37. JUDGMENT TESTS OF AIRCRAFT NOISE

By Karl D. Kryter, Paul J. Johnson,
and James R. Young

Stanford Research Institute

SUMMARY

Psychological tests were conducted to determine the ability of various physical units to predict the judged perceived noisiness of the sounds from a variety of aircraft operating under landing and take-off power. The best agreement between objective measures and subjective judgments of the noisiness or unacceptability of aircraft noise of all types is generally found by calculating the tone-corrected effective perceived noise level in $EPNdB_t$ from 1/3-octave band spectra taken every 1/2 second during the noise cycle. When low-frequency aircraft noise is compared with high-frequency aircraft noise, a systematic overestimation of the perceived noisiness of the lower frequency noise is obtained by the various PNdB units.

INTRODUCTION

Over the past 20 to 30 years, research has led to the development of a number of ways of measuring noise that purport to be correlated with the measurement or response of man's auditory system to noise. These methods run the gamut from simple peak sound-level meter readings to weightings of 1/3-octave band spectra taken every 1/2 second during the duration of a noise occurrence. Although the measurement procedures vary, the purpose and goal of making these measurements is to estimate or predict how unwanted, unacceptable, or noisy the sounds being measured are perceived to be by people.

Most of the psychological tests concerned with establishing the relationships between perceived noisiness and the physical aspects of aircraft noise have been conducted in the laboratory with recordings of the noise; relatively small groups of subjects and often a restricted variety of aircraft noises were used. The present tests were designed to permit as valid an examination as possible of the ability to predict from physical measurements the judged acceptability to people of a wide variety of actual aircraft noise when heard in or outside typical homes. In addition, a number of ancillary questions related to the generation and propagation of the noise and the reaction of house structures to the noise were to be investigated. These tests were recently conducted at the NASA Wallops Station.

The present report is concerned primarily with the relation between the psychological judgment of people outdoors and physical measurements made from recordings of the noise from microphones located outdoors. Measurements of house vibrations at Wallops

Station are reported in reference 1. Additional technical reports by the National Aeronautics and Space Administration and Stanford Research Institute on various aspects of the studies will be prepared later as further analyses of the data are completed.

PROCEDURE

A frame house, a brick-veneer house, and a large yard near one of the houses, all located in a residential area of Wallops Station, were chosen as the test sites. The subjects were adults, primarily housewives, selected from communities in the local area. Figure 1 is a photograph of some of the subjects as seated for the tests. The subjects were tested with an audiometer and all were found to have normal hearing (±15 dB from audiometric zero). The subjects were paid and given careful instructions prior to and during the tests as to the importance of the tests and the nature of the task they were to perform.

The fundamental task of the subjects was to mark on an answer sheet which of two aircraft sounds presented to them in a brief period of time they considered to be the least acceptable, assuming these noises were heard in or near their home 20 to 30 times per day. The subjects also rated each noise on a scale ranging from completely acceptable to completely unacceptable.

Because of the very large number of aircraft tested, it was not possible to pair, for the judgments, each aircraft noise with every other aircraft noise. Instead, two of the aircraft were chosen to provide a reference, or standard, noise, and this noise was paired with the noise from each of the other aircraft when operating under landing and take-off power. The reference, or standard, aircraft chosen were the turbojet 880 and the 1049G (Super Constellation), a propeller-driven aircraft with reciprocating engines (recip-prop).

The aircraft were flown so that about 5 minutes elapsed between pairs of noises, and usually 1 minute elapsed between each member of a pair. The altitude for the operational conditions for each aircraft was carefully monitored and controlled for all flights. An attempt was made to operate the reference aircraft at an altitude that would be reasonable for that aircraft at about 2 to 3 miles from an airport. Figure 2 is a schematic illustration of the flight paths followed by aircraft for the tests. Table I summarizes the aircraft tested and number of overflights made per aircraft.

