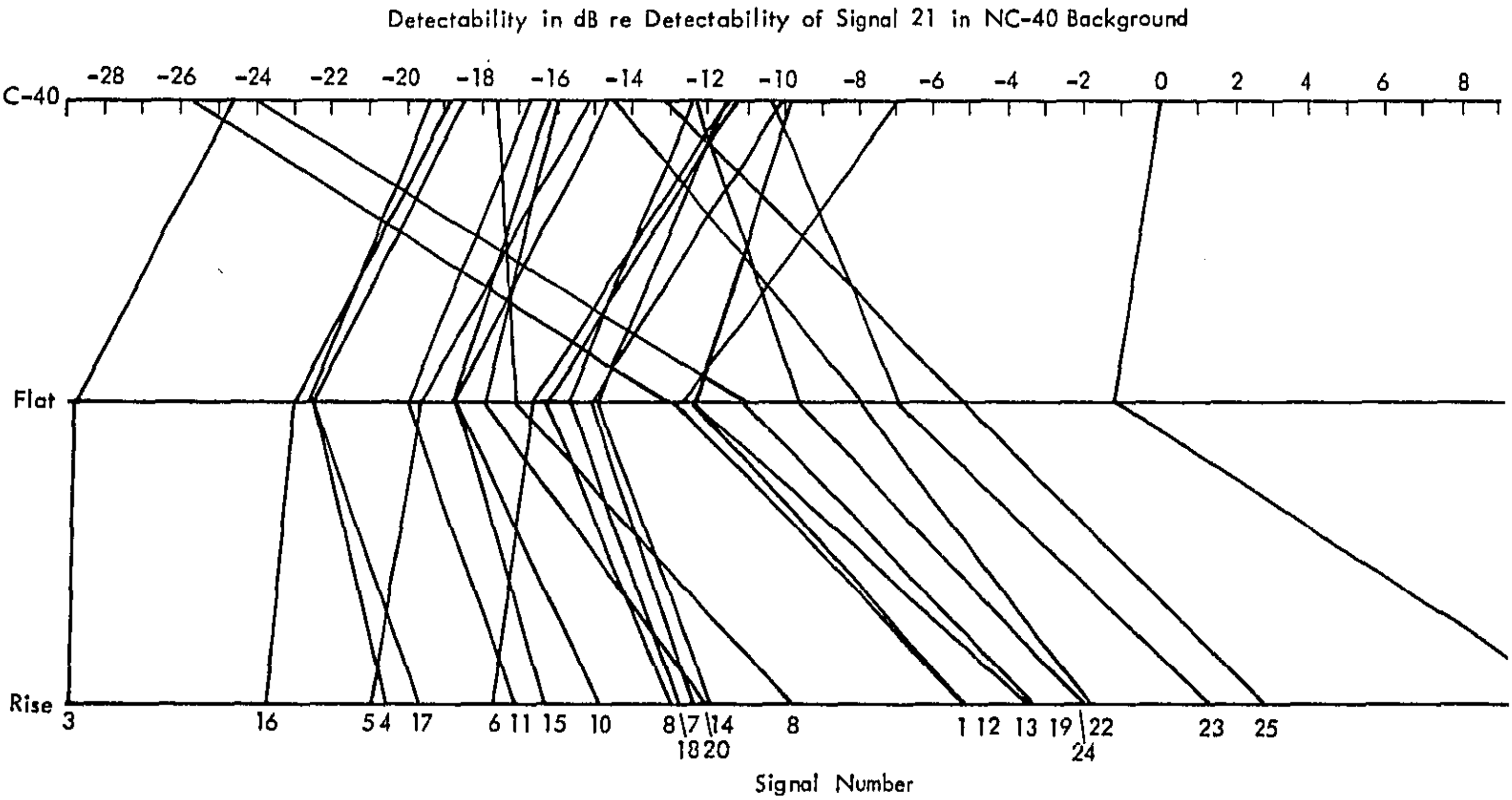


FIGURE 1. PREDICTED DETECTABILITIES OF 24 SIGNALS IN 3 BACKGROUND NOISE ENVIRONMENTS EXPRESSED IN dB RE d'_{max} OF SIGNAL 21 IN PNC-40 BACKGROUND FOR ALL CONSTANT A-LEVELS

B. Data Collection

1. Pilot Study

Thirteen audiometrically screened individuals were used as subjects. Each individually compared the annoyance of twenty four signals heard at fixed levels in an anechoic chamber with the annoyance of an octave band of noise centered at 1 kHz. The comparisons were made by a procedure known as Parameter Estimation by Sequential Testing (PEST), described in Appendix A. Briefly, the procedure required subjects to push a button corresponding to the more annoying of a pair of two sounds, each four seconds in duration. A laboratory computer controlling the equipment that generated the sounds adjusted the level of the 1 kHz band of noise until it determined that the sounds were equally annoying within 1 dB.

Instructions to test subjects may also be found in Appendix A. Subjects were trained in the trial procedures until their judgments of the annoyance of a signal compared with itself were no more deviant than 1 dB. This training usually required 20 or fewer paired comparison judgments. The training period was completed in about one half hour. Subjects were familiarized with all of the sounds in a given session before starting to make annoyance judgments.

The adaptive paired comparison judgments were made in sessions composed of six randomly intermingled "runs"; i.e., six separate determinations of points of subjective equality. Four of the six runs in a session were determinations of the point of subjective equality of signals with the 1 kHz band of noise. The other two were runs used to check the validity and reliability of the subject's judgments. Selection of the four signals heard in each session was counterbalanced over groups of six subjects to avoid potential sequential effects. Thus, five counterbalancings were employed for the 13 subjects. Spectra of the test signals may be found in Appendix B.

One of the check runs was a comparison of the annoyance of one of the four signals with itself. The other check run in each session was a repeat of one of the other runs. The former type of check run was termed a "self test" (for validity assessment) while the latter type of check run was termed a "test-retest" (for reliability assessment).

Six sessions were required to complete the testing of the 24 test signals in each of three noise backgrounds. Subjects were permitted short rest periods after each half hour of testing, and never made judgments in more than one noise background on a given day. The order in which subjects encountered background noise conditions was also counterbalanced. The background noise environment was always present at any time a test subject was in the anechoic chamber.

All of the signals were heard in each background noise environment. The A-weighted level of each of the three background noise environments was 50 dB. Additionally, the twelve odd numbered signals were heard in all three background noise environments 10 dB lower in level (40 dB(A)) and the twelve even numbered signals were heard in all three background noise environments 10 dB higher in level (60 dB(A)). The signal to noise ratios under all level conditions were constant, however, since the signal levels and the background noise levels were amplified or attenuated together.

2. Main Study

Thirty audiometrically screened subjects compared the annoyance of the twenty four test signals to the octave band of noise at 1 kHz in the three background noise environments at a fixed level of 50 dB(A). All procedures were identical to those of the pilot study; the only difference was that the annoyance judgments were collected at the signal levels of Table I in background noise environments of one level (50 dB(A)).

C. Instrumentation

Figure 2 is a schematic representation of the equipment used to generate and present test signals. There were four signal sources: specialized circuitry that generated phase locked harmonics of 60 Hz to simulate an electrical transformer; a band pass filtered noise generator; a cartridge magnetic tape machine; and a reel-to-reel magnetic tape deck.

Twenty three of the test signals were produced under computer control by the cartridge tape machine. The background noises were produced by the reel-to-reel machine. Signal conditioning circuitry (gates, electronic switches, attenuators, etc.) was used to minimize crosstalk, hum, and other extraneous noises, and to control the rise and decay times of all signals (250 msec.).

End-to-end electrical calibration of the computer-controlled interface was accomplished by daily monitoring of voltages produced across the loudspeaker terminals in the anechoic chamber by tones of known level. Acoustic re-calibration was performed at several times during the course of experimentation

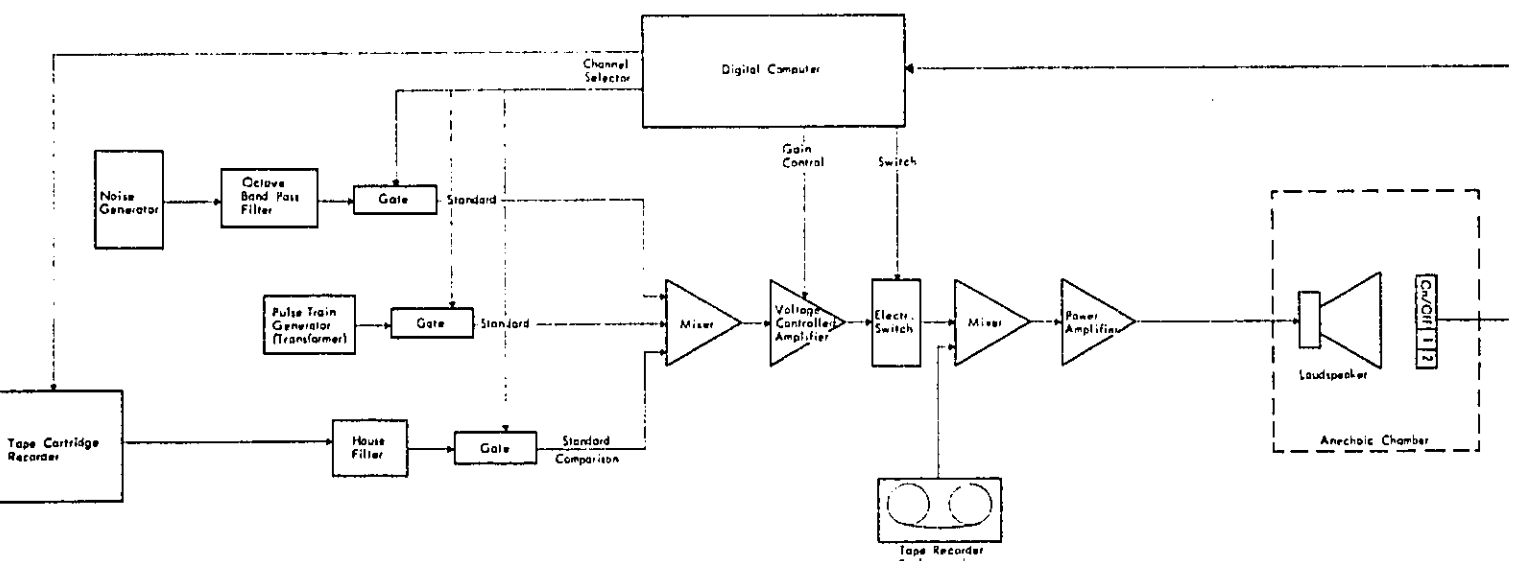


FIGURE 2. BLOCK DIAGRAM OF APPARATUS

as well. It is unlikely that random or systematic errors in levels greater than $\pm$ .5 dB could have occurred during the many weeks of data collecton.

IV. RESULTS

A. Pilot Study

Since the current study was among the first to concern itself with the annoyance of intrusive signals, one goal of the pilot study was to explore the sensitivity of annoyance judgments to the absolute level at which test signals were heard. If annoyance judgments depended critically upon absolute level, then one would expect the annoyance of the 24 test signals to change greatly as the presentation levels of the signals varied. If, on the other hand, annoyance judgments did not depend greatly upon absolute level, but upon relative level (the difference between signal and background noise levels), then one would expect little change in the relative annoyance of the set of 24 signals as the background level and signal levels varied together.

As outlined in the Method section, the effects of absolute level on annoyance judgments were assessed by presenting test signals at constant signal to noise ratios in background noise environments of three different absolute levels: 40 dB(A), 50 dB(A), and 60 dB(A). As the background noise level changed over this 20 dB range, so did the absolute levels of each test signal. Note that the range of levels within the set of 24 signals (28 dB) was unaffected by this manipulation. The highest signal level heard by subjects (signal 21 in the 60 dB(A) background condition) was 86.3 dB(A), while the lowest signal levels (signals 1 and 3 in the 40 dB(A) background condition) were 38.4 dB(A).

The basic measure of annoyance of each signal was the level to which an octave band of noise centered at 1 kHz was adjusted by PEST in response to the subjects' judgments. These annoyance values were averaged over all thirteen subjects within background noise spectra (PNC-40, flat, and rising) and background noise presentation levels (40, 50, and 60 dB(A)). Product-moment correlations were then computed for these averaged annoyance values between the 40 and 50 dB(A) presentation levels and the 60 and 50 dB(A) presentations. No comparison was possible between the 40 and 60 dB(A) presentation levels since the same signal was never heard in both background noise environments. These correlations may be seen in Table II.

TABLE II. PRODUCT-MOMENT CORRELATIONS BETWEEN MEAN LEVELS OF THE 1 kHz BAND OF NOISE AT THE POINT OF SUBJECTIVE EQUALITY WITH TWENTY FOUR SIGNALS IN THREE BACKGROUNDS

	BACKGROUND		
ABSOLUTE LEVEL	PNC-40	FLAT	RISING
50 dB(A) vs. 40 dB(A)	.93	.82	.88
50 dB(A) vs. 60 dB(A)	.92	.95	.91

The correlations seen in Table II are all significantly different from zero (i.e., unlikely to have arisen by chance alone) but not significantly different from one another. Their absolute values are so high that there can be little doubt that the absolute levels of the background noise environment had essentially no effect on annoyance judgments.

A major effect of presentation levels on the *variability* of annoyance judgments was observed, however. Figure 3 shows a strong inverse relationship between the standard deviations of

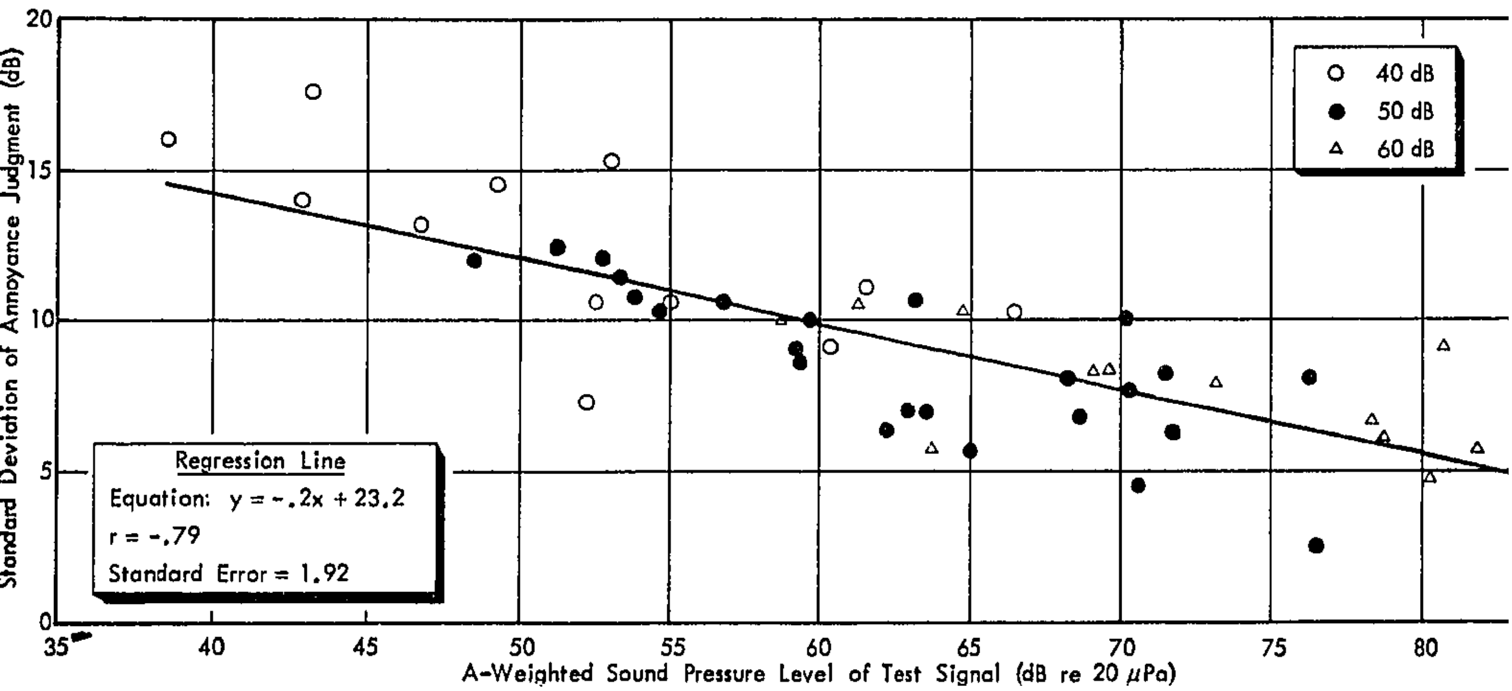


FIGURE 3. RELATIONSHIP BETWEEN ABSOLUTE PRESENTATION LEVEL OF SIGNAL AND STANDARD DEVIATION OF ANNOYANCE JUDGMENTS FOR THIRTEEN OBSERVERS IN PNC-40 BACKGROUND

the thirteen subjects' annoyance judgments and the signals' presentation levels: as the presentation level increased, the standard deviations of the subjects' judgments decreased. In other words, subjects agreed with one another far more about the annoyance of high level signals than they did about the annoyance of low level signals. Relationships similar to that seen in Figure 3 were observed in all three background noise spectra, and as a function of judged annoyance as well as presentation level.

B. Main Study

The thirty test subjects made over 3200 determinations of points of subjective equality of annoyance of 24 test signals with an octave band of noise centered at 1 kHz, including all training, validity, and reliability checks. The 720 judgments in each of three background spectra (2160 in toto) are discussed first.

1. Overview of Results

Table III displays sound pressure levels of the octave band of noise at the point of equal annoyance with each signal in each background. These figures are averaged over all subjects, and reported for the sake of simplicity in A-weighted units. Note that the annoyance of some signals changes little over backgrounds, while the annoyance of others changes considerably.

Since each signal was heard at the same level in all noise backgrounds, methods of predicting the annoyance of signals that are based exclusively on absolute levels would predict no significant change in annoyance from background to background. Methods of predicting the annoyance of signals that consider the

TABLE III ANNOYANCE OF 24 SIGNALS IN THREE BACKGROUND NOISE SPECTRA, EXPRESSED AS MEAN VALUES (FOR 30 SUBJECTS) OF THE LEVEL OF AN EQUALLY ANNOYING OCTAVE BAND OF NOISE AT 1 kHz IN dB(A)

	NC-40 BACKGROUND		FLAT BACKGROUND			RISING BACKGROUND		
SIGNAL	MEAN ANNOYANCE	STD. DEV.	MEAN ANNOYANCE	STD. DEV.	CHANGE FROM BKG. 1	MEAN ANNOYANCE	STD. DEV.	CHANGE FRO BKG. 1
1	58.9	10.4	69.1	9.1	+10.2	68.1	11.4	+9.2
3	55.5	9.9	51.8	8.0	- 3.7	52.4	13.8	-3.1
4	65.3	9.5	62.2	9.4	- 3.1	61.8	9.0	-3.5
5	74.7	8.4	72.5	10.1	- 2.2	71.4	9.8	-3.3
6	77.6	10.0	77.4	11.1	- 0.2	73.3	10.2	+4.3
7	79.2	7.1	79.7	6.7	+ 0.5	77.1	9.8	-2.1
8	67.3	10.5	69.2	10.5	+ 1.9	68.0	10.0	+0.7
9	75.3	9.9	74.2	7.3	- 1.1	71.6	9.5	-3.7
10	69.3	8.4	67.5	7.5	- 1.8	68.2	8.3	-1.1
11	70.5	9.5	66.6	10.0	- 3.9	66.2	8.3	-4.3
12	81.9	9.6	83.0	7.1	+ 1.1	83.3	7.6	+1.4
13	76.2	10.3	76.7	9.8	+ 0.5	76.6	9.4	+0.4
14	79.8	8.0	78.8	8.8	- 1.0	77.1	8.3	-2.7
15	67.7	12.7	69.4	11.7	+ 1.7	67.7	11.2	0.0
16	66.7	12.9	69.8	12.2	+ 3.1	65.9	13.2	-0.8
17	65.8	12.1	66.9	12.3	+ 1.1	66.1	14.0	+0.3
18	72.7	9.5	73.2	9.2	+ 0.5	71.1	8.1	-1.6
19	56.8	9.1	63.1	9.1	+ 6.3	59.3	8.3	+2.5
20	64.9	8.4	68.5	7.9	+ 3.6	65.8	9.4	+0.9
21	77.6	11.9	80.6	11.3	+ 3.0	80.3	11.7	+2.7
22	70.1	12.7	72.5	11.4	+ 2.4	69.7	10.6	-0.4
23	74.0	10.1	75.2	9.6	+ 1.2	75.6	10.0	+1.6
24	74.6	9.4	75.2	11.4	+ 0.6	74.7	10.1	+0.1
25	75.5	8.6	76.9	9.9	+ 1.4	75.2	8.9	-0.3

effects of background noises on the annoyance of sounds would, on the other hand, predict a specifiable pattern of changes in level. The predicted pattern of change and the observed pattern of change in level are most simply compared by a sign test. A non-parametric test of this sort is of low power, serving primarily as a screening tool to determine whether more detailed analyses are worthwhile.

The sign test is conducted by assigning a predicted direction of change in annoyance (no change (0), more annoying (+), or less annoying (-)) for each signal in the flat and rising backgrounds relative to its annoyance in the PNC-40 background. A tolerance of .2 dB was used for each category of prediction. The number of congruences between predicted and observed directions of change that would be expected by chance alone can then be compared with the observed number of congruences. If the number of congruences is significantly greater than would be expected by chance alone, it can be concluded that the background noise in which a signal is heard does indeed affect its annoyance.

Since the probability of obtaining by chance alone as many congruences as were actually observed was less than .02, it was concluded that background noise does influence annoyance judgments, and more detailed analyses were undertaken.

2. Correlational Analyses

Figures 4, 5, and 6 compare observed annoyance with annoyance predictions made by Eq. 2 and by A-level. The correlations between predicted and observed annoyance are also summarized in Table IV.

