United States
Environmental Protection
Agency

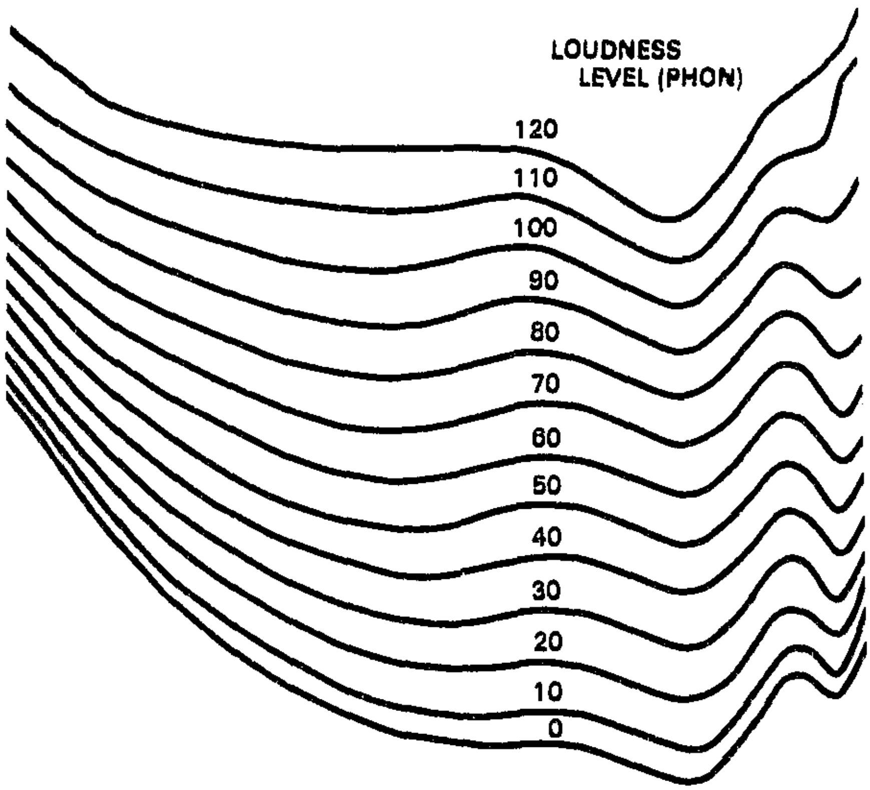
Office of
Noise Abatement and Control
Washington, D.C. 20460

EPA 550/9-79-102 N-96-01
November 1979 II-A-87

Noise

Comparison of Various Methods for Predicting the Loudness and Acceptability of Noise

Part II: Effects of Spectral Pattern and Tonal Components



N-96-01
II-A-57

COMPARISON OF VARIOUS METHODS FOR PREDICTING THE LOUDNESS AND ACCEPTABILITY OF NOISE

Part II

EFFECTS OF SPECTRAL PATTERN AND TONAL COMPONENTS

November 1979

Prepared by

B. Scharf and R. Hellman

Auditory Perception Laboratory
Northeastern University
Boston, Massachusetts 02115

Prepared for

U.S. Environmental Protection Agency
Office of Noise Abatement and Control
Washington, D.C. 20460

PREFACE

The United States Environmental Protection Agency (EPA) was charged by Congress in the Noise Control Act of 1972, as amended by the Quiet Communities Act of 1978, to conduct or finance research to investigate "...the psychological and physiological effects of noise on humans and the effects of noise on domestic animals, wildlife, and property, and the determination of dose/response relationships suitable for use in decision making..." (Section 14(b)(1)).

Pursuant to and as part of this mandate, EPA has undertaken investigations to determine and quantify subjective reactions of individuals and communities to different noise environments and sources of noise. A specific series of studies has been initiated to determine the best methods for evaluating subjective magnitude and aversiveness to noise on the basis of spectral and temporal properties, and to ascertain the importance of and means for including nonacoustical factors in the evaluation of general aversion to noise. The overall purpose of this line of research is to derive a more solid basis for assessing the aversiveness of noise and the benefits of noise control. The program calls for detailed analysis and evaluation of available data from both the laboratory and the field to assess the relative validity and predictiveness of various subjective acoustic ratings (spectral weightings and calculation schemes), as well as to acquire new data where appropriate.