Physical Measurements

Tape recordings made of the aircraft noises reaching the ground were played through 1/3-octave band filters. Each filter output was passed through an envelope detector, smoothed, and sampled every 1/2 second. These samples were digitized and stored on magnetic tape and were used for subsequent calculation of the following units:

(1) Max dB(A), dB(B), dB(C), dB(N)

(2) Max phons (Stevens), PNdB, $PNdB_{t_1}$, and $PNdB_{t_2}$

(3) Peak phons (Stevens), PNdB, $PNdB_{t_1}$, and $PNdB_{t_2}$

(4) E (effective) dB(A), dB(N), PNdB, $PNdB_{t_1}$, and $PNdB_{t_2}$

(5) EE (estimated effective) PNdB, $PNdB_{t_1}$, and $PNdB_{t_2}$

References 2 and 3 give a detailed description of these units. These units of noise measurement appear at the present time to be the best available units for the evaluation of the perceived noisiness of the sound from aircraft. An additional method is that of calculating loudness levels, in phons, by the 1/3-octave band method of Zwicker (ref. 4). It is planned to obtain max phons (Zwicker) for these aircraft noises for their inclusion in the final report of the tests conducted at Wallops Station.

One-third—octave band spectra present outdoors when the overall sound pressure level (SPL) was at max dB(C) are shown in figure 3 for representative flights for each of the aircraft tested.

Psychological Measurements

The paired-comparison tests which are believed to provide the most essential psychological data from this study are scored and interpreted as follows:

Step 1.- The percent of listeners in a group (for this report a group of 33 people) who preferred the reference aircraft noise when it appeared first in a given pair and when it appeared second in the same pair are averaged.

Step 2.- The percent obtained in step 1 is plotted against the level, measured by a given physical unit, of the comparison aircraft noise. Inasmuch as the level of the comparison noise was systematically varied, the percent of people, in general, who preferred the reference noise increased as the level of the comparison noise increased. An attempt was made to have the comparison noise vary over a range that caused the percent of people preferring the reference aircraft noise to change from near zero percent to near 100 percent. Sample plots of the data are shown in figure 4.

Step 3.- On each function, such as those shown in figure 4, a perpendicular is dropped to the abscissa from the point where the 50-percent line crosses the curve drawn through the data points.

Step 4.- The value obtained in step 3 is taken as the level, for the given unit of measurement, required for the comparison noise if and when it is to be perceived as equal to the reference aircraft noise in unacceptability or noisiness.

Step 5.- The difference, if any, between the reference and comparison noises when judged to be equal is taken as the index of the ability of each of the physical units to properly measure or indicate the perceived noisiness of each pair of sounds.

RESULTS AND DISCUSSION

Table II presents a summary of averaged data obtained when the noise from jet aircraft was judged against that from other jet aircraft or from turboprop aircraft, recip-prop aircraft, or helicopters. Two equally important indicators that show the accuracy with which an objective measure predicts the subjective judgment data are shown in table II and are described as follows:

(1) The average of the differences between an objective measure for the reference and comparison aircraft noise when they are judged to be equally noisy. If the physical units were perfectly correlated with the psychological data, the average of the differences would be zero for each pair; that is, when the aircraft noises were judged to be equal, they would be measured physically as being equal.

(2) The range of the differences between an objective measure for the reference and comparison aircraft noise when they are judged to be equally noisy. It is important that an objective measure be reasonably accurate with respect to the most common or most prevalent aircraft noises evaluated. This accuracy is reflected in the average of the differences between the objective measures of any two noises judged to be subjectively equal.

Two particular conclusions can be drawn from table II. First, the average differences for the better units are significantly smaller when the comparison and reference noises are both high-frequency jet noises than when one of the noises is a low-frequency propeller-aircraft noise. These differences are about 1.5 dB when the comparison and reference noises are both high-frequency jet noises and about 3.0 dB when one is a low-frequency propeller-aircraft noise. Appropriate modifications to procedures for measuring noise in the lower frequency bands should reduce the differences between the physical measures and judged noisiness and thereby provide an effective perceived noise level that would predict, with an average accuracy of about 1.0 dB, judgments of the noisiness or unacceptability of aircraft noise regardless of its source, spectral complexity, and, within limits, duration. Second, when all types of aircraft noises and both the average differences and the total range of differences between the values for the references and comparison noises are considered, the most accurate units of measurement are usually EPNdB_t.

The relative accuracy of the various objective measures is illustrated in table III, which shows how each unit ranks with regard to the average of the average differences and the range of differences. A unit with the rank of one would have the best agreement with the subjective judgments.

Two other recent studies (see refs. 5 and 6) have been made of subjective judgments of aircraft noise in which all, or nearly all, of the objective measures used in the present study were evaluated. Table IV shows that EPNdB has the highest rank in each of the studies. It should be pointed out, however, that the practical significance between some of the differences in these measures, an amount equivalent to less than 1.0 dB in level in some cases, is perhaps questionable.