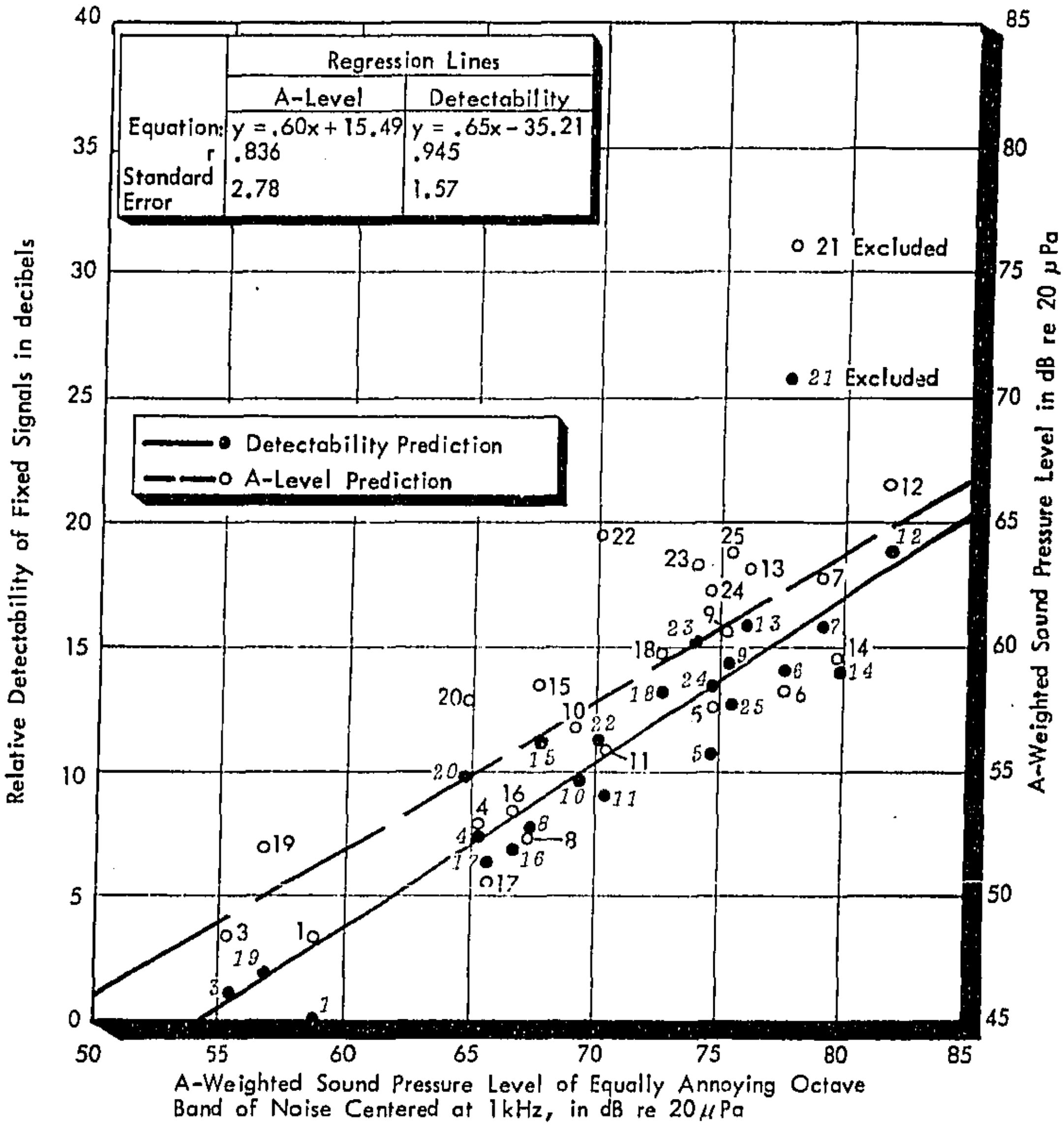


FIGURE 4. PREDICTED VS. OBSERVED ANNOYANCE OF TEST SIGNALS IN PNC-40 BACKGROUND NOISE SPECTRUM. MEAN DATA FOR 30 SUBJECTS

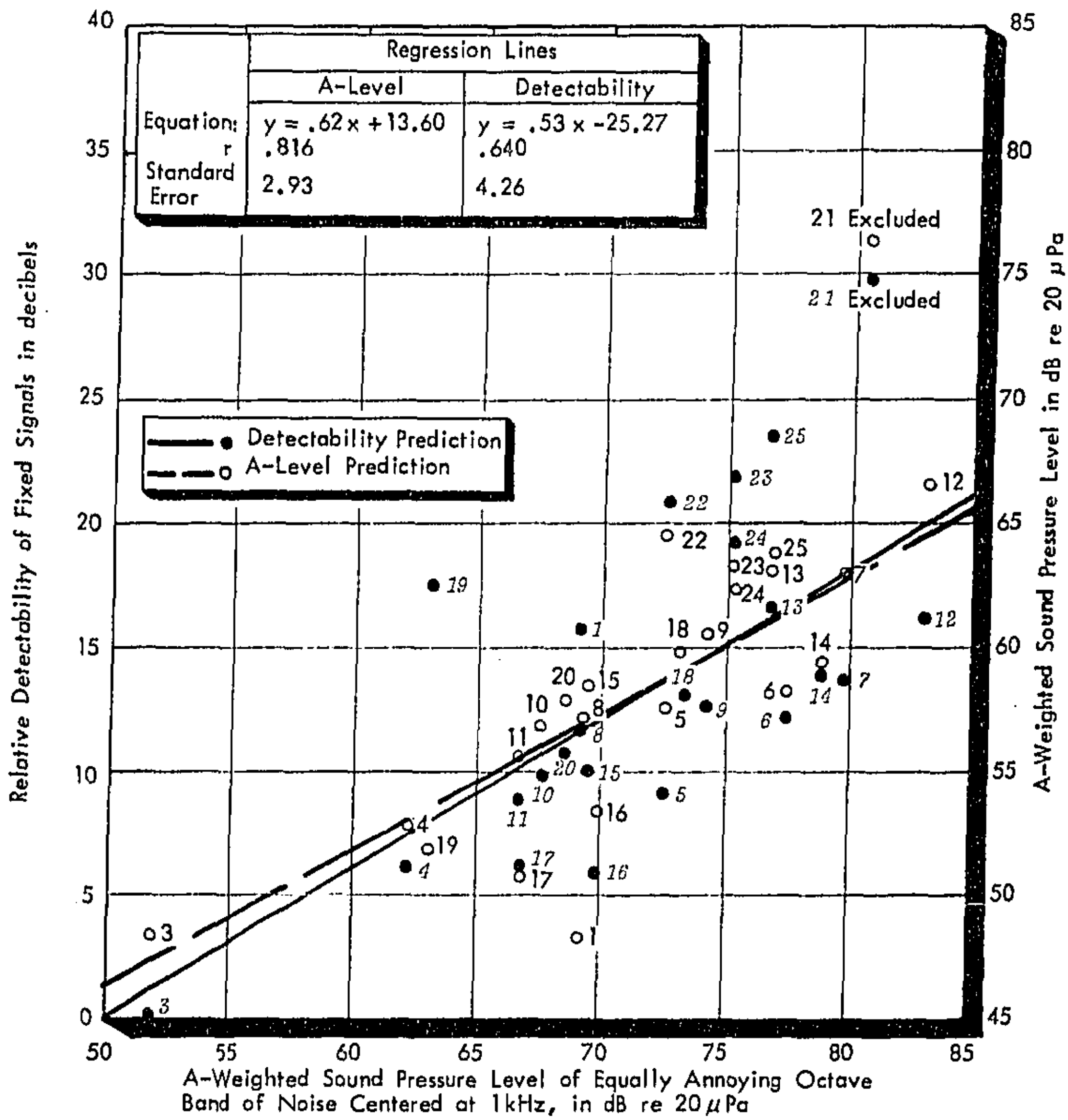


FIGURE 5. PREDICTED VS. OBSERVED ANNOYANCE OF TEST SIGNALS IN FLAT BACKGROUND NOISE SPECTRUM. MEAN DATA FOR 30 SUBJECTS.

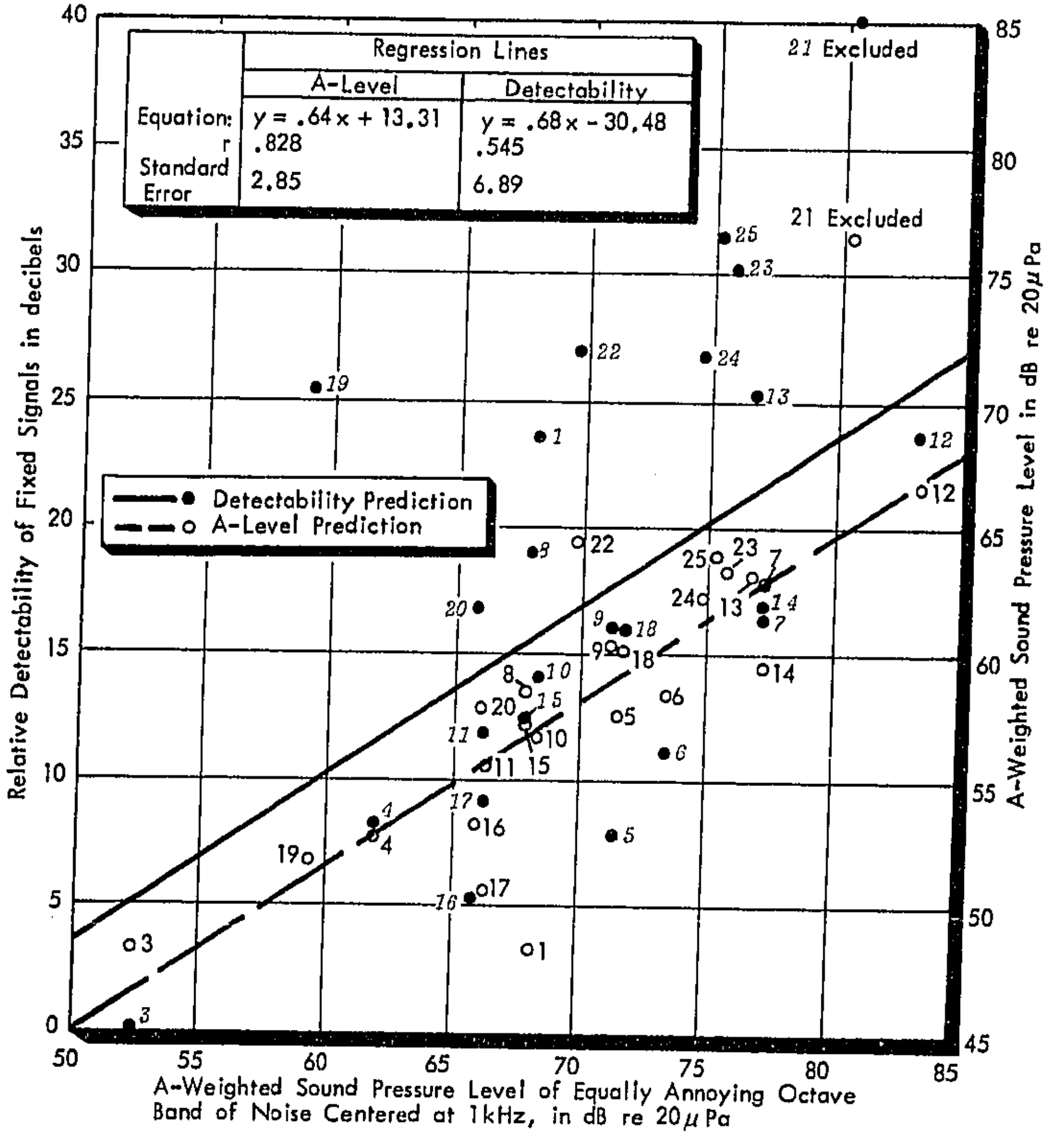


FIGURE 6. PREDICTED VS. OBSERVED ANNOYANCE OF TEST SIGNALS IN RISING BACKGROUND NOISE SPECTRUM. MEAN DATA FOR 30 SUBJECTS

The correlations reported in Table IV exclude one notable discrepant signal, #21 (a railroad train). This distinctive signal was judged much less annoying than either its A-level or signal to noise ratio (26.3 dB) would suggest. Including Signal 21 would lower the correlations of Table IV slightly under most conditions: for example, from .945 to .887 in the case of Eq. 3 predictions in the PNC-40 background, from .838 to .780 for A-level in the same background, and from .861 to .801 for D-level in the same background. The discussion section speculates on the nature of this discrepancy in the annoyance judgments.

TABLE IV. CORRELATIONS BETWEEN PREDICTED AND OBSERVED ANNOYANCE IN THREE BACKGROUNDS*

PREDICTION METHOD	BACKGROUND		
	PNC-40	FLAT	RISING
d'_{max} (Eq. 3)	.945	.640	.545
d'_{vector} (Eq. 4)	.892	.696	.550
Overall Level	.440	.642	.641
A-Level	.836	.816	.828
D-Level	.861	.822	.825
PNL	.837	.770	.779
Loudness Level (Stevens, Mark VI)	.807	.745	.752

*Signal 21 excluded

All of the correlations in Table IV (other than the overall level correlations in the PNC-40 background) differ significantly from zero at the .01 level of significance. The d'_{max} correlation of .945 is significantly higher than the A-level correlation in the NC-40 background by a one-tailed test based on the Fisher r-to-z transform. The other correlations do not differ significantly from one another.

3. Reliability and Validity of Annoyance Judgments

Two of the six PEST runs in each session were reserved for purposes of checking the meaningfulness and repeatability of annoyance judgments. Data from these runs were combined for all 30 subjects for a total of 540 test-retest runs and 540 self-test runs. The mean difference in self-test (signal compared with itself) judgments over these 540 runs was 0.88 dB, with a standard deviation of 0.62 dB. The comparable mean absolute difference in test-retest judgments was 4.17 dB, with a standard deviation of 1.45 dB.

Subjective discrimination of annoyance derived from the current procedure would therefore seem to have a resolution of about ± half a decibel, which is equivalent to the resolution of the test procedure. The repeatability of annoyance judgments by individual subjects was approximately ± 2 dB.

V. DISCUSSION

A. Major Findings

The major finding of this study is that predictions of the annoyance of a set of low level sounds based on their detectabilities correlate virtually perfectly (r = .945) with their judged annoyance in a conventional background noise environment (one dominated by low frequency energy). In terms of variance accounted for, the correlation of detectability-based predictions with judged annoyance accounts for 20% more variance than the correlation of A-level based predictions with observed annoyance.

This result is hardly surprising, in that detectability-based predictions of annoyance consider an additional parameter that A-level predictions do not consider: the relationship of a signal to the background noise in which it is heard. In a sense, this explicit consideration of the effects of audibility of sounds on their annoyance represents a return to the original philosophy from which A-level measurements were first proposed.

A-level, as explained by Galt (1930), was derived from application of inverse contours of equal loudness to sounds about 55 dB above the threshold of hearing, while B-level measurements were intended for sounds on the order of 70 dB above threshold, and C-level measurements were intended for yet higher levels. In other words, the various weighting networks were proposed to reflect the audibility of sounds half a century ago, at a time when the effects of masking by background noise on audibility of sounds were not appreciated in detail.

B. Extension of Relationship to Other Data

An independent set of annoyance judgments of 17 low level signals heard in a PNC-20 noise background was made by 32 observers for purposes unrelated to the present study (Pearsons et al., 1978). These annoyance judgments were made by the same experimental procedures described in Section III, at about the same time as the present data were collected. As a further check on the generality of the major findings of the current study, detectability-based predictions of annoyance were also made for this additional body of data.

Table V contains the sounds and their presentation levels. Figure 7 plots the relationships between detectability-based and A-level predictions of annoyance with observed annoyance for this set of data. The correlation of detectability based predictions with observed annoyance is .89, whereas the correlation of A-level predictions of annoyance with observed annoyance is only .42. For 17 signals, the A-level correlation does not even differ significantly from 0 (no correlation at all), while the correlation of detectability-based predictions and observed annoyance is significantly different from 0 at the .01 level of significance.

The failure of the A-level predictions to account for the observed annoyance is attributable primarily to some glaring mispredictions of the annoyance of signals with concentrations of energy at low and high frequencies. Because of the smaller effects of masking at high frequencies relative to low frequencies

TABLE V. LIST OF SIGNALS AND PRESENTATION LEVELS IN INDEPENDENT STUDY

SIGNAL #	DESCRIPTION	A-WEIGHTED PRESENTATION LEVEL (dB)
1	Simulated Power Transformer (1)	variable
2	Octave Band of Noise Centered at 2 kHz	variable
3	Transformer (1)	50.0
4	Electric Power Line (1)	45.2
5	Transformer (2)	50.0
6	Electric Power Line (2)	44.6
7	Simulated Power Line	45.0
8	Transformer Shaped Noise	49.8
9	Simulated Power Transformer (2)	49.8
10	Low Pass Filtered Power Line	37.9
11	High Pass Filtered Power Line	41.9
12	Rain	40.1
13	Traffic	45.2
14	Babble	55.3
15	Heavy Tractor	54.4
16	Lawnmower	49.9
17	Dishwasher	50.1
18	Air Conditioner	44.7
19	Noy Curve	44.7

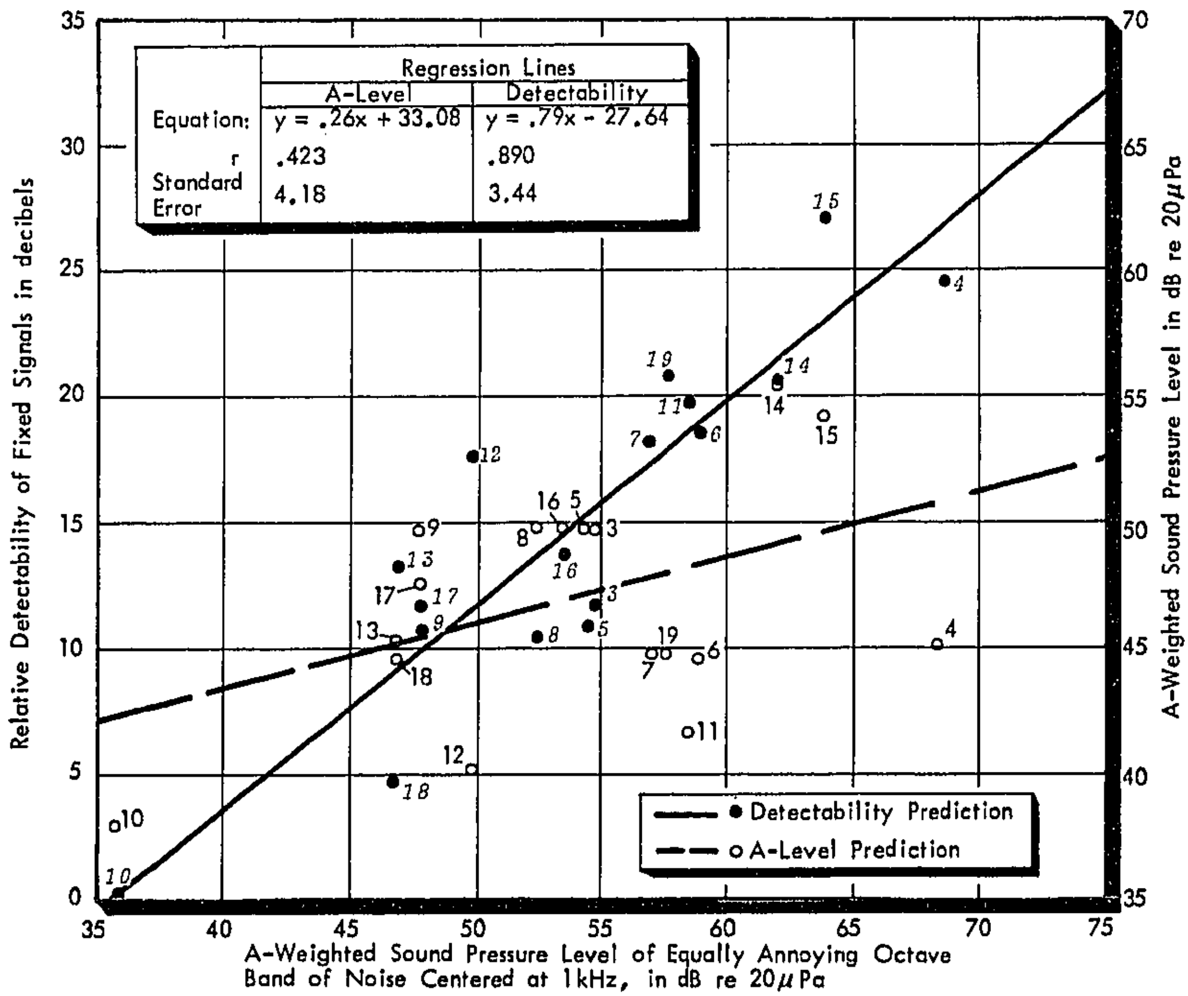


FIGURE 7. PREDICTED VS. OBSERVED ANNOYANCE OF TEST SIGNALS IN PNC-20 BACKGROUND NOISE SPECTRUM. MEAN DATA FOR 32 SUBJECTS

in the NC-20 background, the detectability-based predictions were capable of reflecting the observed annoyance judgments more accurately.

C. Limits of Application of Detectability Predictions

Equation 2, upon which the current predictions of detectability are based, is a general expression for predicting the detectability of long duration, relatively steady state broadband signals in broadband noise. Detectabilities of other unusual signals (for example, signals of very brief duration, or of very narrow band frequency composition, or of extremely low frequency content) may not be well predicted by this equation, for a variety of technical reasons (Green and Swets, 1966). Nonetheless, the model accounted reasonably well for the annoyance of impulsive type typewriter noise (Signal 15) and two other non-steady state spectra (Signals 16 and 17). It did, however, predict that highly detectable low frequency tones (such as the pure tone at 120 Hz of Signal 1) would be much more annoying than subjects actually found them to be.

D. Alternative Predictions of Annoyance Based on Signal to Noise Ratio

As is evident in Table IV, the correlations of detectability-based predictions of annoyance with the observed annoyance values are inversely related to the slope of the masking noise spectrum: the relationship is strongest in the familiar falling spectrum typical of community noise in urban areas, but decreases in the flat spectrum, and decreases yet further in the rising spectrum.