Findings have been published previously in EPA Report No. 550/9-77-101 entitled "Comparison of Various Methods for Predicting the Loudness and Acceptability of Noise." That report dealt with the ability of commonly employed frequency weightings and calculation schemes to predict and quantify subjective aspects of sound. The results of the study showed the calculation schemes to be superior in predictive capability to the frequency weightings.

The D- and E-frequency weightings were significantly better than the B- and C-weightings. The A-weighting was only slightly more variable than the D- and E-weightings. All frequency weightings were level dependent with the predictive capability worse at higher levels. Analysis of the results with regard to the type of noise and the presence of tonal components was not conclusive due to a limited amount of available data.

The purpose of the investigation described in this report was to undertake a more detailed, rigorous, and systematic analysis of the previously compiled psychoacoustic data in order to (a) account for certain apparent anomalies in the data analyzed earlier as part of this program, (b) examine the sensitivity of various frequency weightings and rating schemes to spectral differences of the sound stimuli used in the investigations, and (c) evaluate subjective response to discrete frequency components superimposed over a background. The results provide partial but needed information on the relative ability of computational procedures and frequency weightings to assess subjective loudness and acceptability of sounds with different spectral shapes, the necessity of tonal corrections at low and high levels of noise, an indication as to the magnitude of a correction, and the overall effectiveness of commonly used tonal correction procedures.

EPA believes that further evaluation of data on the subjective effects of noise will foster the development of techniques to demonstrate additional benefits of noise control beyond that exhibited by currently used procedures. Fulfillment of this objective awaits further study within this series. The results published in this report, however, do provide an important step toward a more complete understanding of the phenomena of human subjective response to noise.

OFFICE OF THE SCIENTIFIC ASSISTANT
TO THE DEPUTY ASSISTANT ADMINISTRATOR
OFFICE OF NOISE ABATEMENT AND CONTROL
U.S. ENVIRONMENTAL PROTECTION AGENCY

Abstract

The present report is a continuation of an earlier report by Scharf, Hellman, and Bauer (1977). The objectives are (1) to determine whether subjective judgments of particular types of noise, categorized by spectral shape, are better approximated by some descriptors (frequency weightings and calculation procedures) than by others, and (2) to investigate the role of tonal components in these studies and to assess the adequacy of several tone-correction procedures. The analysis of data by spectral shape produced a mixed outcome. Results showed that no overall advantage would accrue from regrouping sets of data across studies on the basis of similar spectral shapes. However, although variability was not reduced when considered across nine spectral categories, the interaction between spectral shape and descriptor was highly significant ($p < .001$). The examination of over 500 spectra with and without tonal components provided only tentative support for the trends noted in the literature. When the judged attribute is either loudness or noisiness, tonal components do not seem to add to the subjective magnitude of broad-band noise below 80 dB sound pressure level. At higher levels, according to one large-scale study, tonal components seemed to add the equivalent of 2 dB to the judged noisiness. No data could be located that would permit adequate assessment of the contribution of tonal components to the "absolute" magnitude of judged annoyance or unacceptability (as distinct from noisiness or loudness). Given the small effect of tonal components in the present group of studies, the evaluation of three different tone-correction procedures (FAR 36, 1969; Kryter and Pearson's, 1965;

and Stevens's, 1970) could not lead to definitive conclusions about their relative merits. Although a small correction may be necessary for the presence of tonal components at high levels, the tone-correction procedures now available cannot be properly evaluated until more appropriate data that demonstrate the need for a tone correction are obtained.