CONCLUSIONS

The following conclusions are based on the results from judgment tests of aircraft noise:

1. The best agreement between objective measures and subjective judgments of the noisiness or unacceptability of aircraft noise of all types is generally found by calculating the tone-corrected effective perceived noise level in EPNdB_t from 1/3-octave band spectra taken every 1/2 second during the noise cycle.

2. When aircraft noise containing its energy predominately in the lower frequencies is compared with predominately high-frequency aircraft noise, a systematic overestimation by about 3.0 dB of the perceived noisiness of the lower frequency noise is obtained by the various PNdB units.

REFERENCES

1. Mayes, William H.; Findley, Donald S.; and Carden, Huey D.: House Vibrations Significant for Indoor Subjective Response. Conference on Progress of NASA Research Relating to Noise Alleviation of Large Subsonic Jet Aircraft, NASA SP-189, 1968. (Paper No. 39 herein.)

2. Kryter, K. D.: Concepts of Perceived Noisiness, Their Implementation and Application. J. Acoust. Soc. Amer., vol. 43, no. 2, Feb. 1968, pp. 344-361.

3. Stevens, S. S.: Procedure for Calculating Loudness: Mark VI. J. Acoust. Soc. Amer., vol. 33, no. 11, Nov. 1961, pp. 1577-1585.

4. Zwicker, E.: A Graphic Method of Determining Loudness and Loudness Level From the "Third-Octave" Level Diagram. Frequenz, vol. 13, no. 8, Aug. 1959, pp. 234-238.

5. Hecker, Michael H. L.; and Kryter, Karl D.: Comparisons Between Subjective Ratings of Aircraft Noise and Various Objective Measures. Tech. Rep. NO-68-33, FAA, Apr. 1968.

6. Kryter, K. D.; Johnson, P. J.; and Young, J. R.: Psychological Experiments on Sonic Booms. Annex B of Sonic Boom Experiments at Edwards Air Force Base, NSBEO-1-67 (Contract AF 49(638)-1758), CFSTI, U.S. Dep. Com., July 28, 1967.

TABLE I.- NUMBER OF OVERFLIGHTS FOR JUDGMENT TESTS

[Total of 189 pairs]

Reference aircraft:	
Turbojet. 880	132
Recip-prop, 1049G	80
Comparison aircraft:	
Turbofan (front). 727	22
Turbofan (front). C-141A	22
Turbofan (aft). 990	28
Turbojet. 720	20
Turbojet. 1329 Jet Star	18
Turbojet (afterburner). F-106	12
Turboprop (STOL). CV-7A	16
Turboshaft (helicopter). 204B	20
Turboshaft (helicopter). CH-47	8

TABLE II.- COMPARISON OF OBJECTIVE MEASURES AND SUBJECTIVE JUDGMENTS

Aircraft noise	Indicator of accuracy	Best units					
		EPNdB	EdB(N)	EPNdB_{t_1}	Peak PNdB_{t_1}	Max dB(B)	Max dB(C)
Jet vs jet	Average difference	1.2	1.4	1.5	2.9	2.0	2.1
	Range of differences	4.5	4.5	7.5	11.0	5.5	7.5
Jet vs turboprop, recip-prop, or helicopter	Average difference	2.4	2.5	3.5	5.6	7.0	7.9
	Range of differences	9.0	8.0	10.0	17.0	20.0	25.0

Aircraft noise	Indicator of accuracy	Best units			Worst units		
		EPNdB EdB(N)	EPNdB_{t_1} EPNdB_{t_2}	EEPNdB	Peak PNdB_{t_1}	Max dB(B)	Max dB(C)
Grand average for all aircraft	Average difference	1.7	2.2	2.8	4.0	4.1	4.6
	Range of differences	8.5	10.3	11.0	19.0	20.0	25.0

TABLE III.- RELATIVE ACCURACY WITH WHICH OBJECTIVE MEASURES PREDICT JUDGMENT DATA OBTAINED AT WALLOPS STATION

[20 units used]