This decrease in strength of relationship appears to be due to the relative detectability of the low and high frequency portions of a signal's spectrum. All other things being equal, the commonplace falling background masks low frequencies most heavily, the flat background masks all frequencies about equally, and the rising background masks high frequencies most heavily. Thus, the same signal is likely to be detectable at high frequencies in the falling background, but at low frequencies in the rising spectrum.

It is well known that low frequencies are not as annoying as high frequencies: this observation is the basis for many frequency weighting procedures. By basing predictions of annoyance on detectability alone, without regard for the frequency band within which detection occurs, a systematic error is induced in detection-based prediction equations. In other words, sounds which are highly detectable in a rising background (by virtue of their unmasked low frequency content), are simply not as annoying as they are when equally detectable in a falling background (by virtue of their unmasked high frequency content).

To demonstrate this effect, another set of detectability predictions was generated by applying Eq. 2 only to signal to noise ratios in one-third octave bands at or above 200 Hz. The improvement in correlation obtained by doing so is clear: in the flat background, the correlation between predicted and observed annoyance increased from .640 to .763, while in the rising background, the correlation increased from .545 to .679.

E. A Notable Anomaly

The greatest discrepancy encountered in predicting annoyance from detectability was the surprisingly small annoyance of Signal 21, a railroad train. Although extremely detectable by virtue of its very high signal to noise ratio (it was heard at a level 26 dB(A) above the level of the background noise), few subjects found the train aversive. Many subjects commented that they enjoyed listening to the train because it was recognizable as a train, because it was "interesting" (contained wheel clicks and other rhythmic and identifiable sounds), because it was "unusual" (different from the other signals), or just because they liked trains. The variance of subjects' judgments of the annoyance of the train was among the highest of any signal in all background noise conditions, despite its high absolute level.

Since A-level predictions of the annoyance of Signal 21 were also considerably in error, there is some reason to believe that the peculiarities of the signal were responsible for its overestimated annoyance. It remains unclear, however, whether the failure of detectability-based prediction of annoyance to account for Signal 21's anomalously low annoyance could also be attributed to its high signal to noise ratio. In other words, the background noise environment may not influence the annoyance of signals that exceed the ambient level by 25 dB or more.

F. Effects of Different Rules for Combination of Detectabilities in Different Spectral Regions

Section II discusses two ways of predicting the detectability of complex signals (Eqs. 3 and 4). The simpler method (Eq. 3), which bases the detectability prediction exclusively upon the most noticeable portion of a signal, seems preferable for present purposes, since it may provide a slightly higher correlation with annoyance judgments than the more complex prediction method (Eq. 4) in conventional noise backgrounds. Furthermore, since there is so little difference in the magnitude of the correlation between observed and predicted annoyance associated with the two prediction rules, parsimony would suggest selection of the simpler rule.

Note that it remains unclear whether Eq. 3 or Eq. 4 better predicts the actual detectability of complex signals, since no effort was made in the current study to determine empirically how detectable the test signals actually were.

G. On Construction of Scales of Intrusiveness

By Stevens' definitions (Stevens, 1951), scales may be categorized as nominal, ordinal, interval, or ratio. Nominal and ordinal scales are of little interest for present purposes, since they provide no convenient quantitative guidance required for noise assessment purposes. Interval scales differ from ratio scales primarily by their arbitrary zero point and units. Technically, predicted detectability can serve as a ratio scale for measurement of intrusiveness, since it has a true zero point (complete undetectability). Pragmatically, however, an arbitrary definition is needed to yield a useful range of values for assessment. For example, a definition of the form "a noise source may be said to be intrusive if 25% of the population would be highly annoyed if exposed to it", although arbitrary, would support a readily interpretable scale.

The current research, which sought primarily to establish whether or not detectability of low level signals was related to annoyance, does not provide any direct information about the levels of detectability that would be most convenient for a definition of intrusiveness. This research does, however, indicate that scales of annoyance may be constructed from physical information about signal and background levels. Such scales would differ only by a constant factor from any detectability-based scale of intrusiveness that might eventually be adopted for assessment purposes.

Selection of the origin for a scale of intrusiveness is not readily justified on the basis of the existing literature on annoyance. In the first place, most of the literature on annoyance from transportation related noise deals with the annoyance of very high level signals, such as aircraft flyovers. It seems safe to assume that such high level signals are intrusive; the problem is to find a lower limit, not an upper limit of intrusiveness. One might therefore turn to adjective scales and categorical judgment data such as that of Pearsons and Horonjeff (1967). A summary figure from Pearsons and Horonjeff (1967) is reproduced here as Figure 8.

It might be argued from Figure 8 that a sound described as "very quiet", "soft", "of no concern", and "pleasant" (with an associated Perceived Noise Level on the order of 60 dB) could not be called intrusive, but that a sound described as "noisy", "loud", "barely acceptable", or "annoying" (with an associated Perceived Noise Level on the order of 90 dB) would have to be considered intrusive. It is doubtful whether this 30 dB range of uncertainty could be meaningfully reduced through inspection of manipulation of other category scale information.

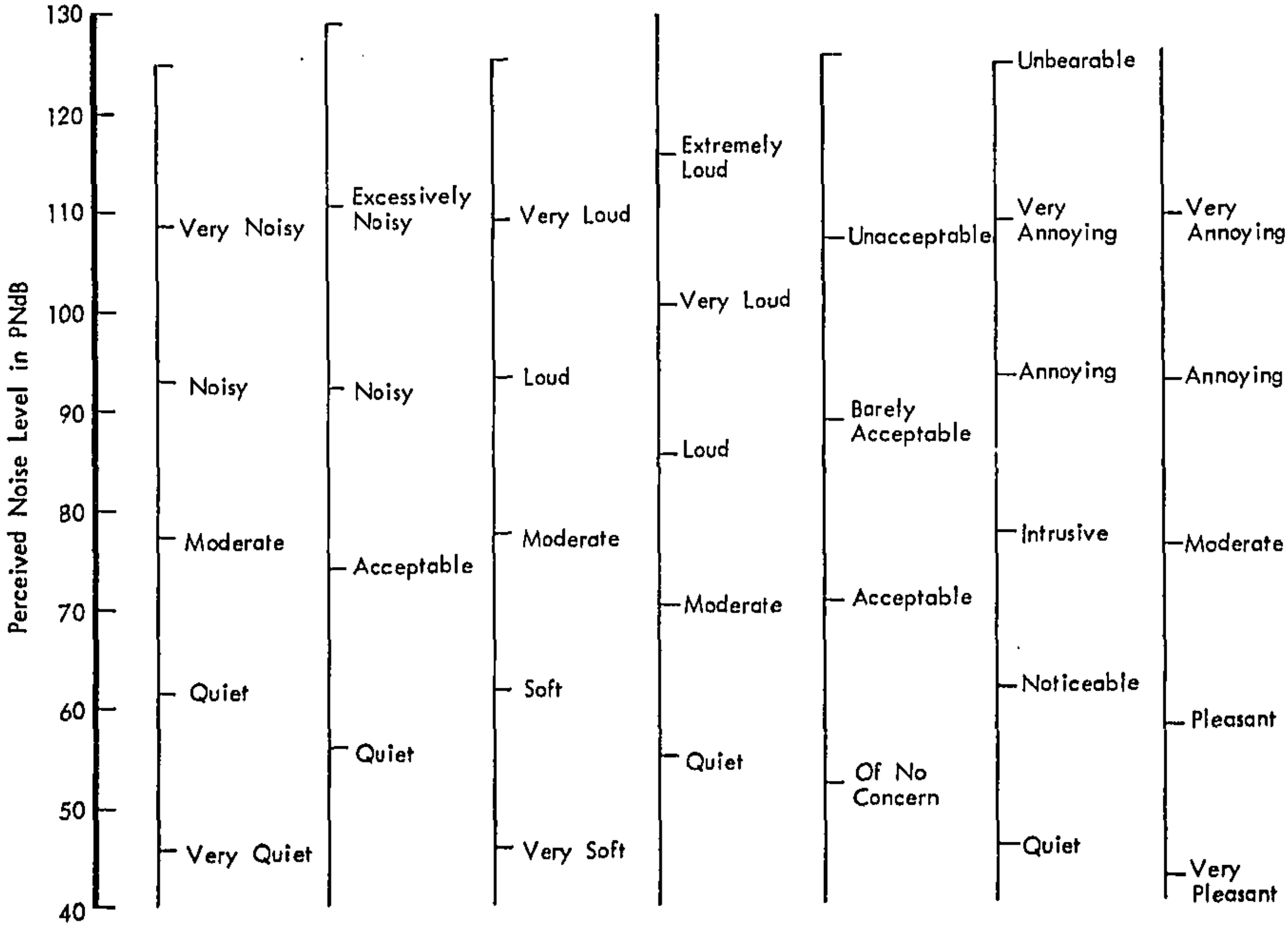


FIGURE 8. JUDGED EQUIVALENT SCALE RATINGS FOR STIMULI EMPLOYED IN ALL LABORATORY TEST SESSIONS

One possible way to produce evidence about the physical characteristics of signals that actually do intrude upon people's awareness sufficiently to be annoying would be to observe in real time the behavior of people in residential settings. Such field measurements could be made by techniques such as those of Fidell et al. (1972), and Purcell (1977).

Alternatively, it may simply be asserted that a signal characterized by a d' value of 40.0 is intrusive. This level of detectability was determined by Fidell (1978) to be a level that attracted the attention of people engaged in a simulated automobile driving task. It might be argued that a signal noticeable enough to cause a driver to divert attention from a complex psychomotor task such as driving is by definition intrusive.

Table VI and Figure 9 illustrate some of the concepts presented in this section as indications of the ways in which the relationships discovered in this research might be applied for assessment purposes.

Table VI. HYPOTHETICAL SCALE OF INTRUSIVENESS IN NC-40 BACKGROUND

Signal	A-Level	Observed[1] Relative Annoyance	Absolute Detectability[2]	Intru-siveness[3]
Transformer	48.4	3.4	28.3	- 1.5
Blender	48.4	0	36.6	- 0.4
Egg Beater	52.9	9.8	152.8	5.8
Carving Knife	57.6	19.2	333.5	9.2
Jig Saw	58.4	22.1	746.7	12.7
Hair Dryer	62.9	23.7	1079.3	14.3
Lathe	57.3	11.8	187.3	6.7
Router	60.7	19.8	780.9	12.9
Belt Sander	56.8	13.8	260.4	8.1
Hand Drill	55.7	15.0	226.8	7.5
Radial Arm Saw	66.6	26.4	2161.1	17.3
Air Compressor	63.2	20.7	1111.0	14.4
Model Airplane	59.5	24.3	734.0	12.6
Typewriter	58.5	12.2	374.2	9.7
Toy Car	53.4	11.2	139.7	5.4
Toy Dog	50.7	10.3	124.6	4.9
Vacuum Cleaner	60.0	17.2	610.5	11.8
Air Condi-tioner	51.9	1.3	43.9	0.4
Garbage Com-pactor	57.9	9.4	272.7	8.3
Train	76.3	22.1	10818.4	24.3
Motorcycle	64.5	14.6	382.9	9.8
Automobile	63.3	18.5	990.2	13.9
Bus	62.4	19.1	634.0	12.0
Lawnmower	63.9	20.0	528.2	11.2

[1]Empirical data, in dB re adjusted level of 1 kHz octave band of noise at point of equal annoyance for Signal 3.

[2]d'_{max} in any 1/3 octave band from 50 Hz to 10 kHz.

[3]d'_{max} re $d' = 40$, in dB.

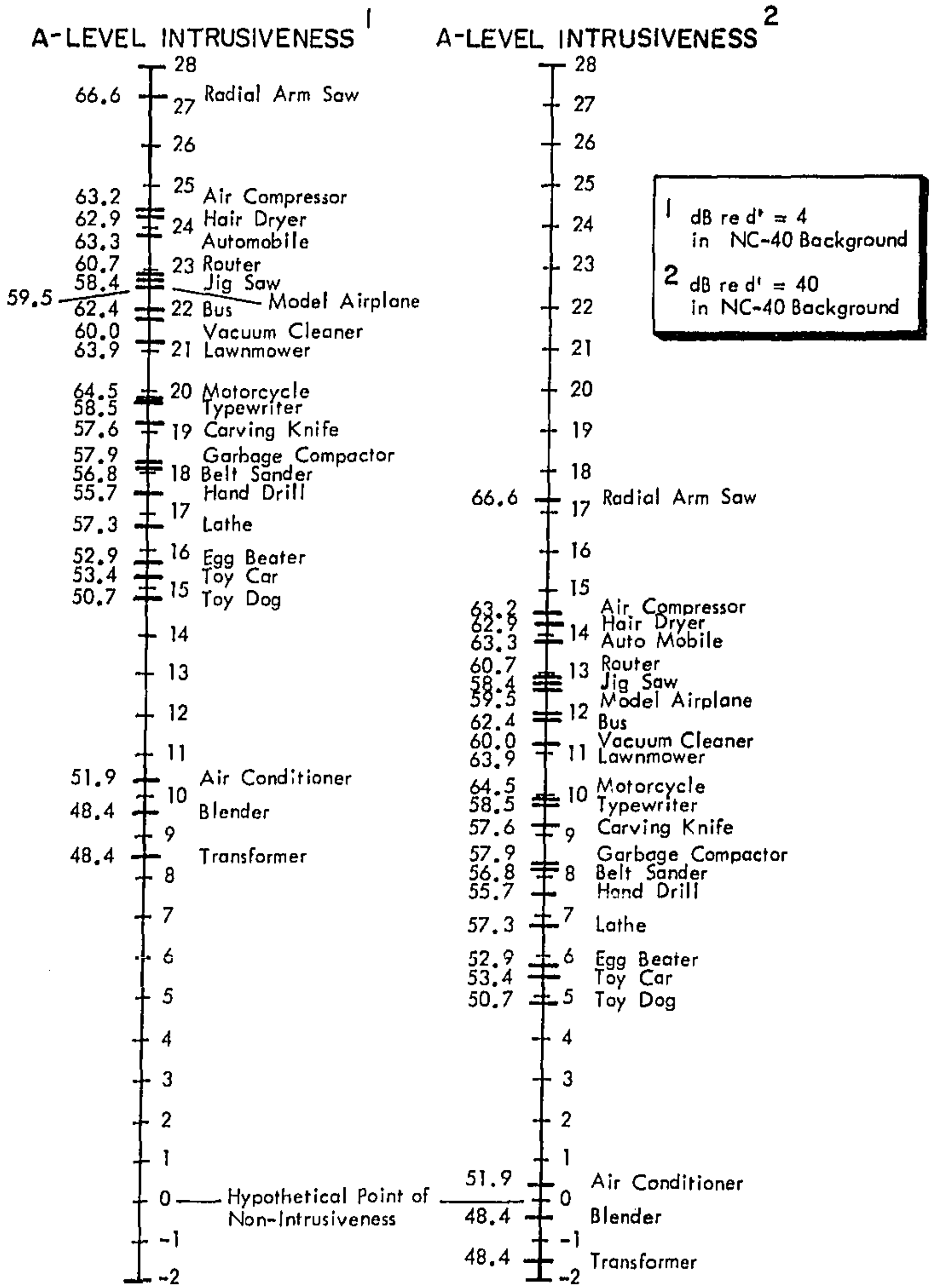


FIGURE 9. ALTERNATIVE SCALES RELATING INTRUSIVENESS TO A-LEVEL OF TEST SIGNALS

VI. CONCLUSIONS

1) The predicted detectability of everyday low level sounds (on the order of 50-70 dB(A)) heard in a noise environment with a commonplace falling spectrum accounts for almost 90% of the variance in subjective judgments of the annoyance of such sounds. The correlation between annoyance and predicted detectability is higher than the correlations of annoyance with conventional frequency-weighted measures of signal level alone under these conditions. The remaining unexplained variance is as likely to be due to random factors, such as the limits of resolution of data collection or the consistency of human annoyance judgments, as to any systematic factors. Thus, the predicted correlation of detectability with annoyance in everyday background noise environments is not likely to be surpassed by any other theoretically based physical measure of acoustic signals.

2) One implication of this finding is that the background noise in which sounds are heard has a considerable influence on their annoyance.

3) The relationship between predicted detectability and annoyance of low level signals noted in 1) above is strong enough to support arbitrary scales of intrusiveness, which differ from a ratio scale of absolute detectability only by constant factors.

4) Variability in subjective judgments of the annoyance of low level signals is negatively correlated with the absolute levels of such signals.

5) Varying the absolute levels of a set of low level signals over a 20 dB range does not affect their relative annoyance if their signal-to-noise ratios are preserved.

REFERENCES

Fidell, S., "Community Reaction to Noise Exposure", Chapter 36, in C. Harris, "Handbook of Noise Control", in press.

Fidell, S., K. Pearsons and R. Bennett, "Prediction of Aural Detectability of Noise Signals", Human Factors 16(4), 373-383, 1974.

Fidell, S., G. J. Jones and K. S. Pearsons, "Feasibility of a Novel Technique for Assessing Noise-Induced Annoyance", Department of Transportation Report DOT-TST-74-3, NTIS PB-225334, September 1973.

Fidell, S., "Effectiveness of Audible Warning Signals for Emergency Vehicles", Human Factors 20(1), 19-26, 1978.

Galt, R. H., "Results of Noise Surveys: Part I: Noise Out-of-Doors", J. Acoust. Soc. Am., 2, 30-58, 1930.

Green, D. M., and J. A. Swets, "Signal Detection Theory and Psychophysics", John Wiley and Sons, Inc., 1966.

Pearsons, K. S. and R. D. Horonjeff, "Category Scaling Judgment Tests on Motor Vehicle and Aircraft Noise", Department of Transportation, Federal Aviation Administration, FAA DS-67-8, July 1967.

Pearsons, Karl S., Ricarda L. Bennett, and Sanford A. Fidell, "The Effects of Transformer and Transmission Line Noise on People - Annoyance", Vol. 1, Bolt Beranek and Newman Report No. 3710, 1978.

Purcell, A. T., and R. H. Thorne, "An Alternative Method for Assessing the Psychological Effects of Noise in the Field", J. Sound and Vibration 55(4), 533-544, 1977.

Stevens, S. S., "Mathematics, Measurement, and Psychophysics", in Handbook of Experimental Psychology, John Wiley and Sons, New York, 1951.

Stevens, S. S., "Procedure for Calculating Loudness: Mark VI", J. Acoust. Soc. Am. 33: 1577-1585, 1961.

Taylor, M. M., and C. D. Creelman, "PEST: Efficient Estimates on Probability Functions", J. Acoust. Soc. Am., 41, 782-787, 1967.

APPENDIX A

APPENDIX A

PEST PROCEDURE EMPLOYED FOR SUBJECTIVE JUDGMENT TESTS

Parameter Estimation by Sequential Testing (PEST) is a computer based adaptive psychophysical procedure which administers an iterative form of the standard paired comparison task to human observers. PEST is called an adaptive procedure because the sequence of signals heard by an observer is not fixed in advance, but rather is determined by his ongoing responses. PEST thus preserves many of the advantages of the paired comparison method while gaining the speed and convenience of an adjustment method.

BBN's implementation of the PEST is based on an interactive teletype conversation between the experimenter and the computer-based system. The system acquires information needed for conduct of an experiment by inquiring of the experimenter the values of a series of parameters which determine the course of the PEST procedure. Initially, the computer requests identification of the observer, the signals employed, and the experimental session. The next questions posed by the computer concern the relative levels at which signals are presented to the observer on subsequent trials.

The experimenter may then specify a standard operating procedure consisting of predetermined values of a dozen parameters such as the intersignal interval, intertrial interval, initial step size, maximum step size, degree of confidence in the observer's responses, anticipated direction of first step, and region of interest of the psychometric function.

A final question serves to delay onset of a trial series until the experimenter and observer are ready to procede. Upon receiving an affirmative response to the question "READY?", the computer types a data heading and awaits final confirmation in the form of "START" switch depression by an observer in an adjacent anechoic chamber.

The trial procedure is a two interval forced choice, in which one signal (the standard) is invariant over trials, while the other signal (the comparison) may change in level. Approximately one second after START switch closure, the computer presents a pair of signals and waits for the observer to decide on his preference for the signal of the first or second interval. Upon receipt of the observer's response, the computer calculates the level at which the comparison signal will be presented on the next trial. After another pause of approximately one second, the computer initiates the next trial by presenting a modified signal pair.

PEST determines the increment in comparison signal level as follows (Taylor and Creelman, 1967).

1. On every reversal of step direction, halve the step size.

2. The second step in a given direction, if called for, is the same size as the first.

3. Whether a third successive step in a given direction is the same as or double the second depends on the sequence of steps leading to the most recent reversal. If the step immediately preceding that reversal resulted from a doubling, then the third step is not doubled; while if the step leading to the most recent reversal was not the result of a doubling, then this third step is double the second.

4. The fourth and subsequent steps in a given direction are each double their predecessor (except that large steps may be disturbing to a human observer and an upper limit on permissible step size of 16 dB is maintained).

The system provides information about the progress of each run in the form of "UP" and "DOWN" lights (signifying the direction of change of comparison signal level on the current trial), and also in two digital counters which cumulate numbers of trials and of decision reversals.

A run, composed of a variable number of trials, is terminated when the system determines that sufficient information has been collected. The general stopping criterion for a run is satisfied when the anticipated step size is 1 dB. When a run terminates, the computer prints the number of the run, the level of the comparison signal on the last trial of the run, the number of trials in the run and the mean response latency. The program is usually set to determine the point of subjective equality, or the level at which observers judged the standard and comparison signals equally noisy.