ACKNOWLEDGEMENTS

We thank Professor Merv Lynch for his aid in statistical analyses, Barbara Kane for implementing the ANOVA programs on the computer, and Harvey Branscomb for writing programs to handle the various tone-correction procedures. We also wish to express our thanks to Jeffrey Goldstein, project director at EPA, for his constructive reviews of initial drafts of this report. A number of undergraduate and graduate students at Northeastern University also helped us with the many details of this report; they include Eleanor Arpino, Angela Ashton, Maureen Hogan, Tom Horton, and Patricia Moran. Correspondence and most of the typing were beautifully handled by Ana Silfer.

TABLE OF CONTENTS

I. INTRODUCTION

A recent report by Scharf, Hellman, and Bauer (1977) examined data from 23 studies in which subjects had judged the subjective magnitude of a large variety of noises. The aim of the Scharf, et. al. (1977) investigation was to determine how well various frequency weightings (presently incorporated or proposed for use on sound level meters) and calculation procedures assess the subjective magnitude of noise. One important conclusion, based on a total of over 600 spectra, was that the calculation procedures predicted subjective magnitude with less variability* and with greater validity** than did the frequency weightings.

Among the six frequency weightings studied, the B- and C-weightings were the poorest predictors of subjective magnitude while the D1-, D2-, and E-weightings were the best predictive weighting functions. It was also noted that the A-weighting was less than 0.5 dB more variable than the D1-, D2-, and E-weightings. Among the five calculation procedures studied, Stevens's Mark VI (1961), Mark VII (1972), and Zwicker's (1958) loudness calculation procedures were the least variable, but Perceived Noise Level (Kryter 1959) was almost as reliable. Tone-corrected Perceived Noise Level (following the FAR 36 procedure, 1969) was a somewhat poorer predictor. Mark VI and Perceived Noise Level yielded the calculated values that were closest, on the average, to the observed or judged values, although all of the frequency weightings and computational procedures examined were about equally variable in this respect.

*The index of variability was the standard deviation of the calculated levels of a group of sounds judged subjectively equal or the standard deviation of differences between calculated and judged levels. These typically ranged from 2 to 4 dB.

**The calculation procedures yielded an absolute calculated level closer to the observed level.

The objectives of the present investigation are (1) to determine whether subjective judgments of particular types of noise, categorized by spectral shape, are better approximated by some descriptors (frequency weightings and calculation procedures) than by others, and (2) to investigate the role of tonal components in these studies and to examine the relevancy of existing tone-correction procedures.

Each of these aims is addressed separately with overall results and conclusions provided in Section IV. Appendices A, B, and C include more detailed analyses.

II. SPECTRAL SHAPE

Particular types of noise often have distinctive spectral characteristics such as the low-frequency spikes of transformer noise, the low-frequency emphasis of vehicular noise, the mid-frequency bulge of many machine noises, the high frequencies of an electric bell, and so forth. In the earlier report (Scharf, et. al., 1977, Table V), twenty of the studies examined were classified according to the specific source or type of noise. The six sources considered were aircraft, industrial, vehicular, and household, as well as artificial and miscellaneous noises. A statistical analysis of the differences among the data for these six noise sources was performed in the present study. For purposes of this analysis, the vehicular noise category, for which there was only one set of data, was combined with the aircraft noise category to form a general transportation noise group, thus yielding a total of five source types. As shown in Table I, a partially hierarchical analysis of variance (ANOVA, Winer, 1962) revealed no significant differences in the predictive ability of the ten descriptors among the five source types. However, the interaction between source type and descriptor was significant ($p < .01$). Despite the statistical significance of this interaction, the differences among the descriptors are too small to provide a basis for concluding that certain types of noises are better assessed in any meaningful, practical sense by one particular descriptor than by another. Moreover, the small number of studies contained within each source type category indicates that noise source type and study are confounded.