Unit name	Rank	Unit name	Rank
EPNdB	1	Max phons	8
EdB(N)	1	Max dB(A)	8
EPNdB_{t_1}	2	Max PNdB_{t_2}	9
EPNdB_{t_2}	2	Max PNdB	9
EEPNdB	3	Max dB(N)	9
EdB(A)	4	Peak PNdB_{t_2}	10
EEPNdB_{t_1}	5	Max PNdB_{t_1}	11
EEPNdB_{t_2}	6	Peak PNdB_{t_1}	12
Peak phons	7	Max dB(B)	13
Peak PNdB	8	Max dB(C)	14

TABLE IV.- BEST AND WORST PREDICTORS OF JUDGMENTS OF AIRCRAFT NOISE

Predictors	Edwards AFB field tests (ref. 5)	FAA laboratory tests (ref. 6)	NASA field tests
Best	$EPNdB_{t_1}$ $EEPNdB_{t_1}$ EEPNdB	$EPNdB_{t_2}$ EPNdB Max dB(N)	EPNdB, EdB(N) $EPNdB_{t_1}$, $EPNdB_{t_2}$ EEPNdB
Worst	Max dB(A) Max dB(C) Max dB(B)	Peak $PNdB_{t_1}$ Max $PNdB_{t_1}$ Max dB(C)	Peak $PNdB_{t_1}$ Max dB(B) Max dB(C)