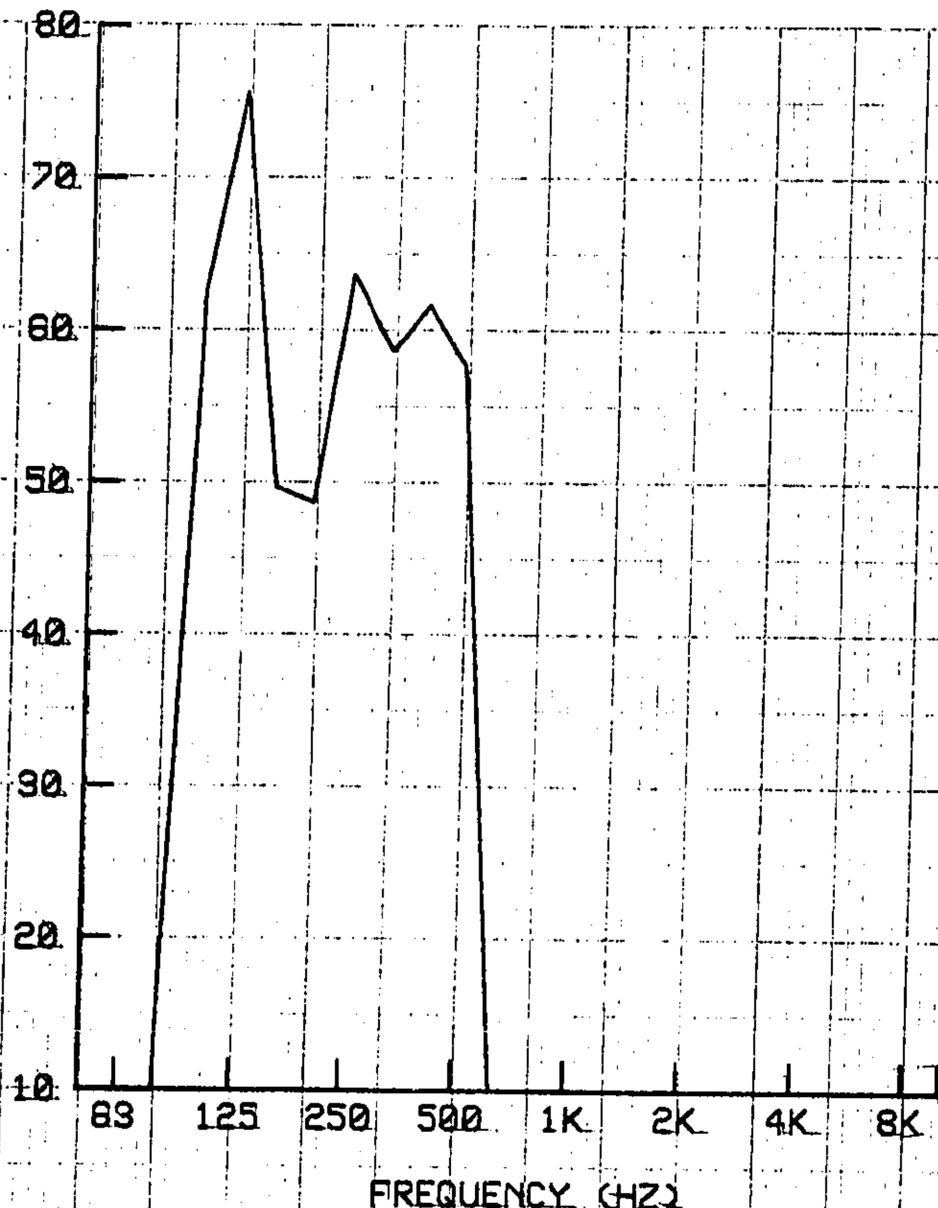
TEST INSTRUCTIONS

During this experiment, you will hear a series of pairs of sounds. Your job will always be the same: to decide which sound of a pair is more annoying. The sounds you hear may vary in level from time to time, based on your opinions about which are more annoying. Sometimes they may be quite loud, and other times they may even be hard to hear.

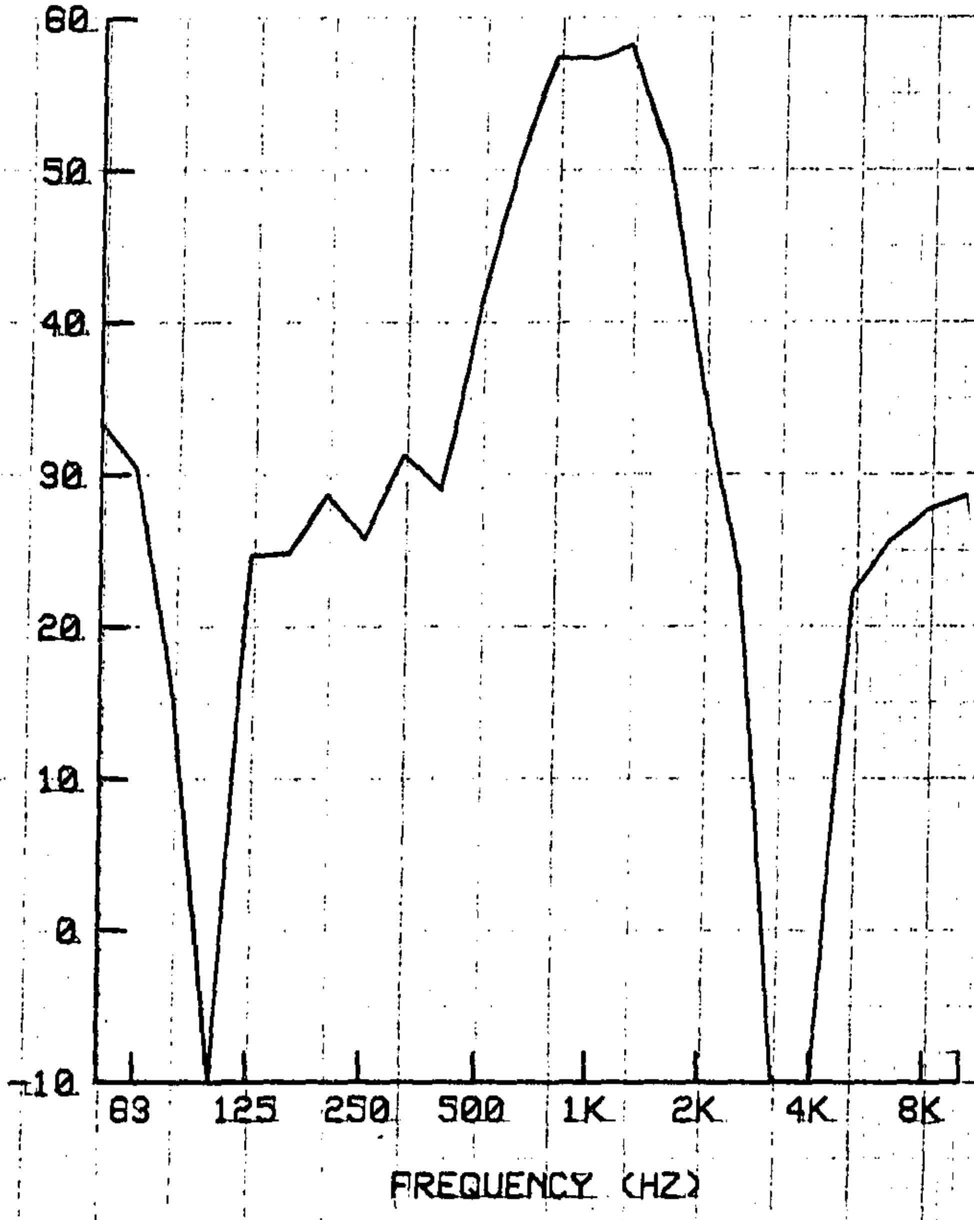
For the computer to keep track of your decisions about which sounds are more annoying, you will have to follow a fixed procedure. A trial starts when the button marked "1" on your response box lights up. As long as Button 1 is lighted, you will be hearing the first sound of a pair. A short while after the light in Button 1 goes out, Button 2 lights up. As long as Button 2 is lighted, you will be hearing the second sound of a pair. When the light in Button 2 goes out, you must press either Button 1 or Button 2 to indicate which sound you felt was more annoying.

The pairs of sounds you will hear will not be presented in any systematic pattern, but will be randomized by the computer. Since there are no "right" or "wrong" answers and since there is no pattern to the order in which you will hear pairs of sounds, there are no fixed strategies to help you make up your mind about which sound of a pair is more annoying. All we *ever* want to know is which sound of a pair is more annoying to you at the moment you make your decision. Please pay careful attention to the sounds at all times. We realize that it may be difficult to make decisions in some cases, but it is very important that you try hard.

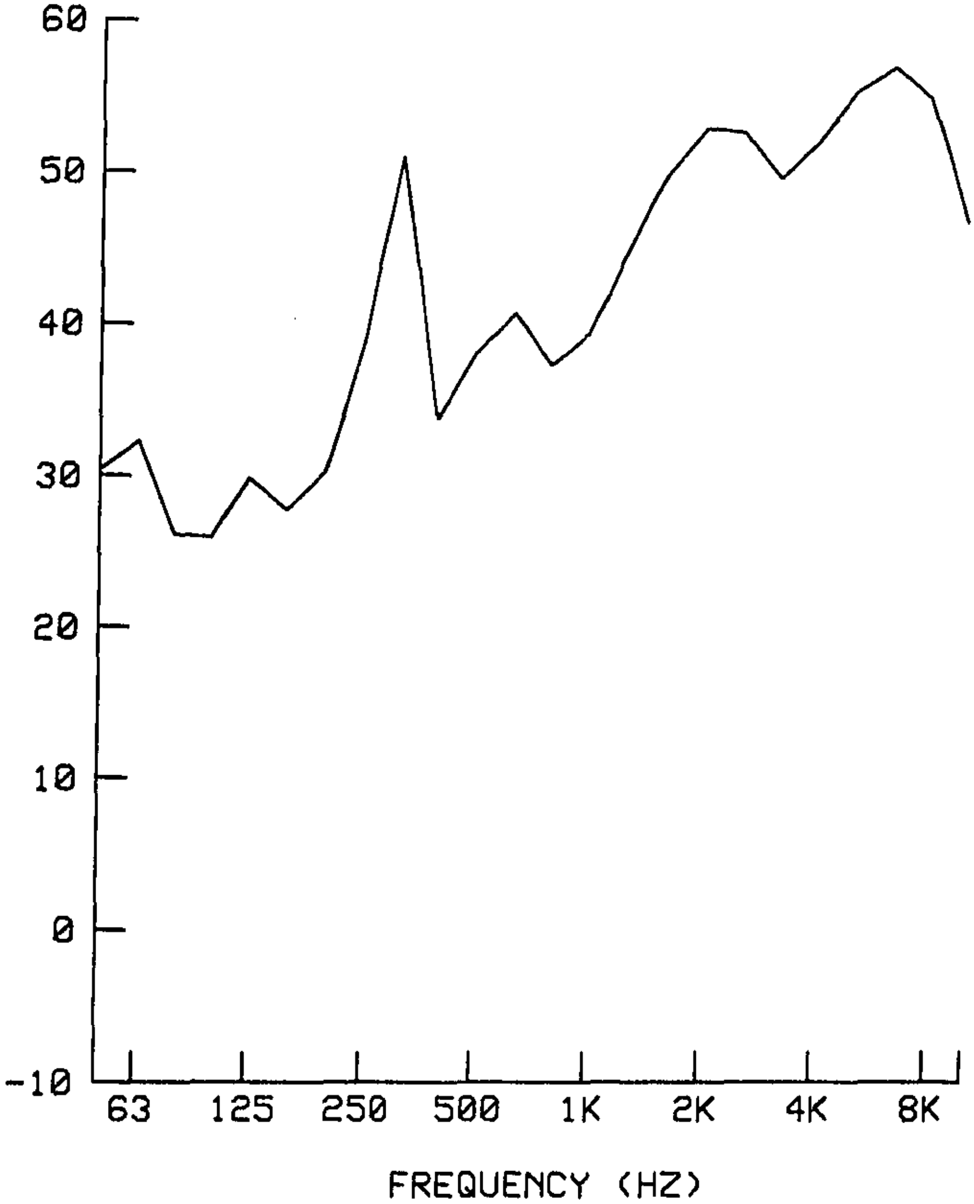
APPENDIX B



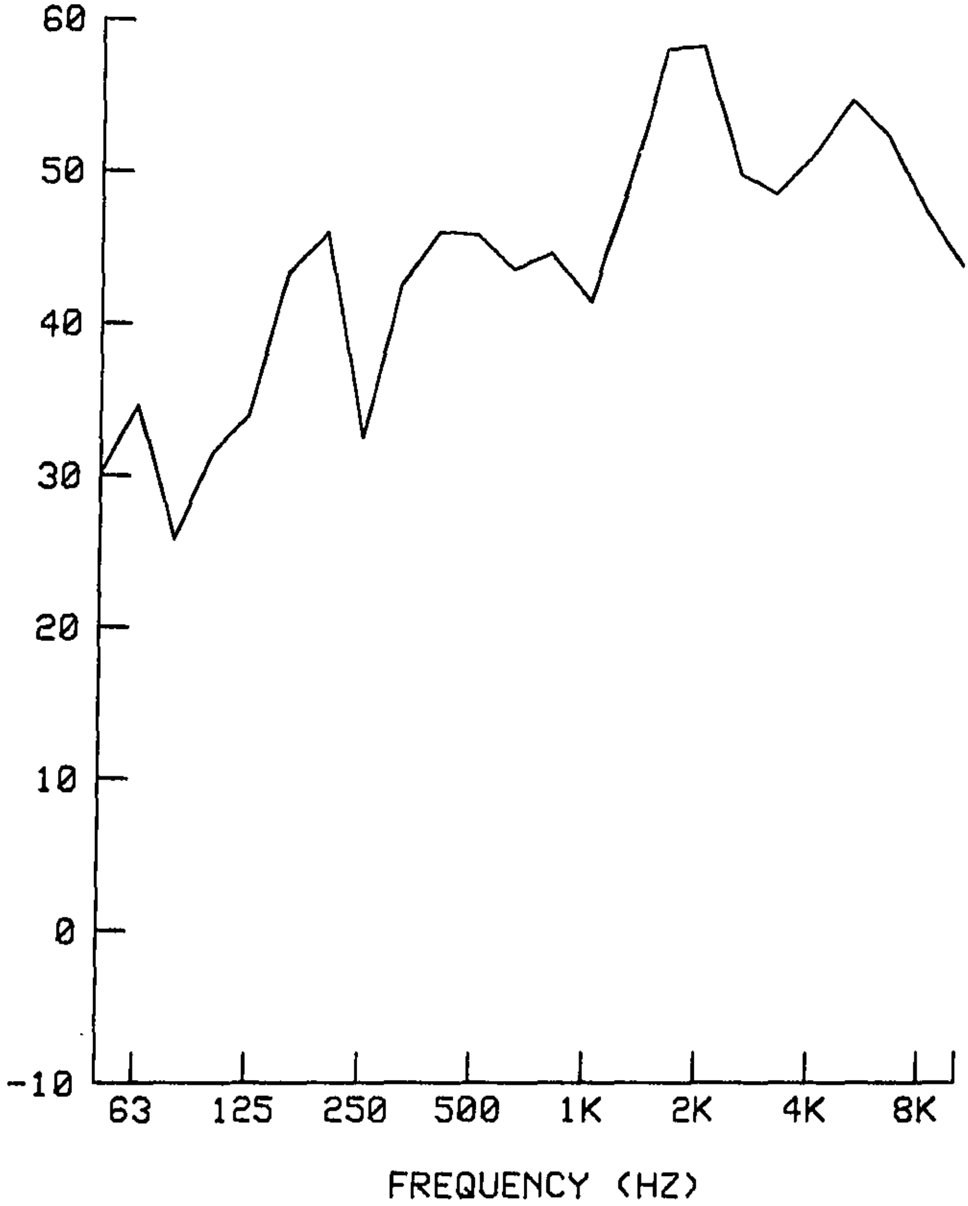
SYNTHETIC TRANSFORMER



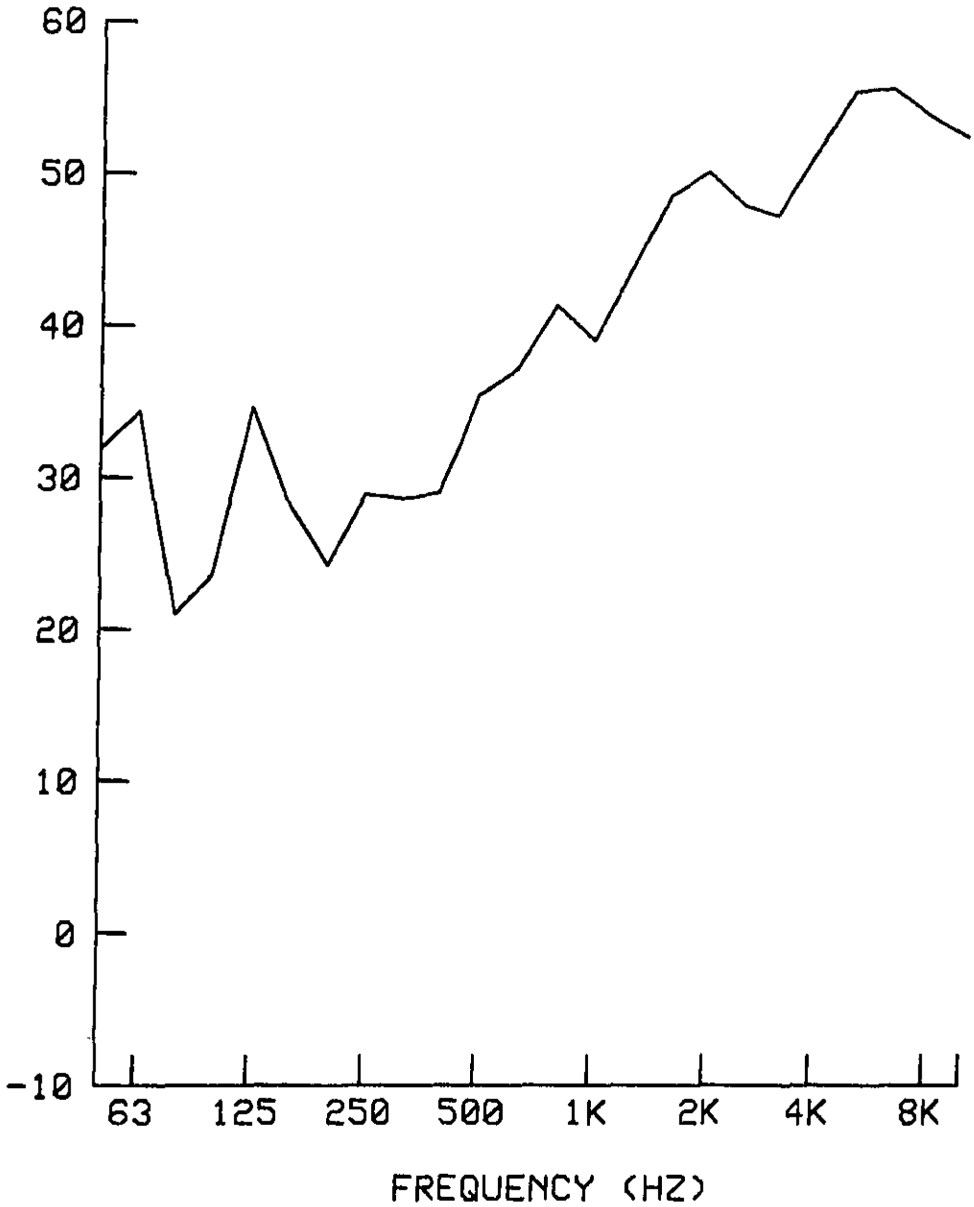
ERA SIGNAL #2 - 1000 HZ OCTAVE BAND OF NOISE