Table I

Summary Table for Partially Hierarchical ANOVA;
Five Source Types by Ten Descriptors (PNLC Has Been Omitted)

Source of Variance	Sum of squares	Degrees of Freedom	Mean Square	F	P
Source type	50.92	4	12.73	1.60	NS
Between groups (error term)	175.20	22	7.96		
Descriptor	63.05	9	7.00	14.21	<.001
Descriptor by source	36.51	36	1.01	2.06	<.01
Within groups (error term)	97.60	198	.49		

The lack of significant differences among source types may, in fact, be attributable to the rather gross classification scheme whereby a wide variety of spectral shapes were included within each source type. Thus, this analysis may have obscured real differences among spectral shapes. A more homogeneous classification can be achieved by regrouping spectra from different studies according to spectral type or shape. It is possible that for certain spectral shapes, particular descriptors (frequency weightings or calculation procedures) predict subjective judgments better than other descriptors. If so, descriptors could then each be applied, in practice, to those spectral shapes to which they are best suited. Accordingly, each noise spectrum within the 19 studies listed in Table II of Scharf, et. al. (1977)* was placed into one of nine spectral categories: (1) negative slope, (2) positive slope, (3) broadband and flat, (4) narrow band, (5) U-shaped, (6) inverted U-shaped, (7) low-frequency peaks or valleys, (8) mid-to-high frequency peaks or valleys, and (9) mixed peaks or valleys. Figures 1 to 9 provide examples of sound spectra from each of the nine main categories. The spectra represent noises from both artificial and natural noise sources. (Appendix A gives more detailed definitions of the spectral shapes and a more detailed breakdown within the main spectral categories.)

Table II presents the standard deviations (SDs) averaged across the nine spectral categories for (1) those sets of subjective data that did not provide judged loudness levels, (2) those that did, and (3) all spectra combined.**

*The data by Pearsons, et. al. (1968) were not included in this analysis. Wells 300 and Wells 400 are counted as one study.

**The SDs for each spectral category are provided in Tables A-2 and A-7 of Appendix A. Note that the means of the SDs were computed without regard to the number of SDs contributed by each category.