Figure 1 **L-68-8585**

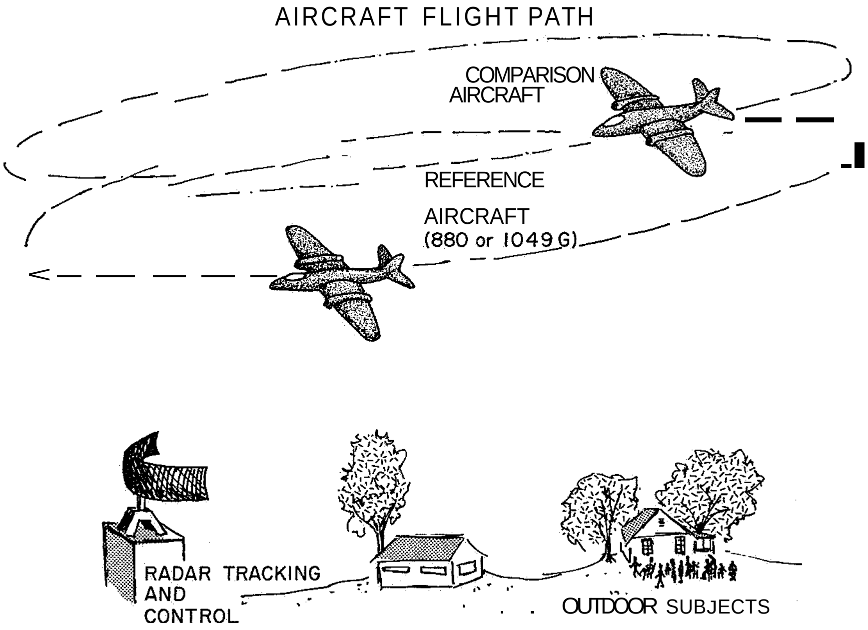

Figure 2

1/3-OCTAVE BAND SPECTRA PRESENT OUTDOORS WHEN OVERALL SPL
(C WEIGHTING) REACHED ITS MAXIMUM

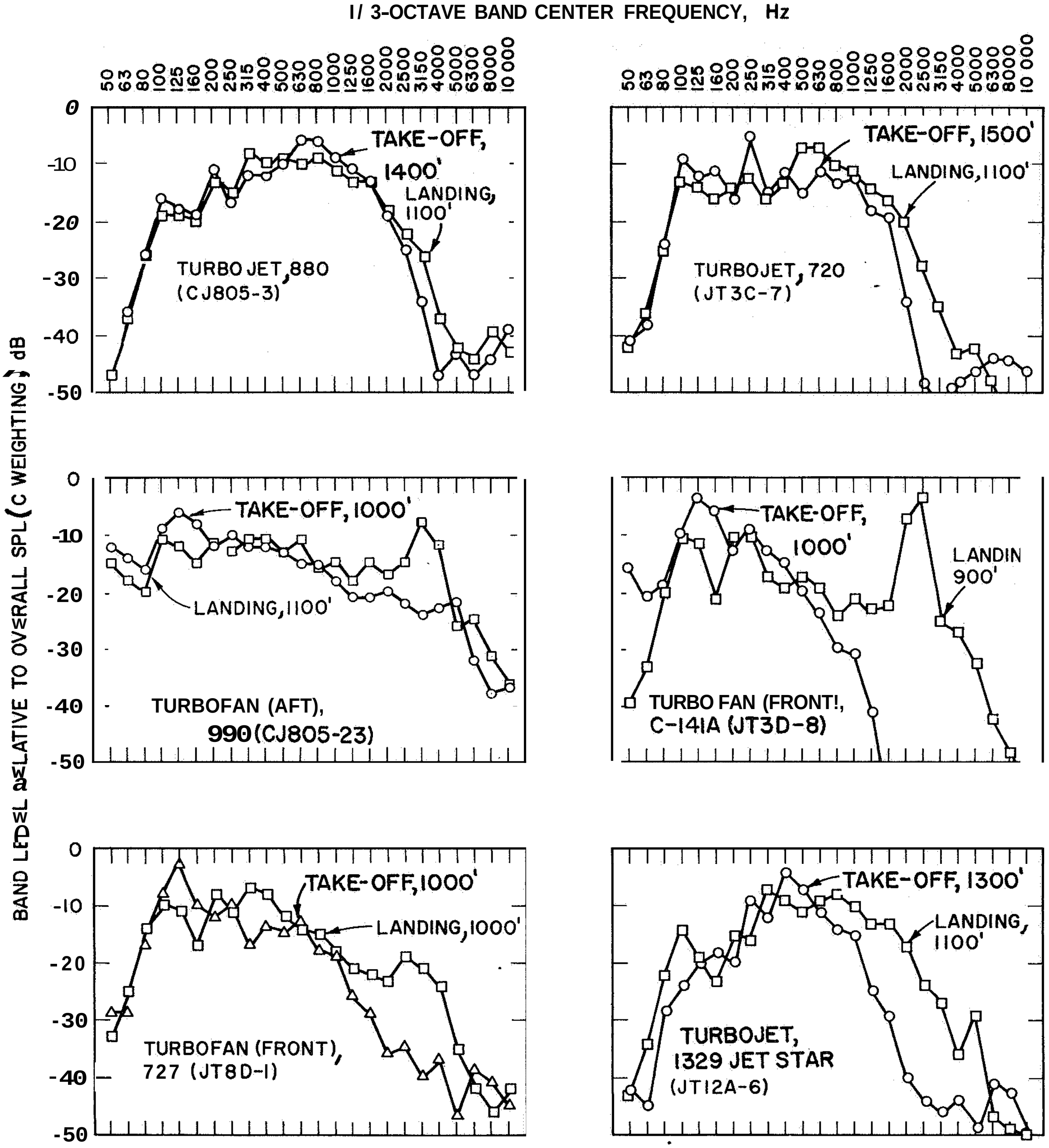

Figure 3(a)