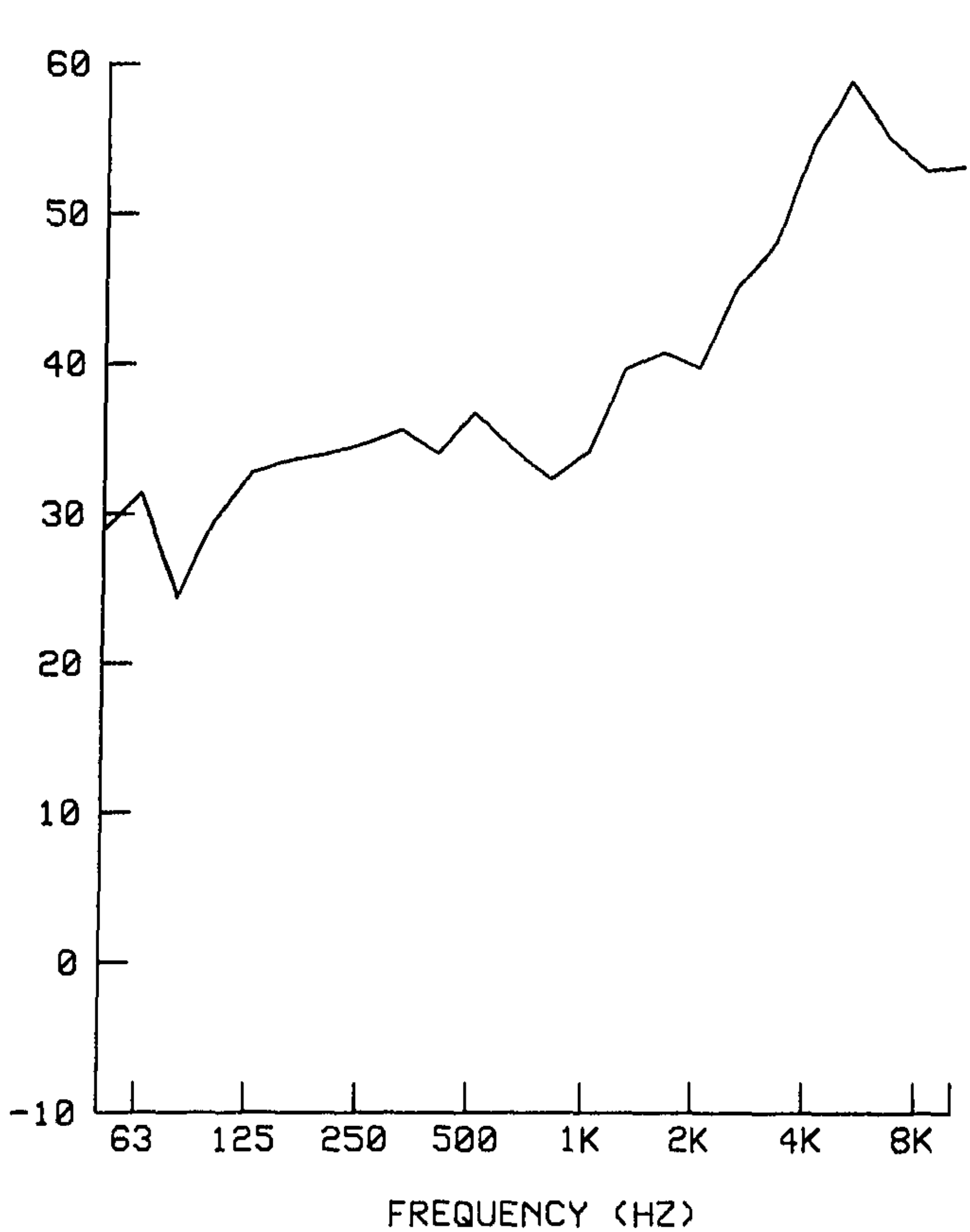
EPA SIGNAL #3, BLENDER WITH WATER



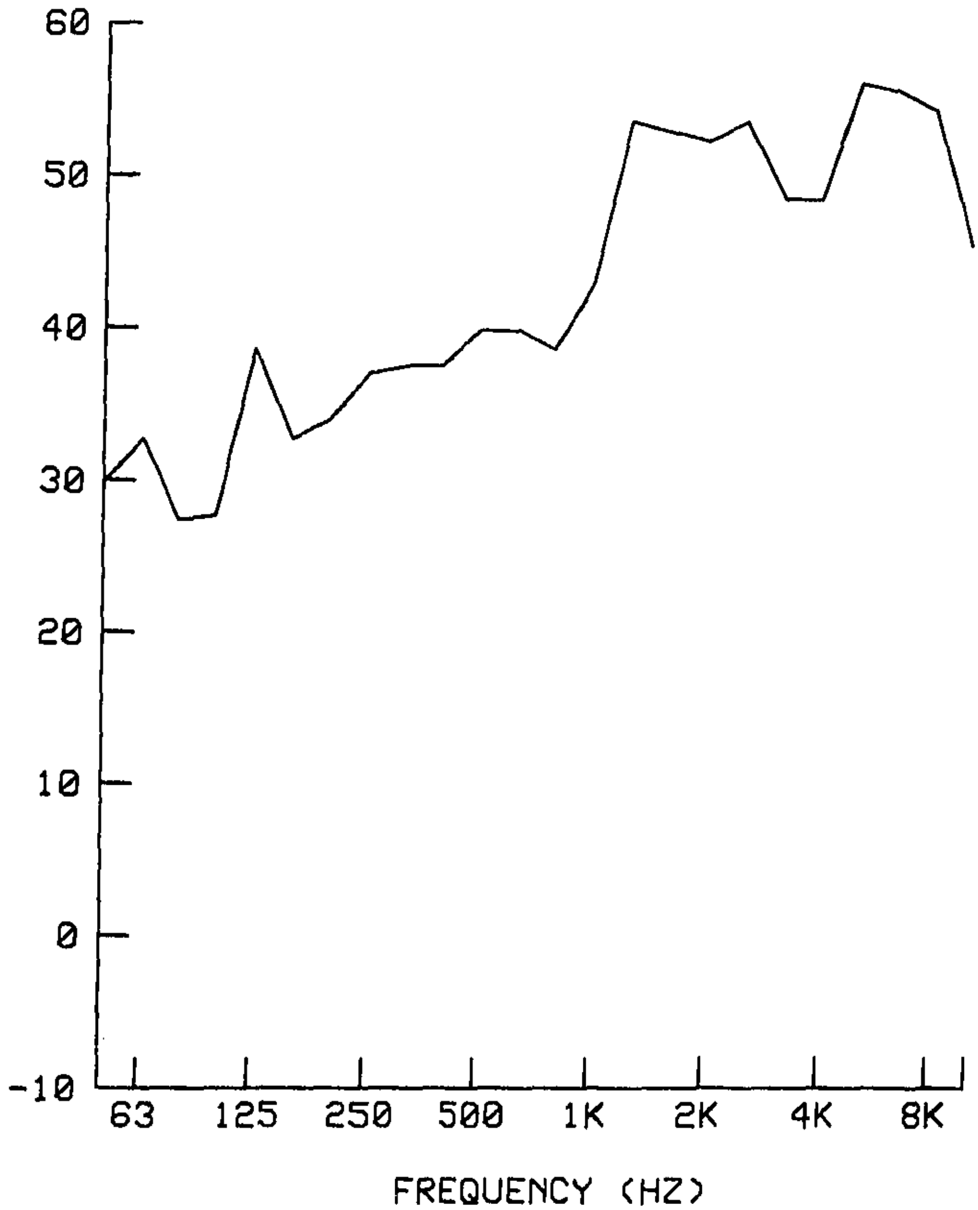
EPA SIGNAL #4. EGG BEATER. 9/9/77



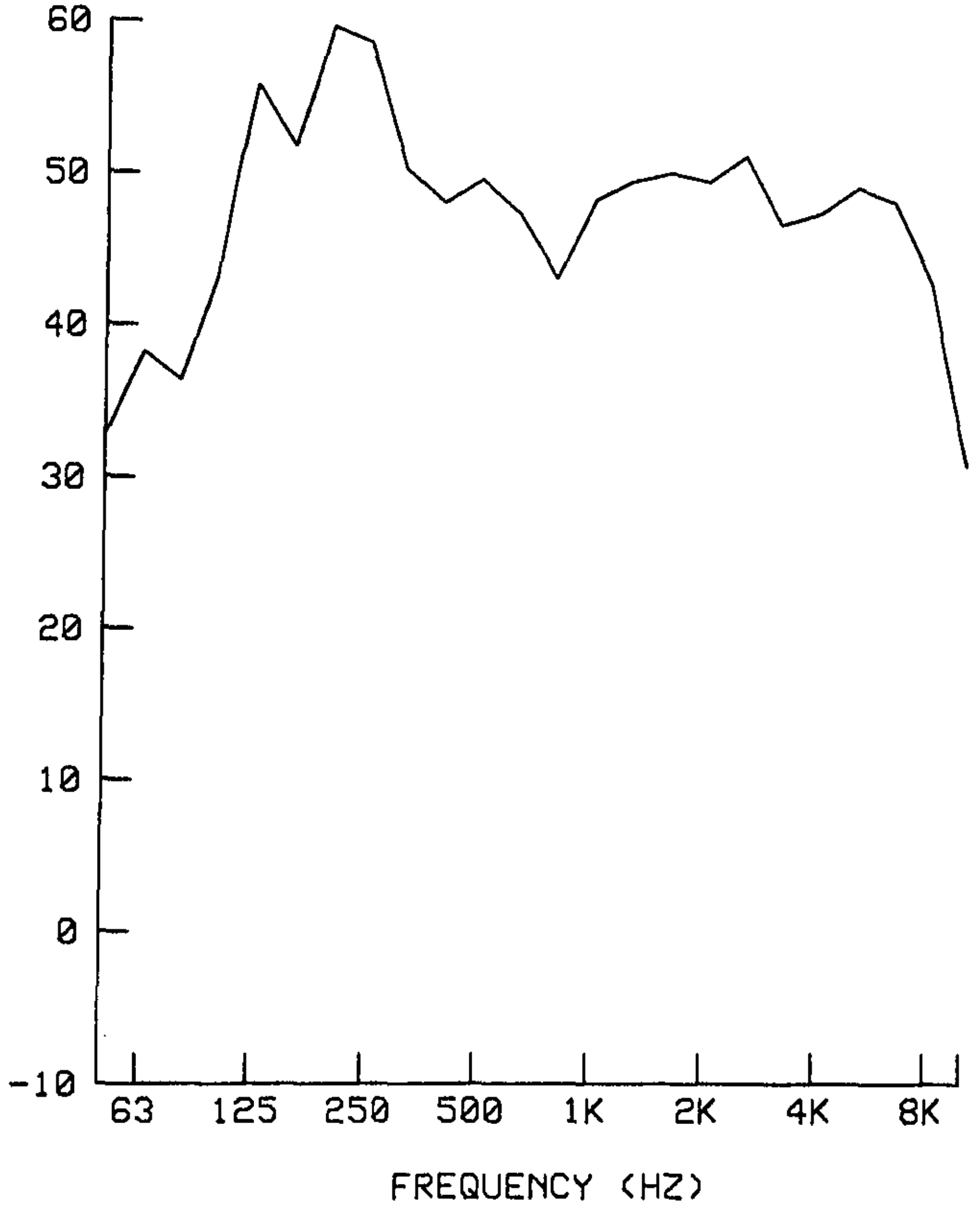
EPA SIGNAL #5, CARVING KNIFE, 9/9/77



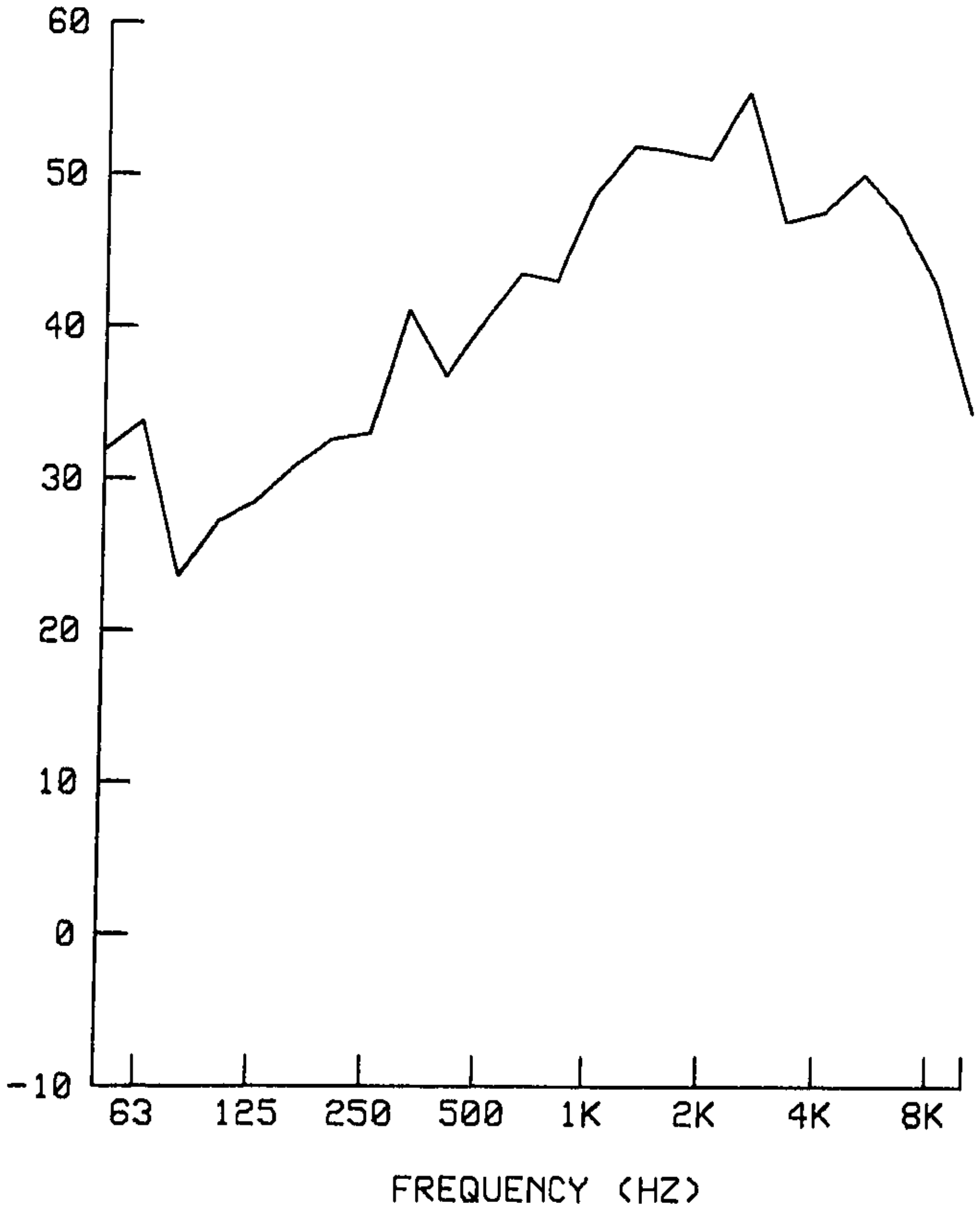
EPA SIGNAL #6, JIG SAW, 9/9/77



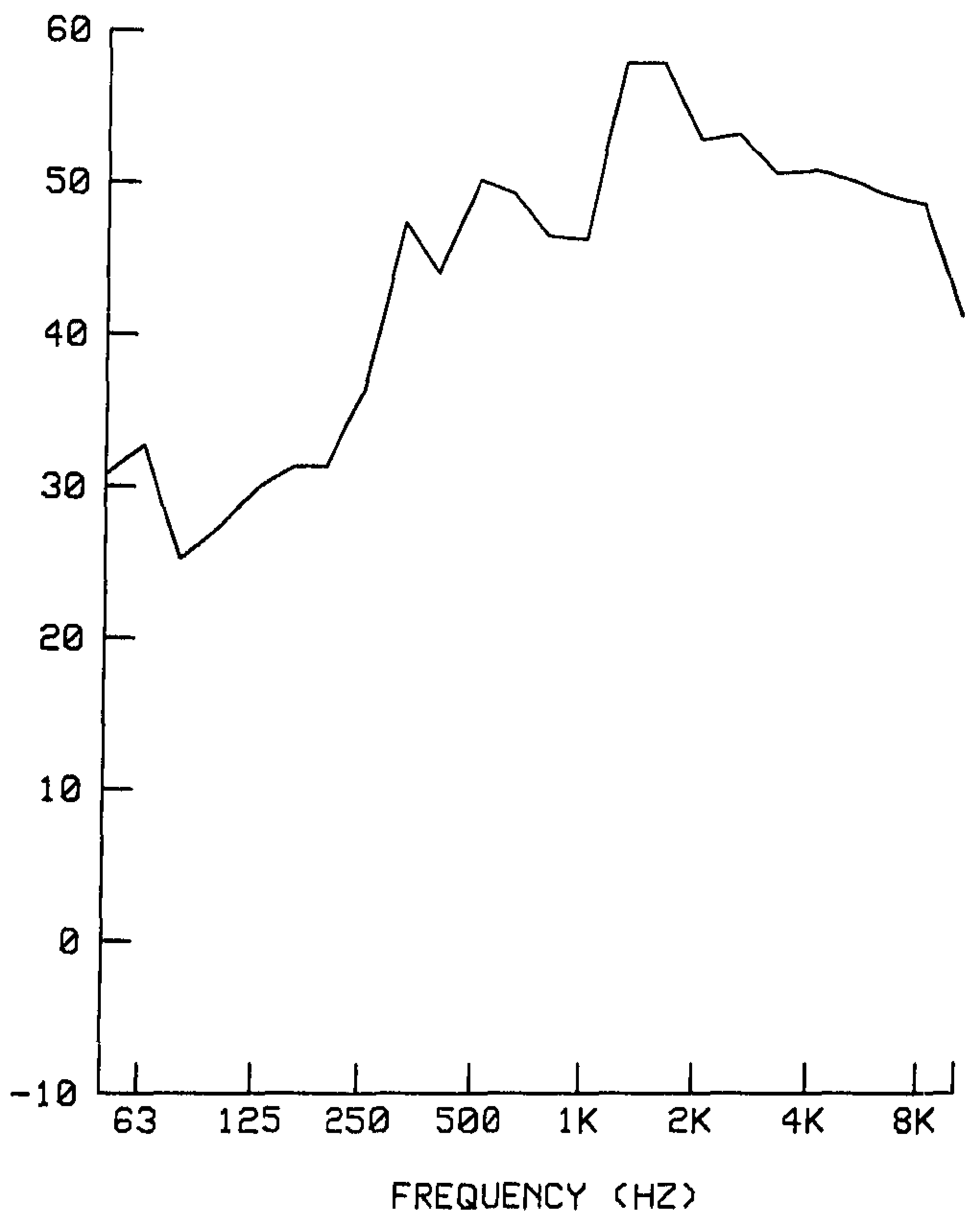
EPA SIGNAL #7, HAIR DRYER, 9/9/77



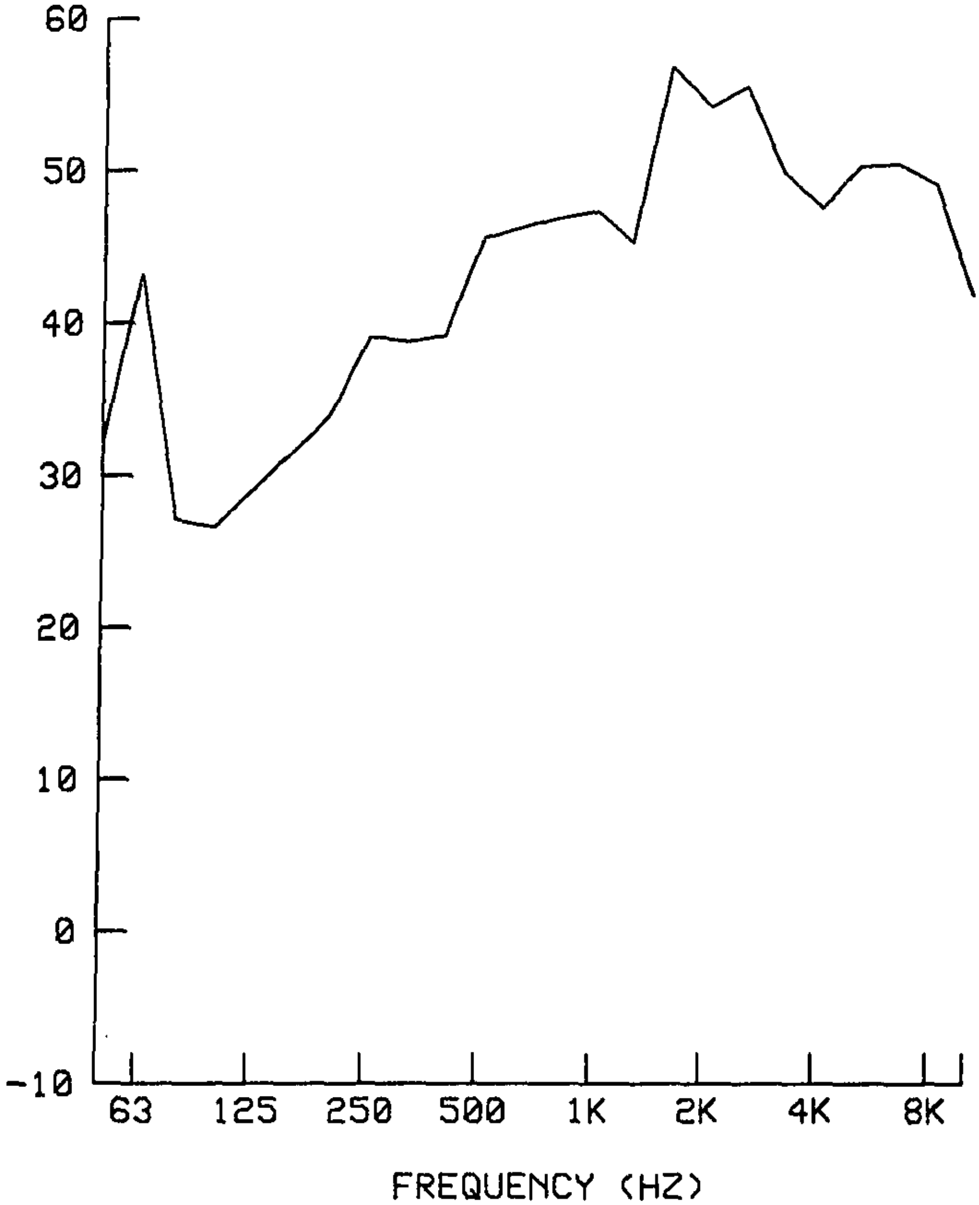
EPA SIGNAL #8, LATHE, 9/9/77