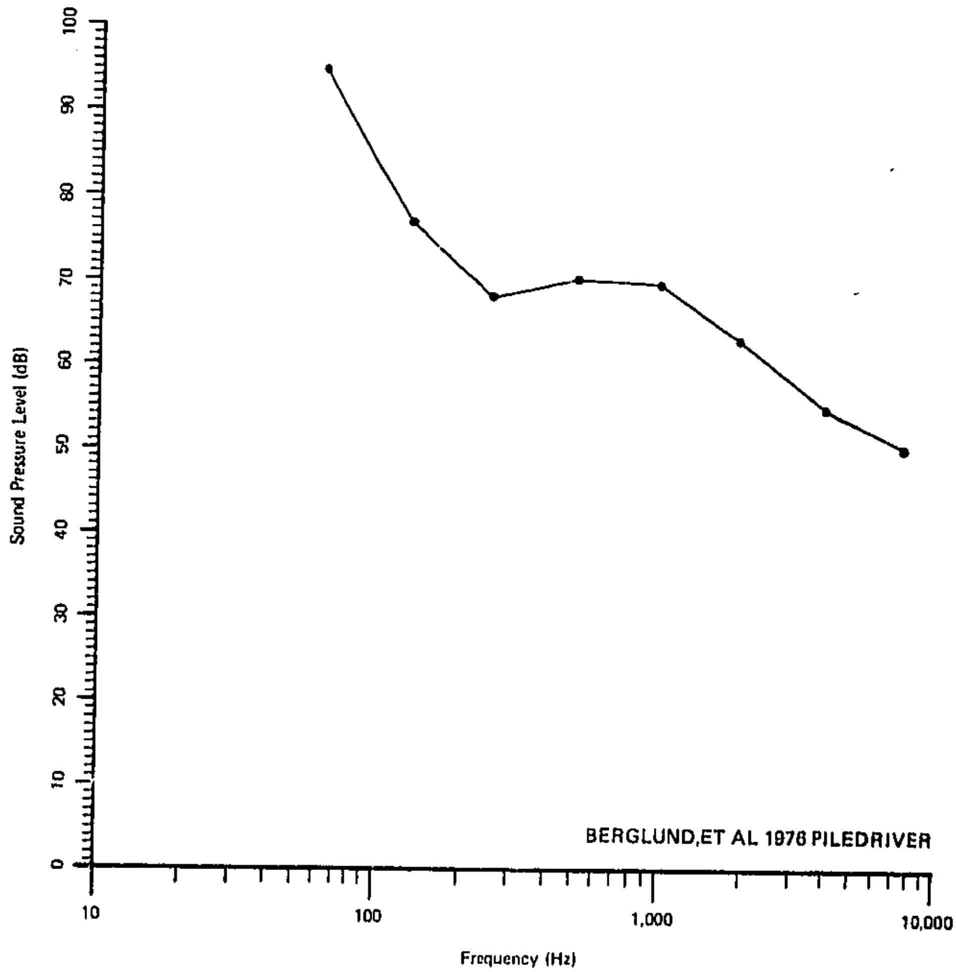


Figure 1. Spectral Category 1A (Strong Negative Slope)