I/3-OCTAVE BAND SPECTRA PRESENT OUTDOORS WHEN OVERALL SPL (C WEIGHTING) REACHED TO MAXIMUM

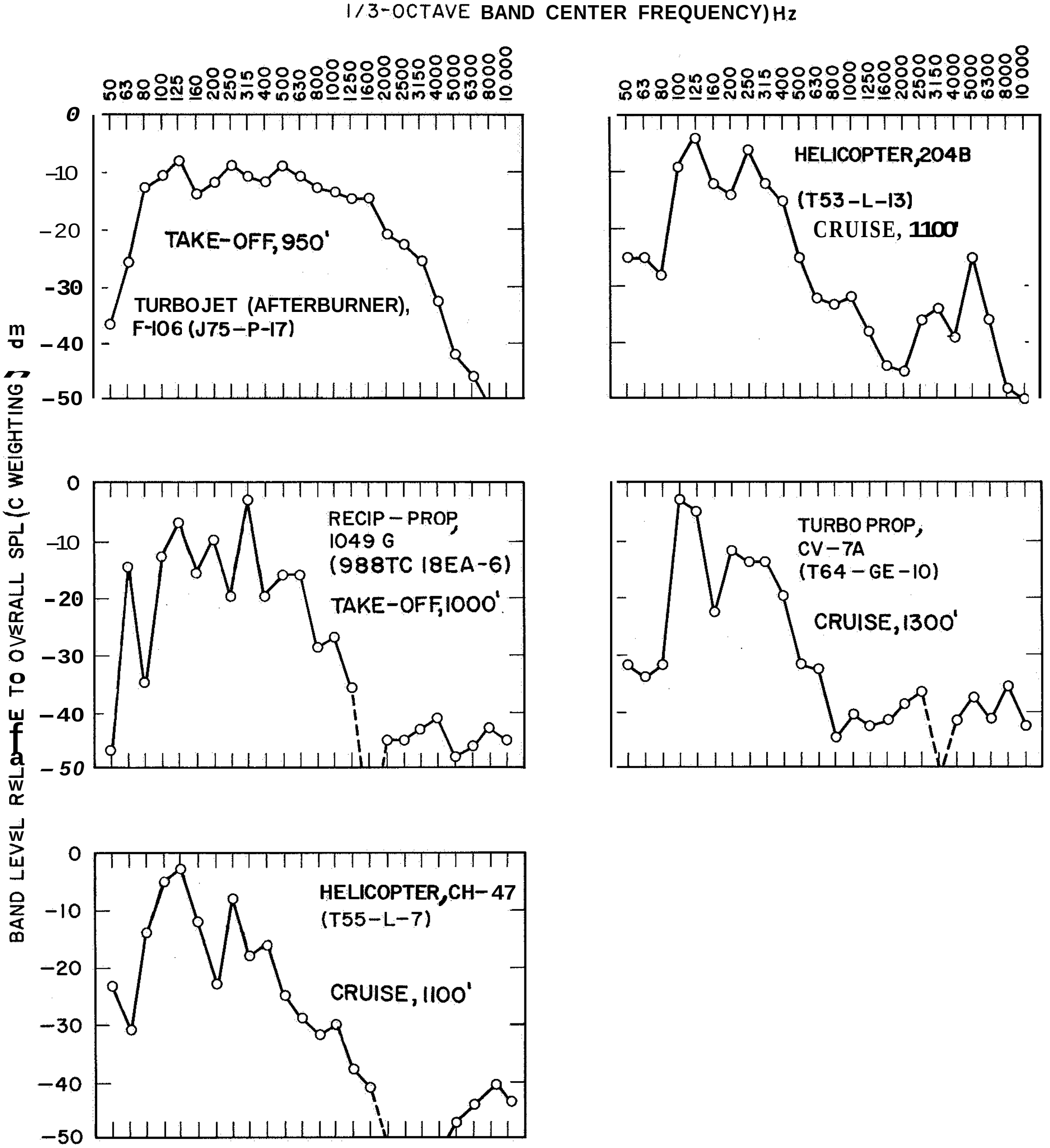

Figure 3(b)

EXAMPLES OF PAIRED-COMPARISON JUDGMENTS OF SUBSONIC NOISE BY OUTDOOR LISTENERS AT WALLOPS STATION

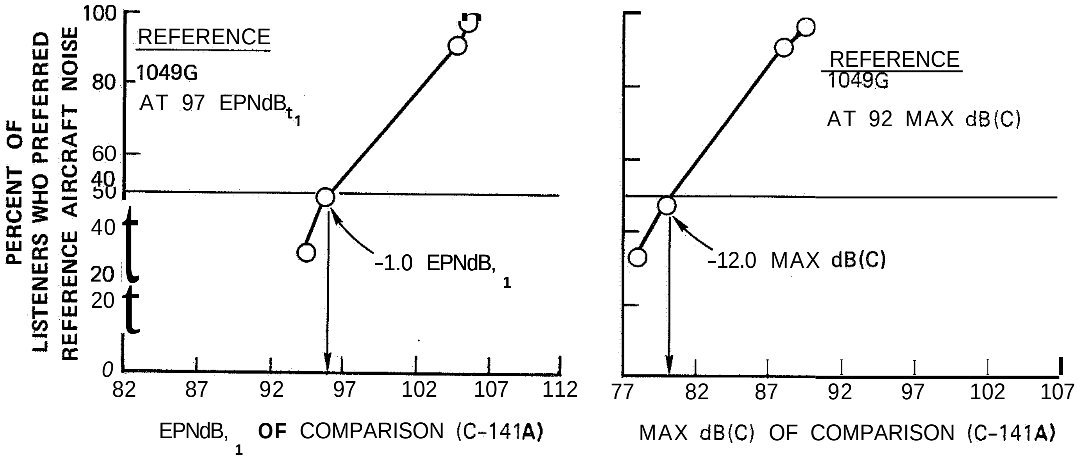

Figure 4