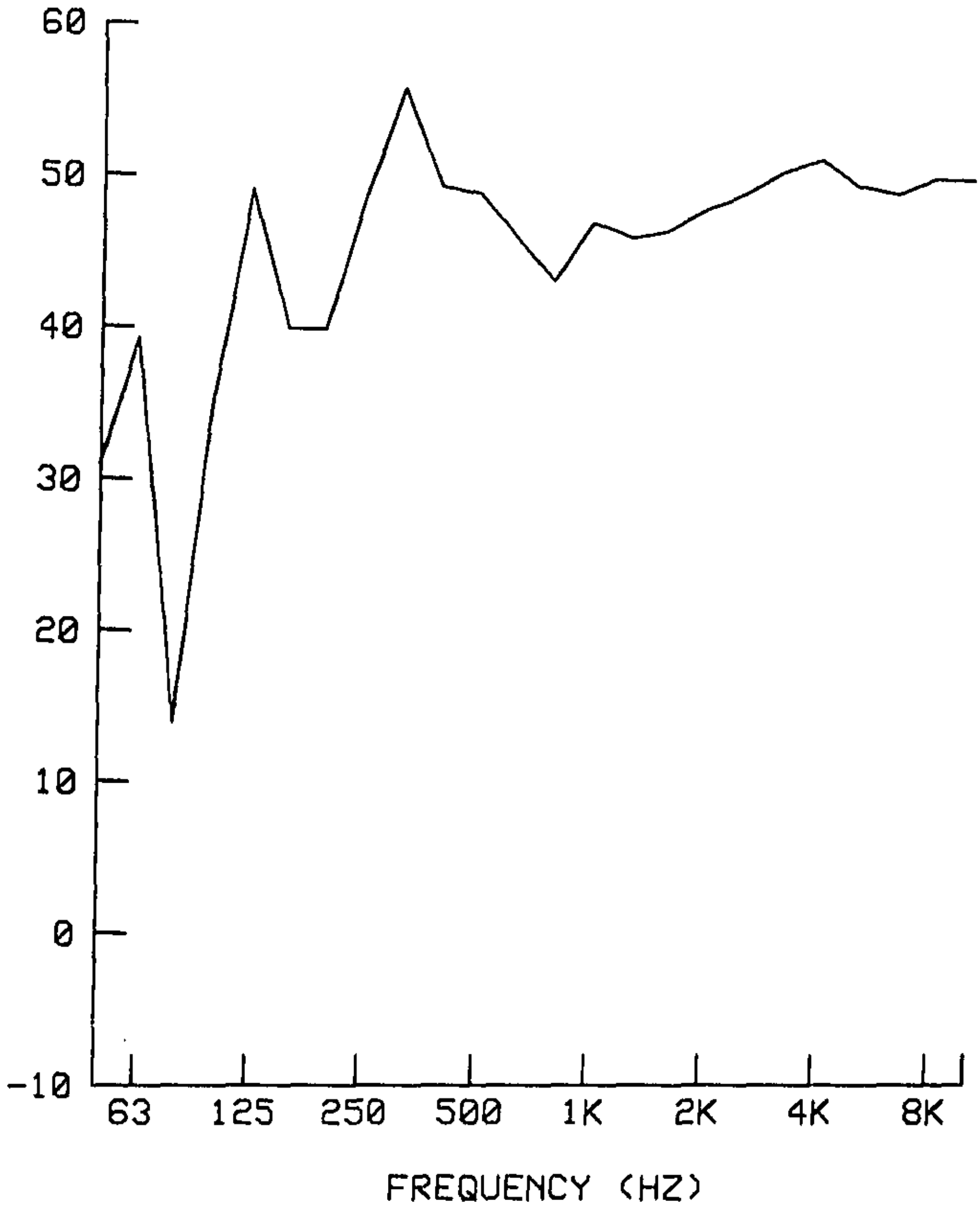
EPA SIGNAL #9, ROUTER, 9/9/77



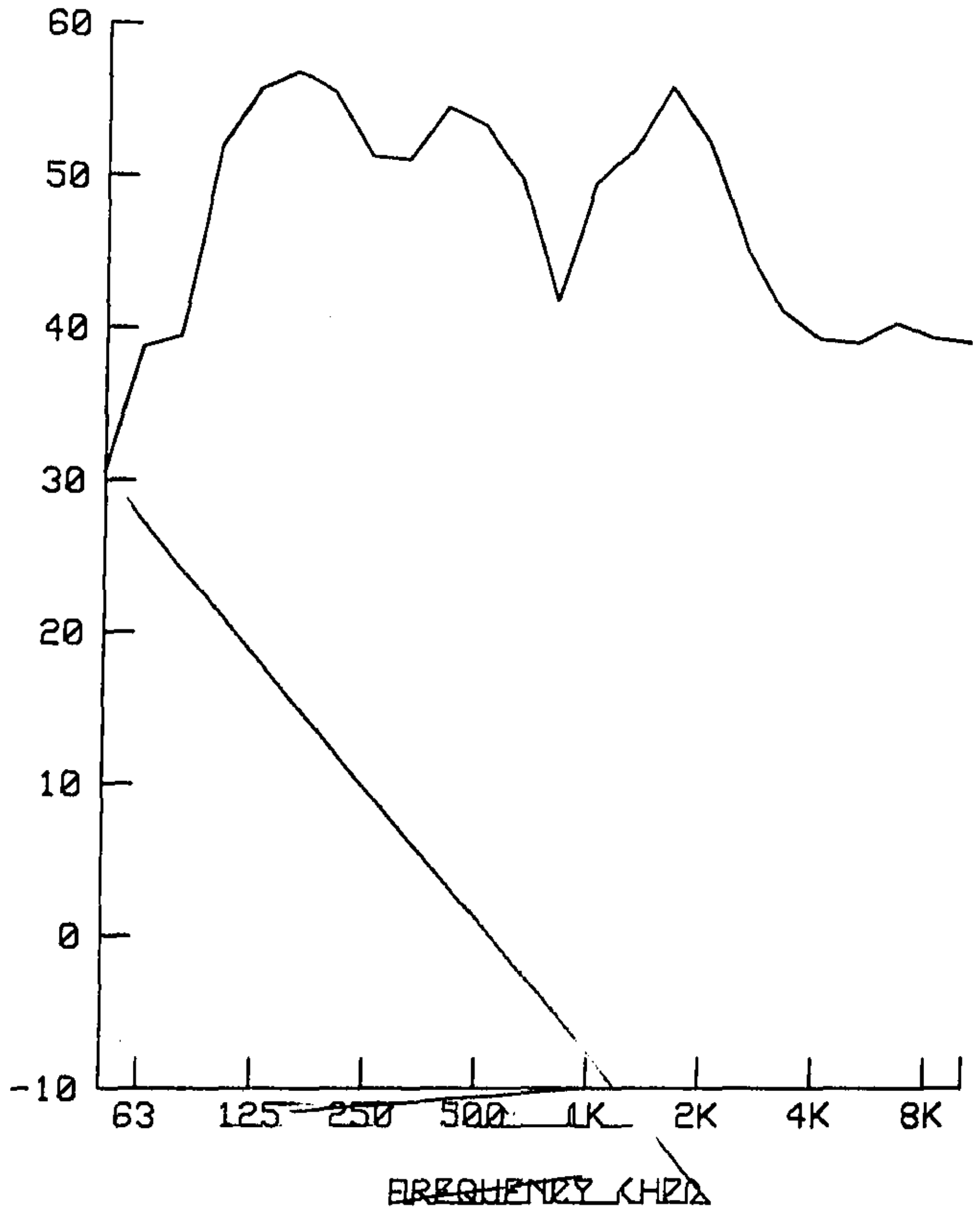
EPA SIGNAL #10, BELT SANDER, 9/9/77



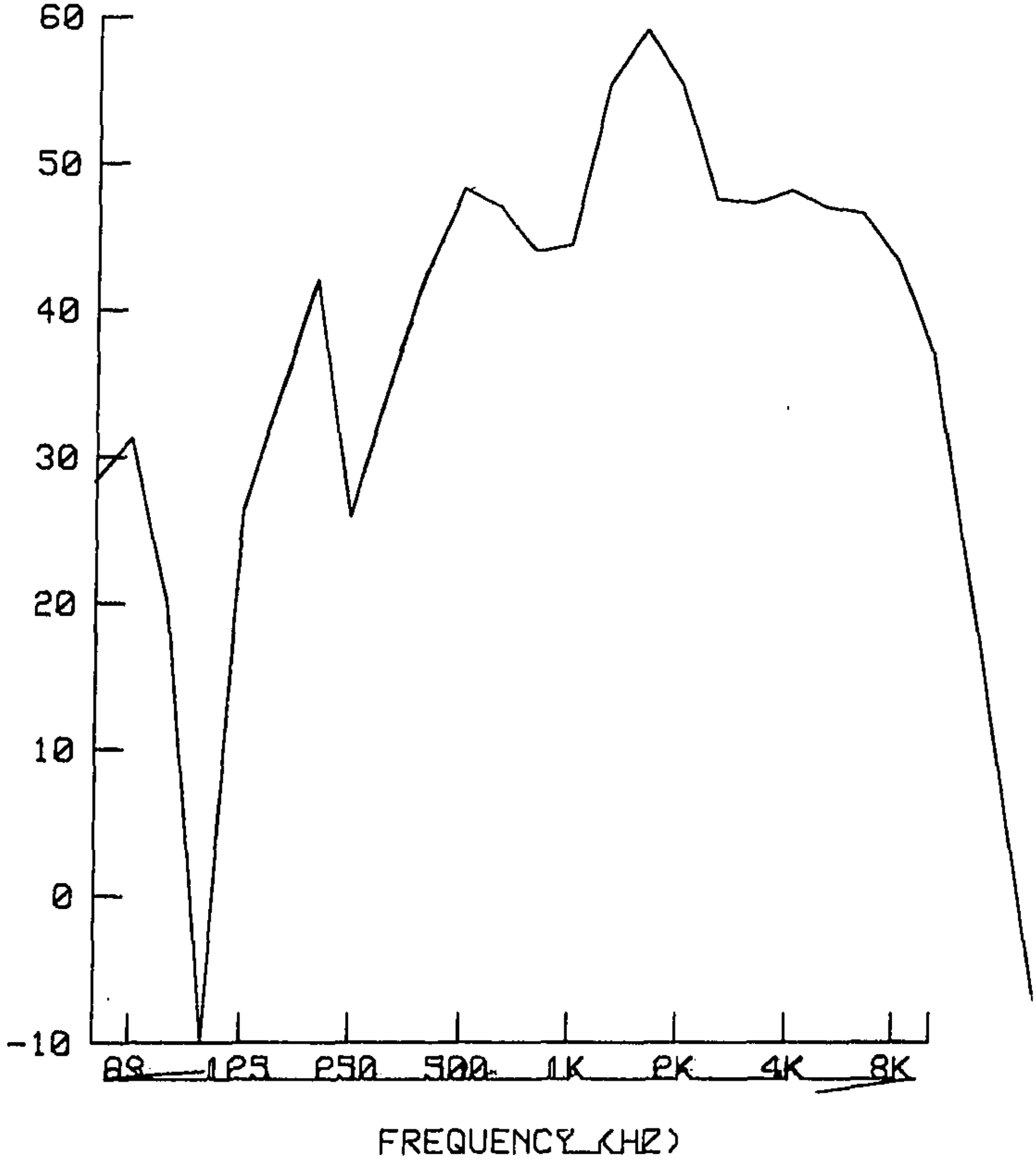
EPA SIGNAL #11, 1/2' DRILL, 9/9/77



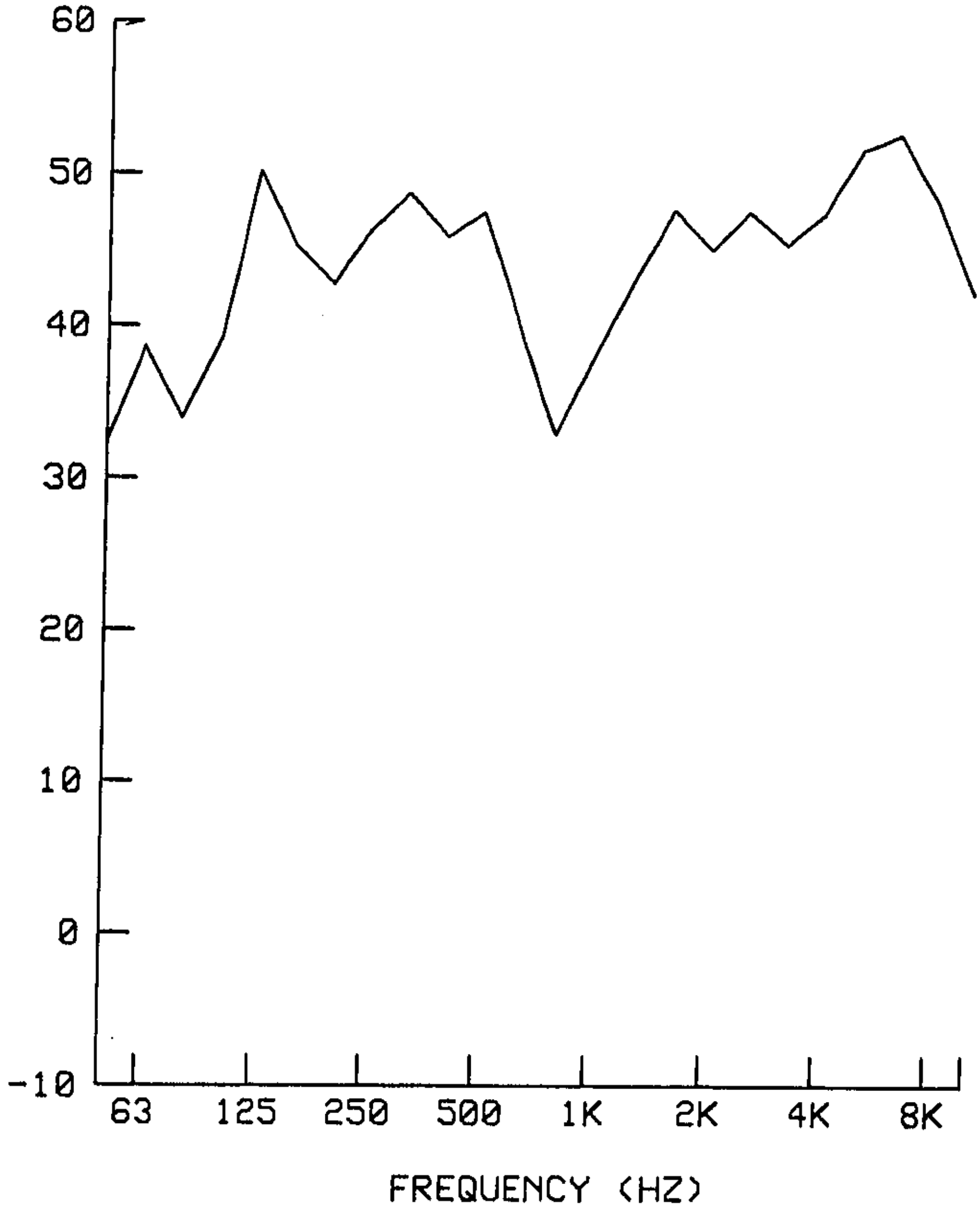
EPA SIGNAL #12, RADIAL ARM SAW, 9/9/77



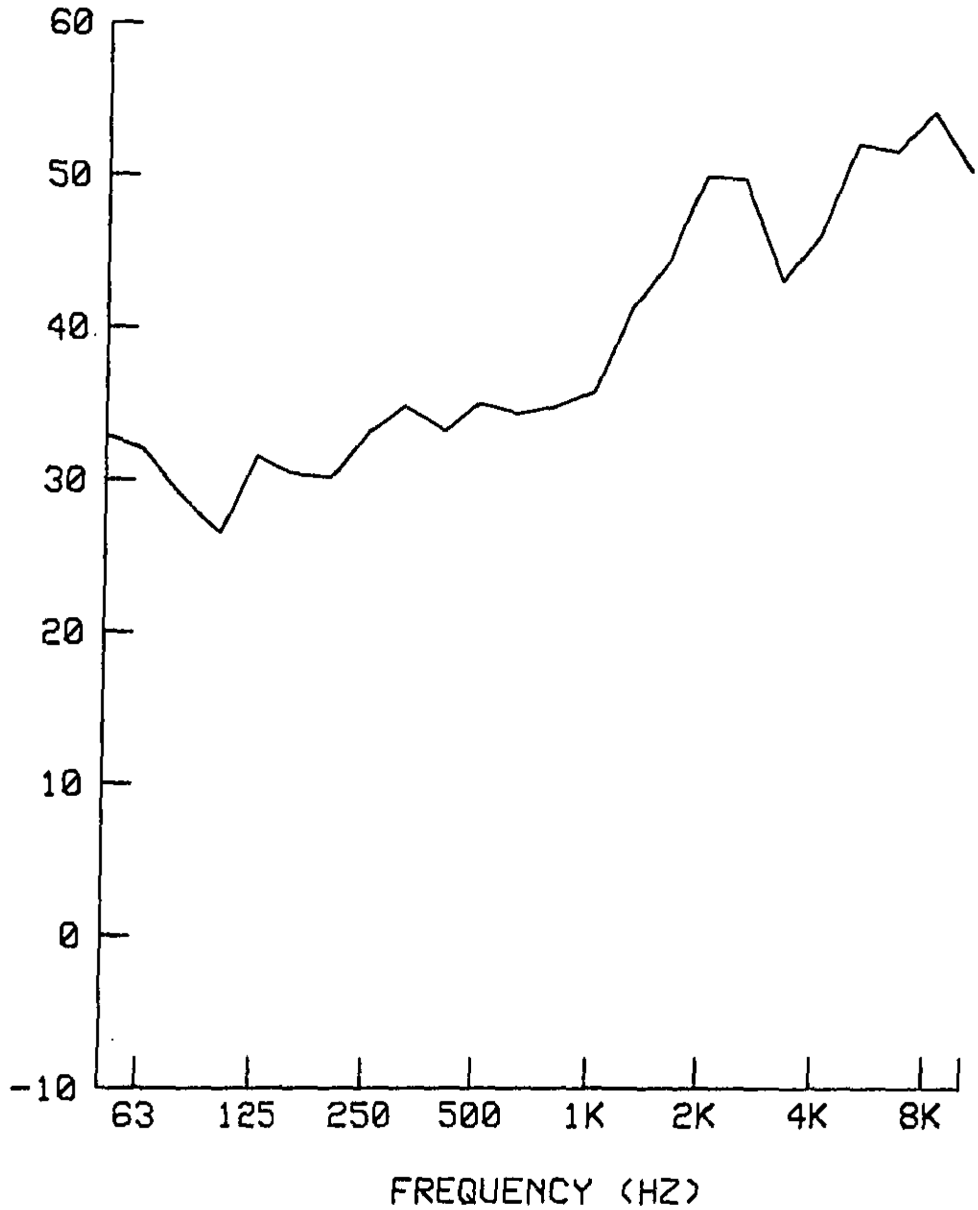
EPA SIGNAL #13, AIR COMPRESSOR, 9/9/77



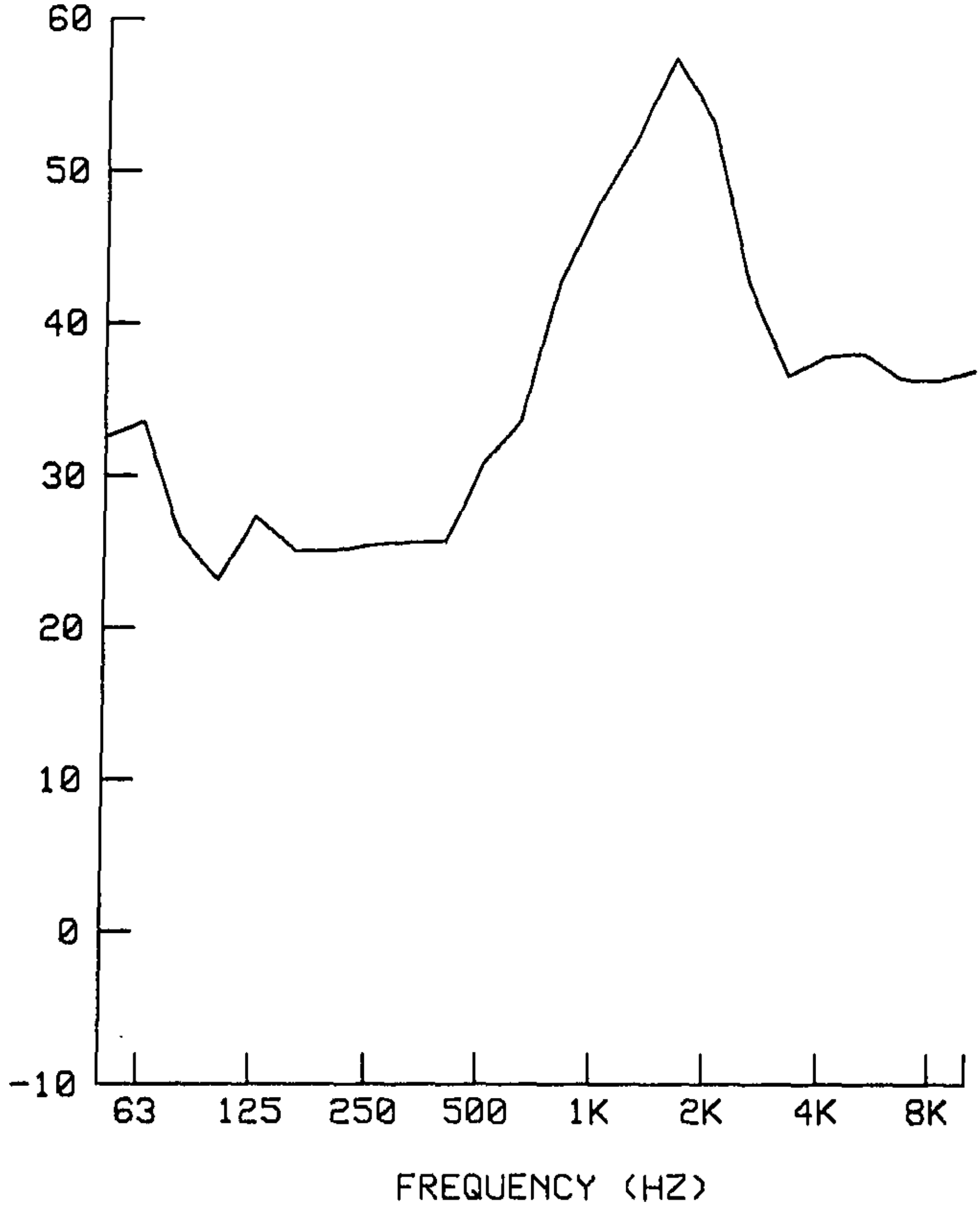
EPA SIGNAL, #14, MODEL AIRPLANE, 9/9/77