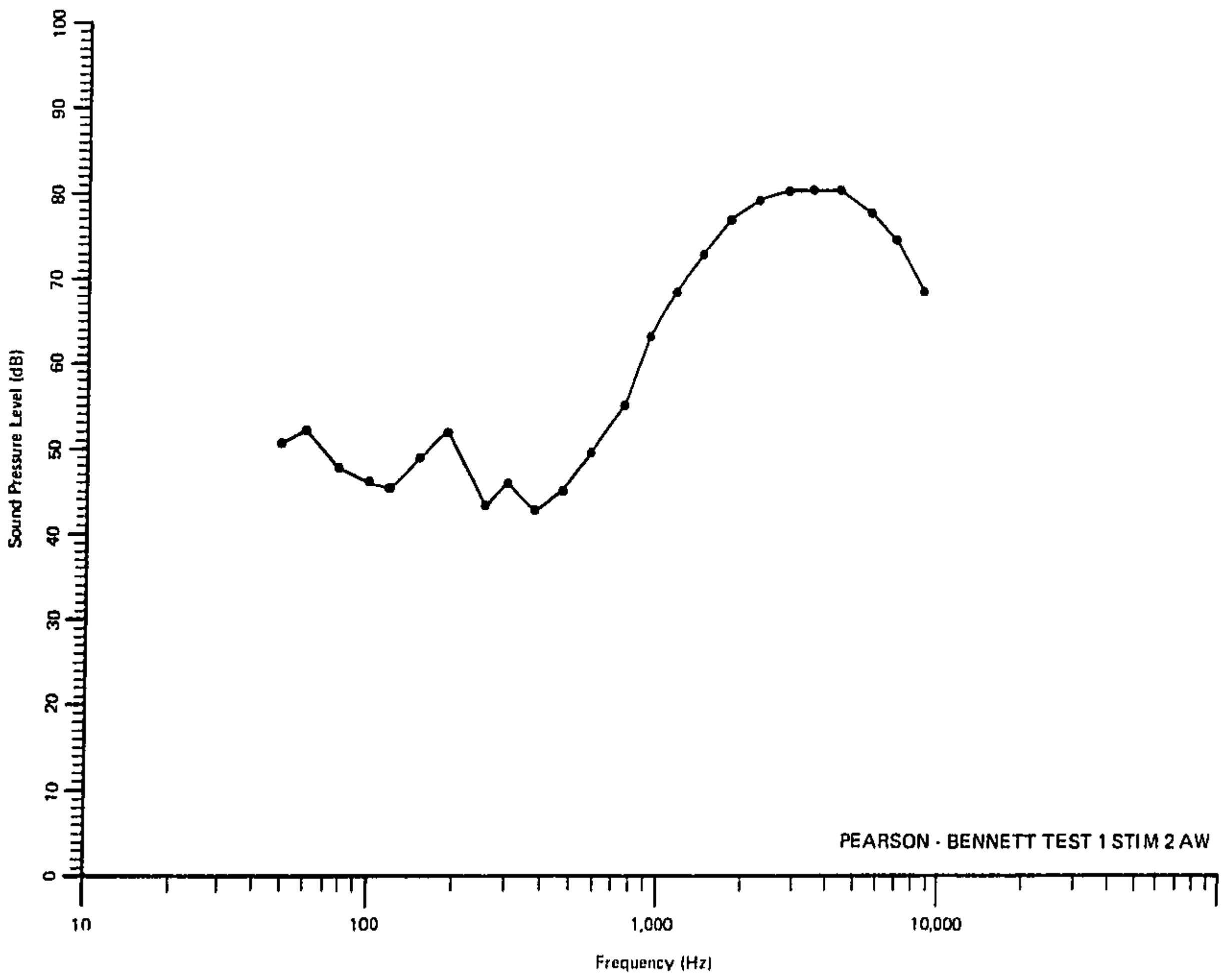


Figure 2. Spectral Category 2 (Positive Slope)