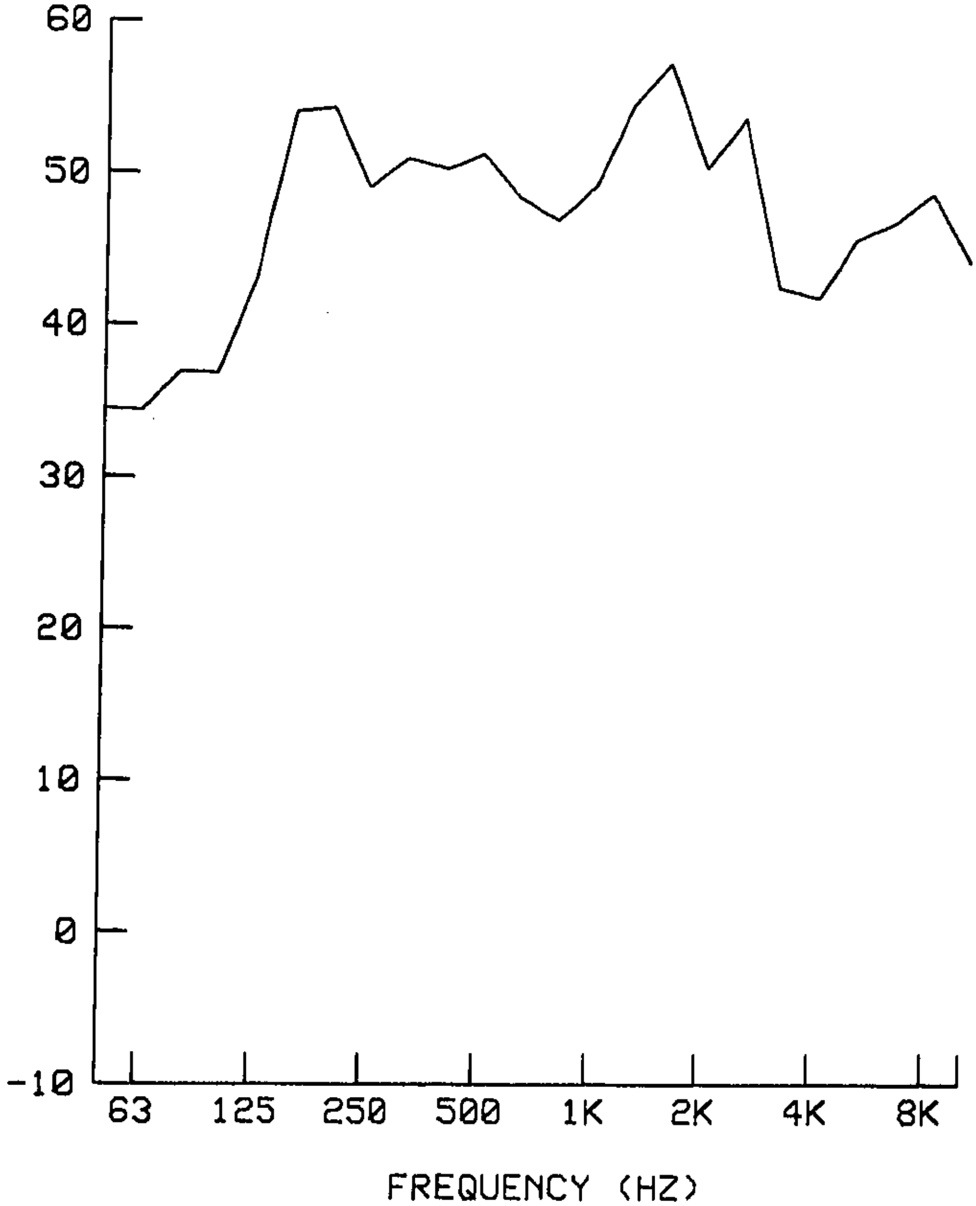
EPA SIGNAL, #15, TYPEWRITER, 9/9/77



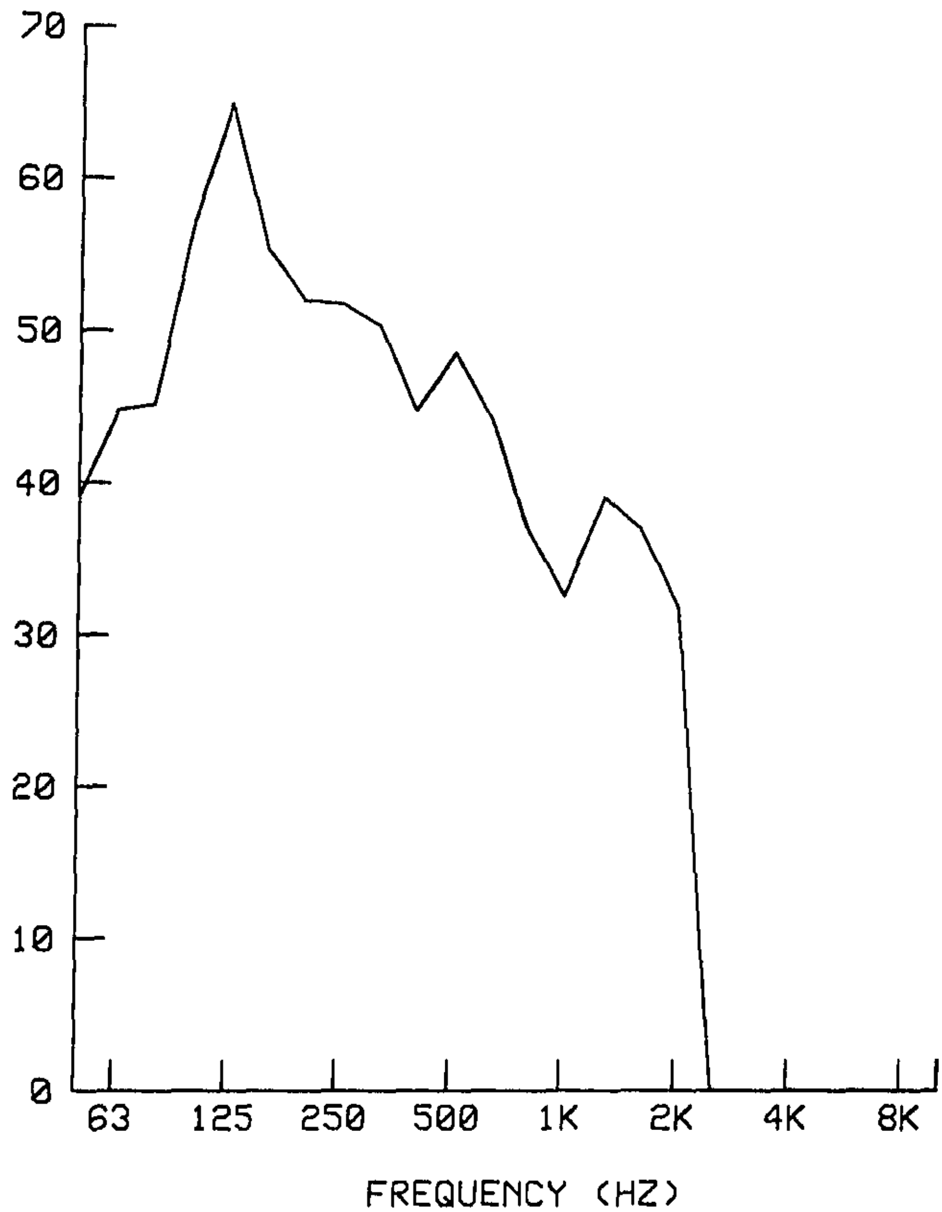
EPA SIGNAL #16, TOY CAR, 9/9/77



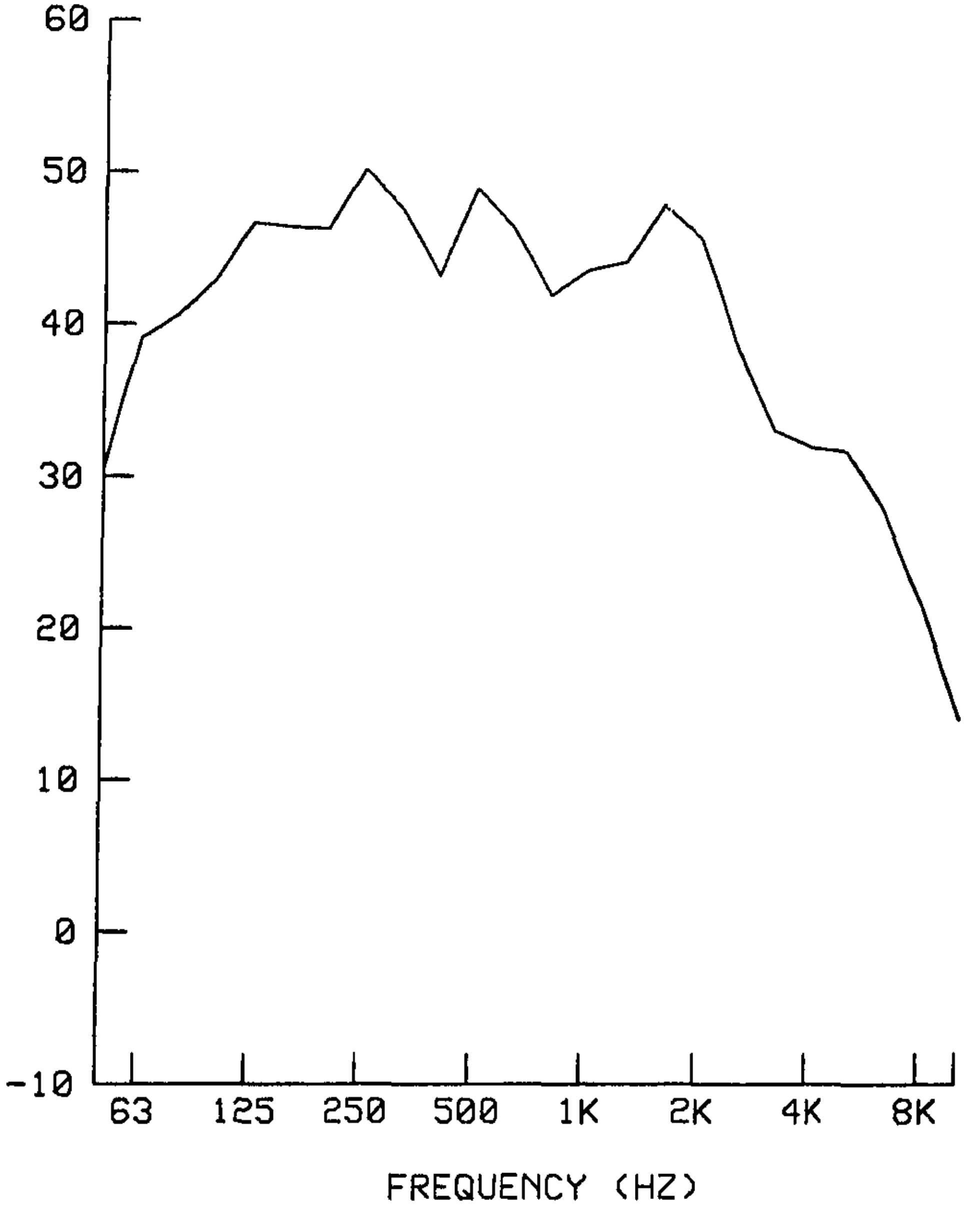
EPA SIGNAL #17. TOY DOG. 9/9/77



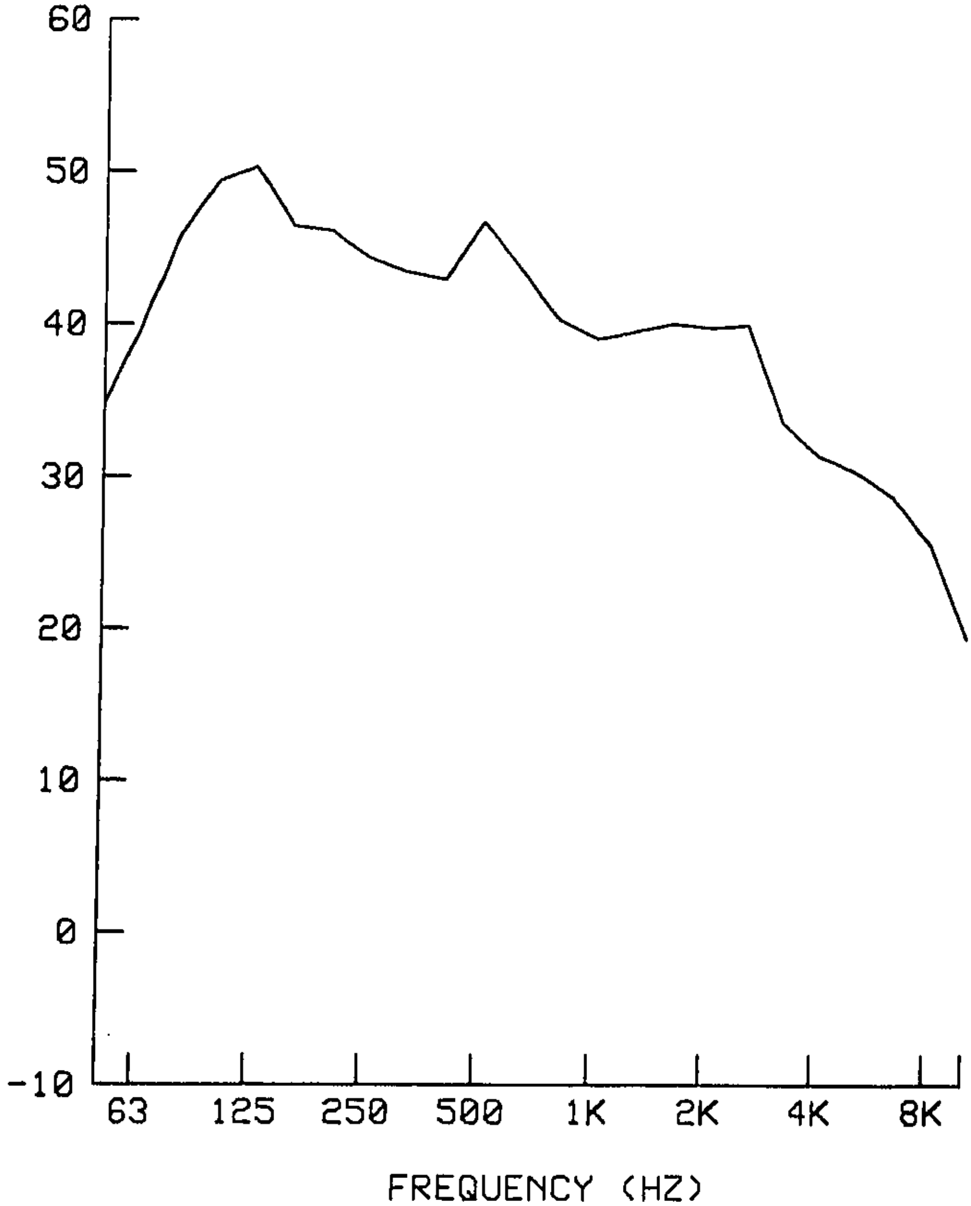
EPA SIGNAL #18, VACUUM CLEANER, 9/9/77



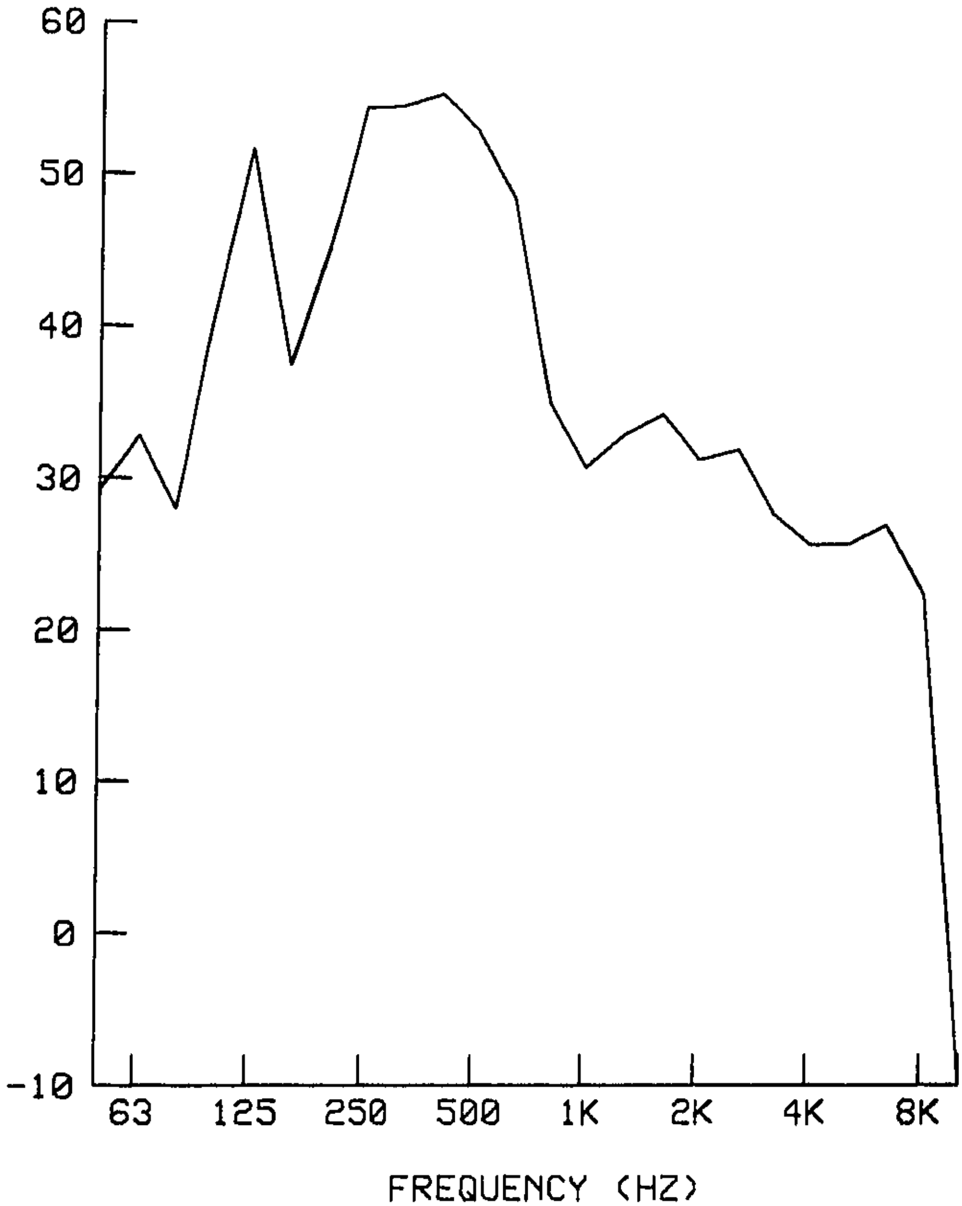
EPA SIGNAL #19. AIR CONDITIONER. 5/9/77



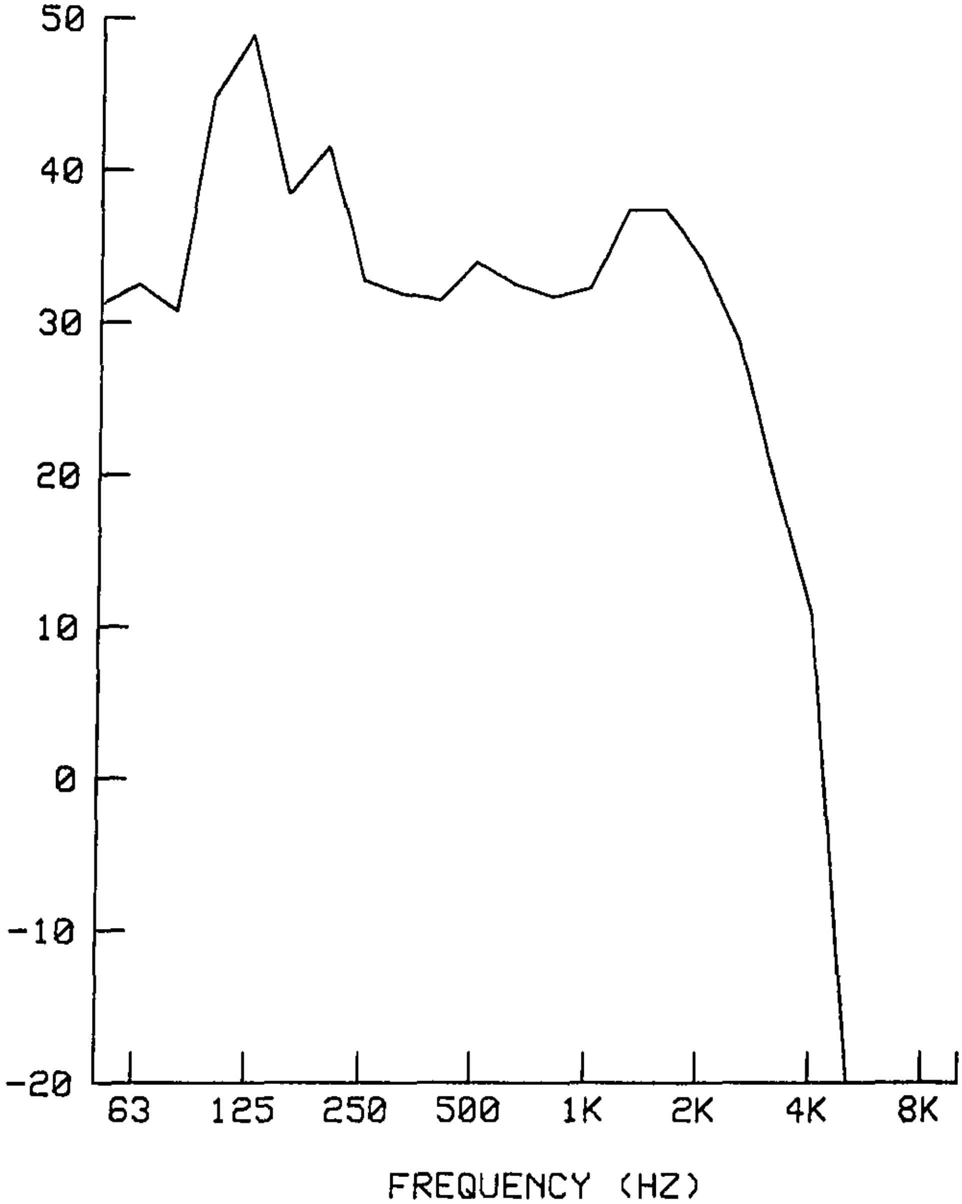
EPA. SIGNAL #20. GARBAGE COMPACTOR. 9/9/77



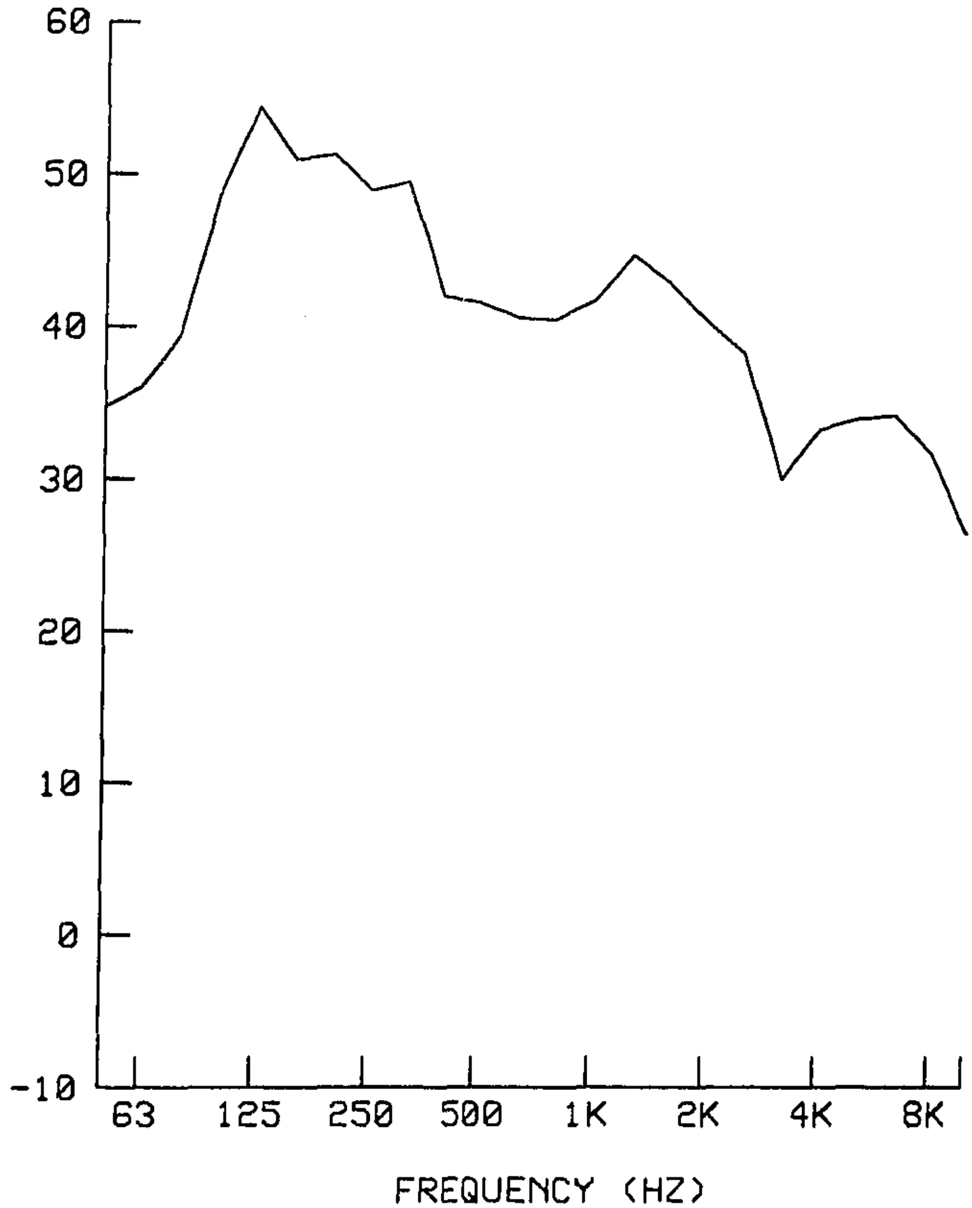
EPA. SIGNAL #21. TRAIN. 9/9/77



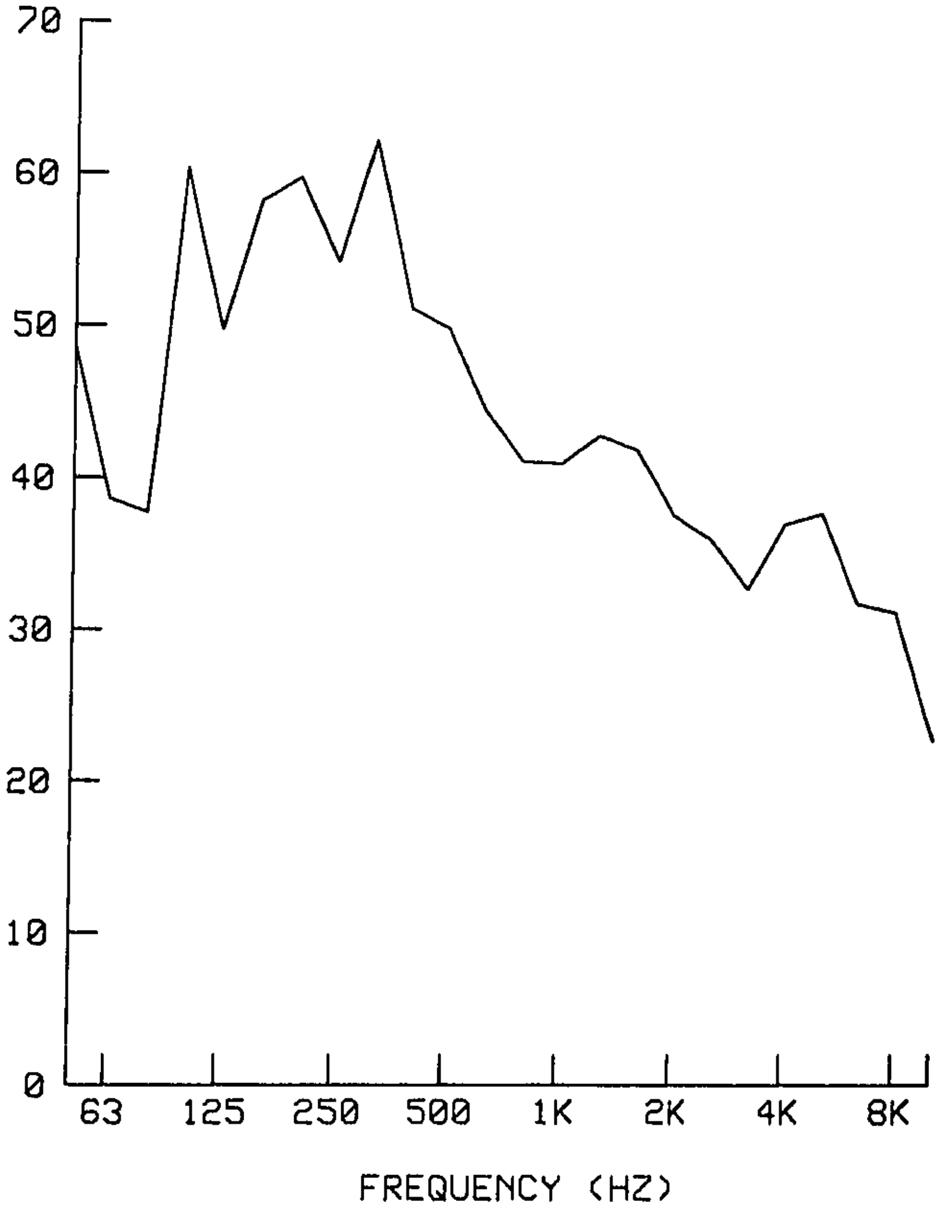
EPA . SIGNAL #22. MOTORCYCLE. 9/9/77



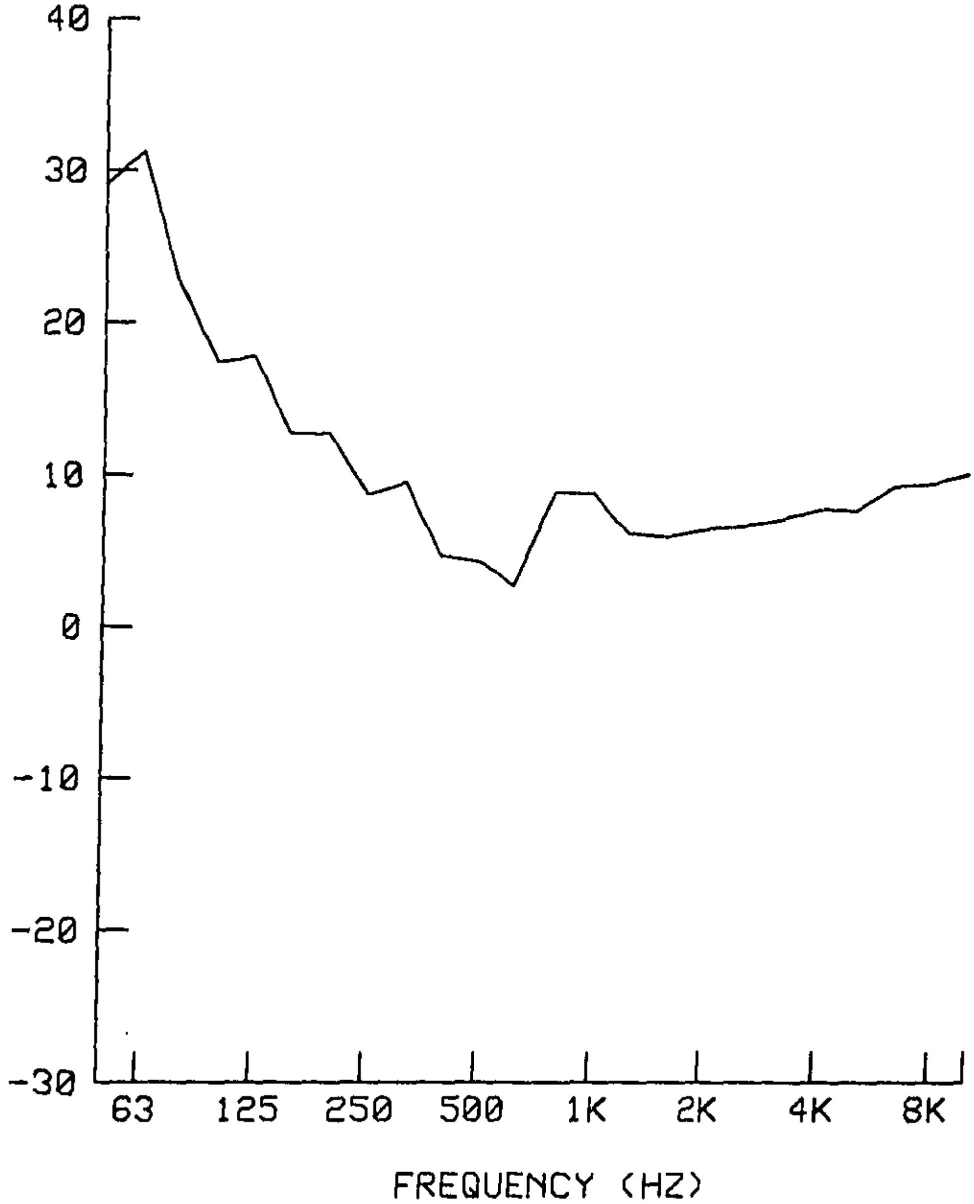
EPA, SIGNAL #23, CAR, 9/9/77



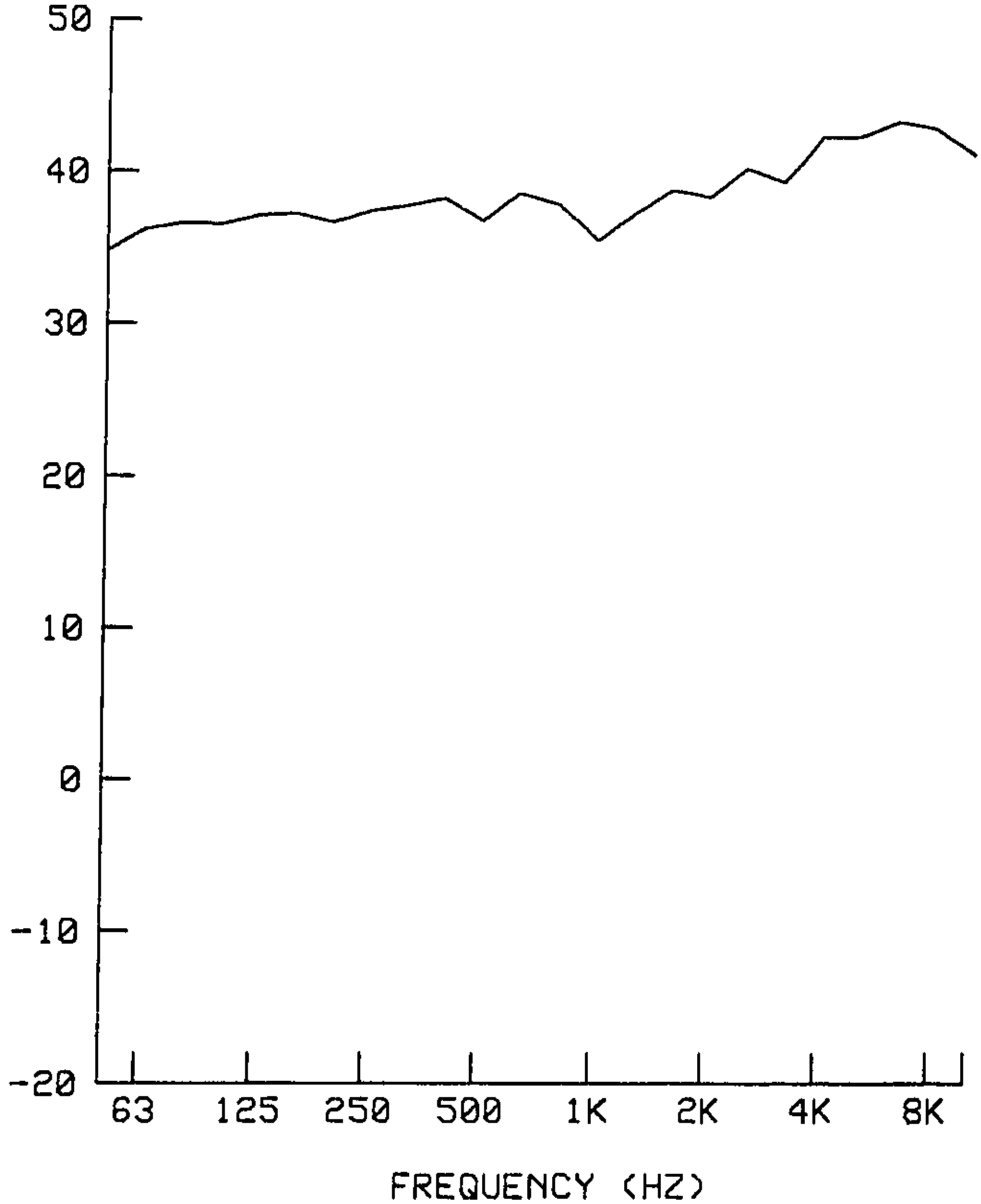
EPA, SIGNAL #24, BUS, 9/9/77



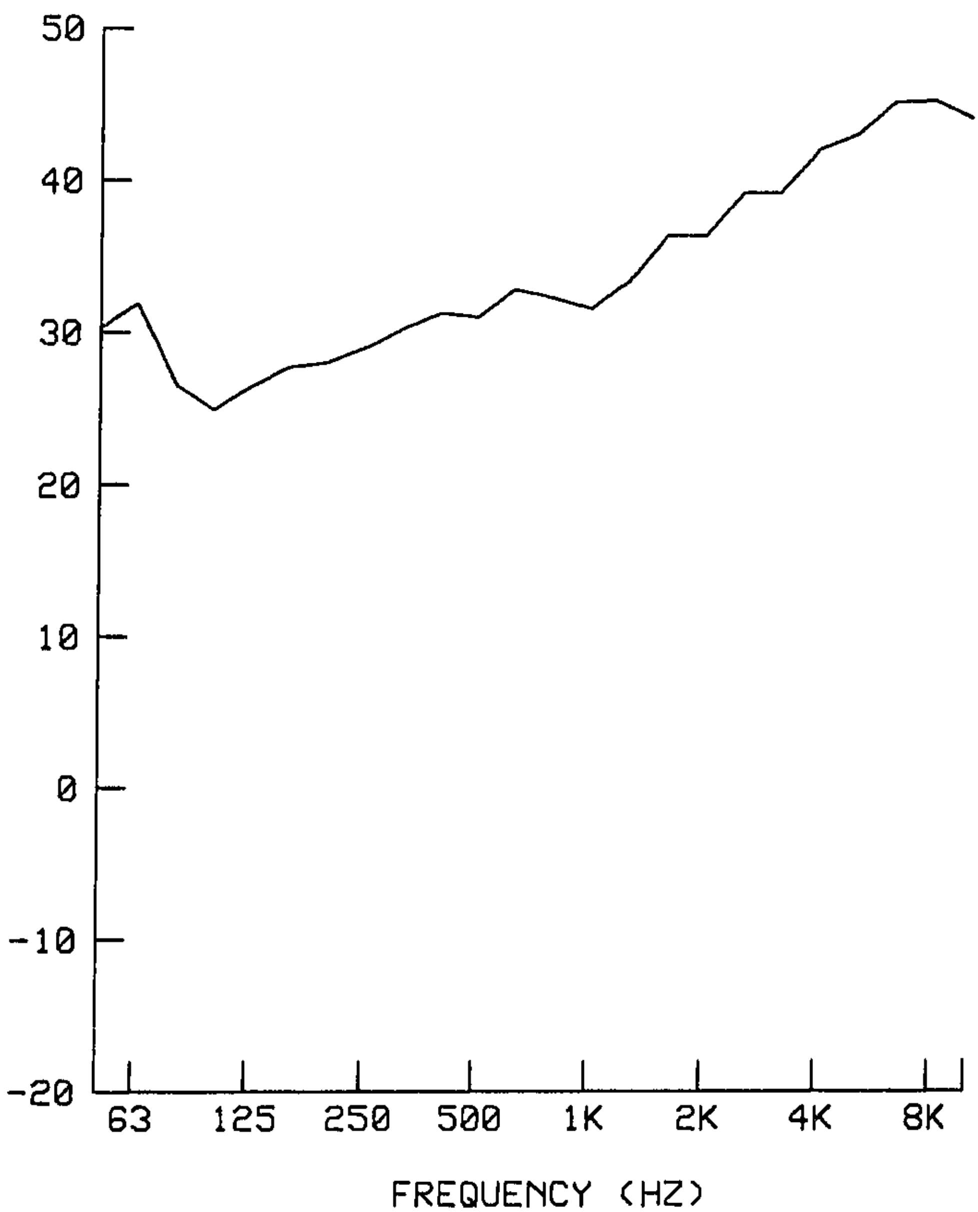
EPA, SIGNAL #25, LAWN MOWER, 9/9/77



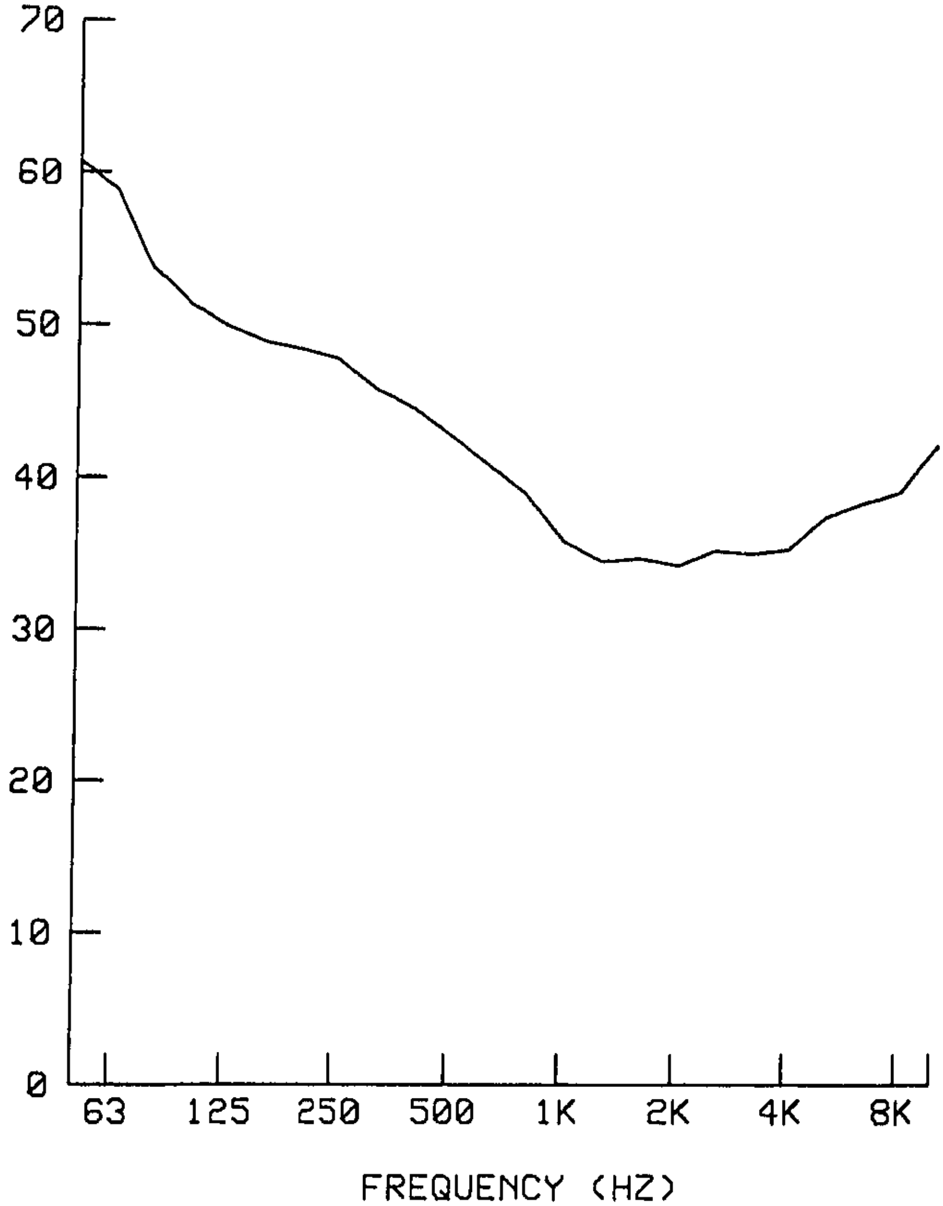
EPA, AMBIENT IN CHAMBER.. 9/9/77



EPA, BKG 3, FLAT SPECTRUM, 9/9/77



EPA, BKG 2, RISING SPECTRUM, 9/9/77



EPA, BKG 1, PNC-40, 9/9/77