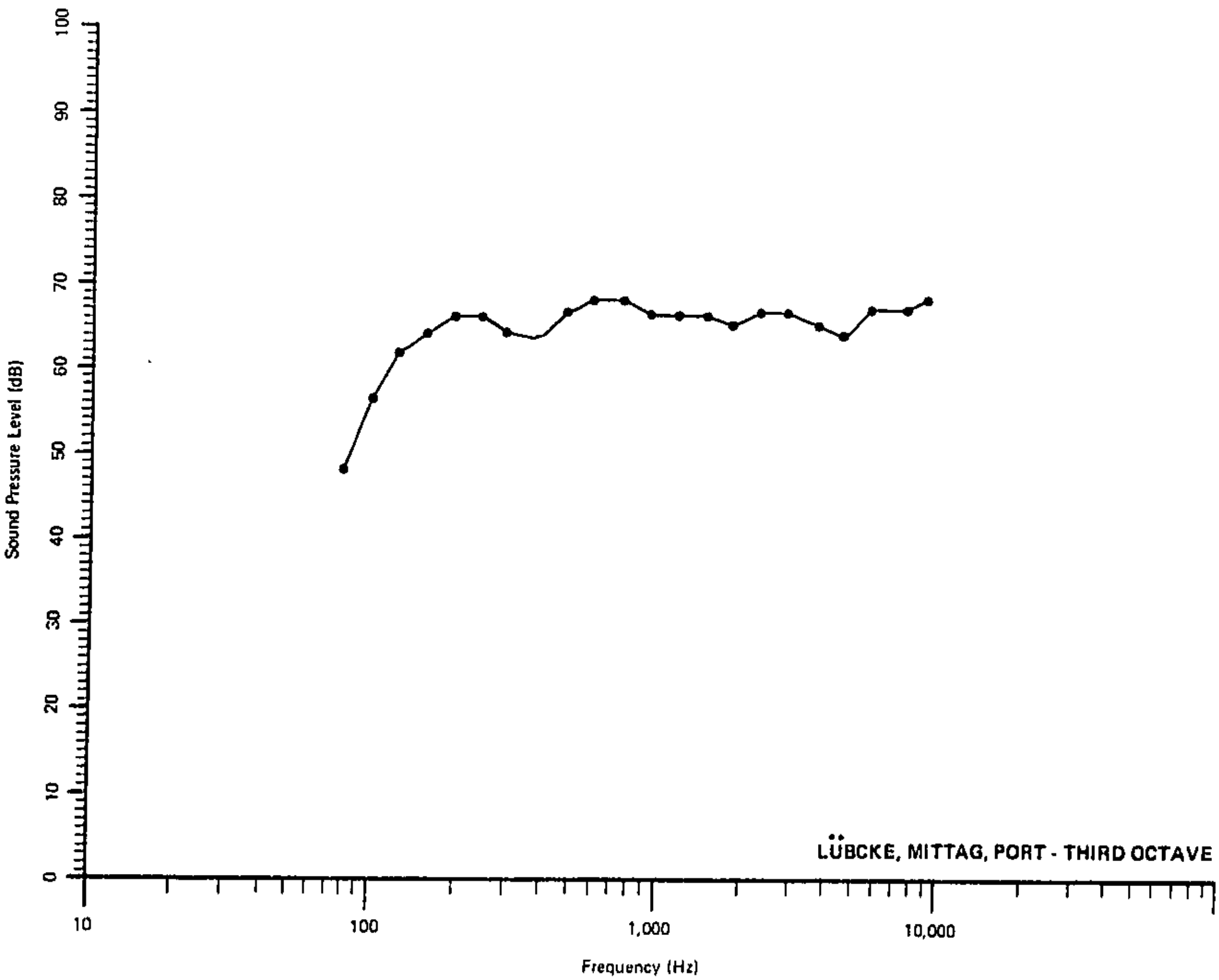


Figure 3. Spectral Category 3 (Broadband Flat)

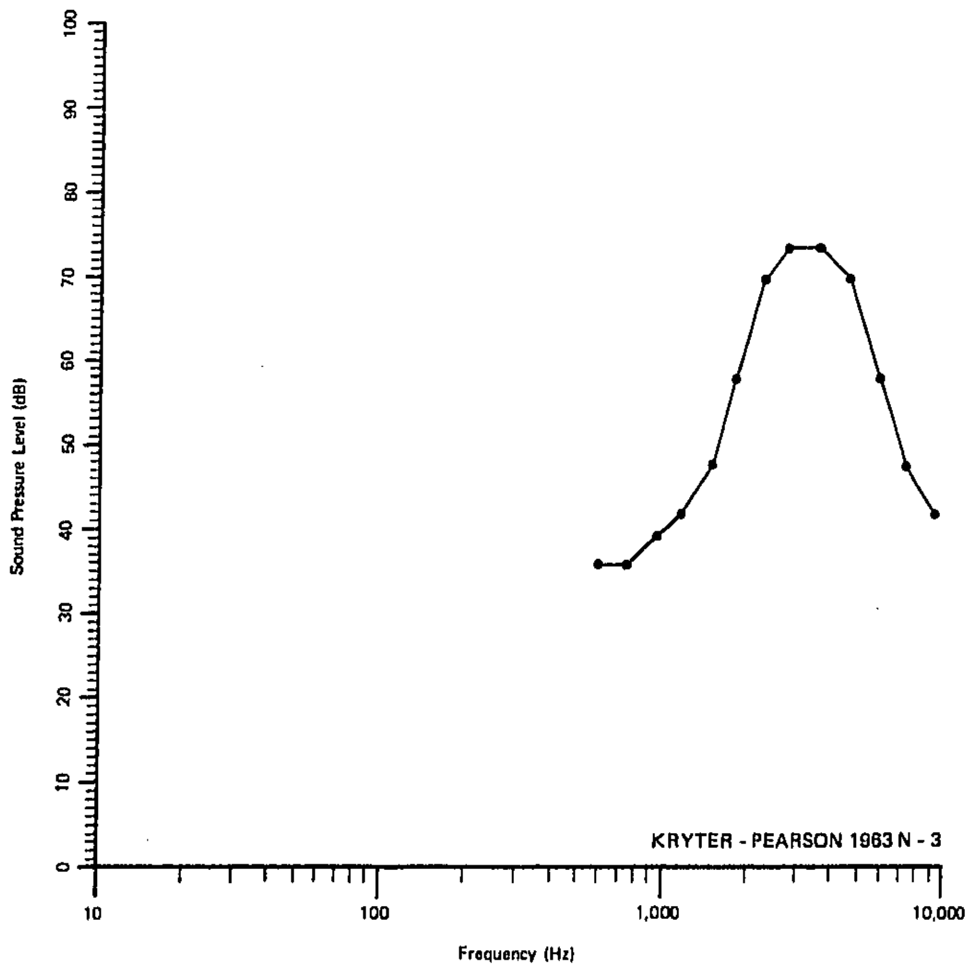


Figure 4. Spectral Category 4 (Narrow-band Noise)

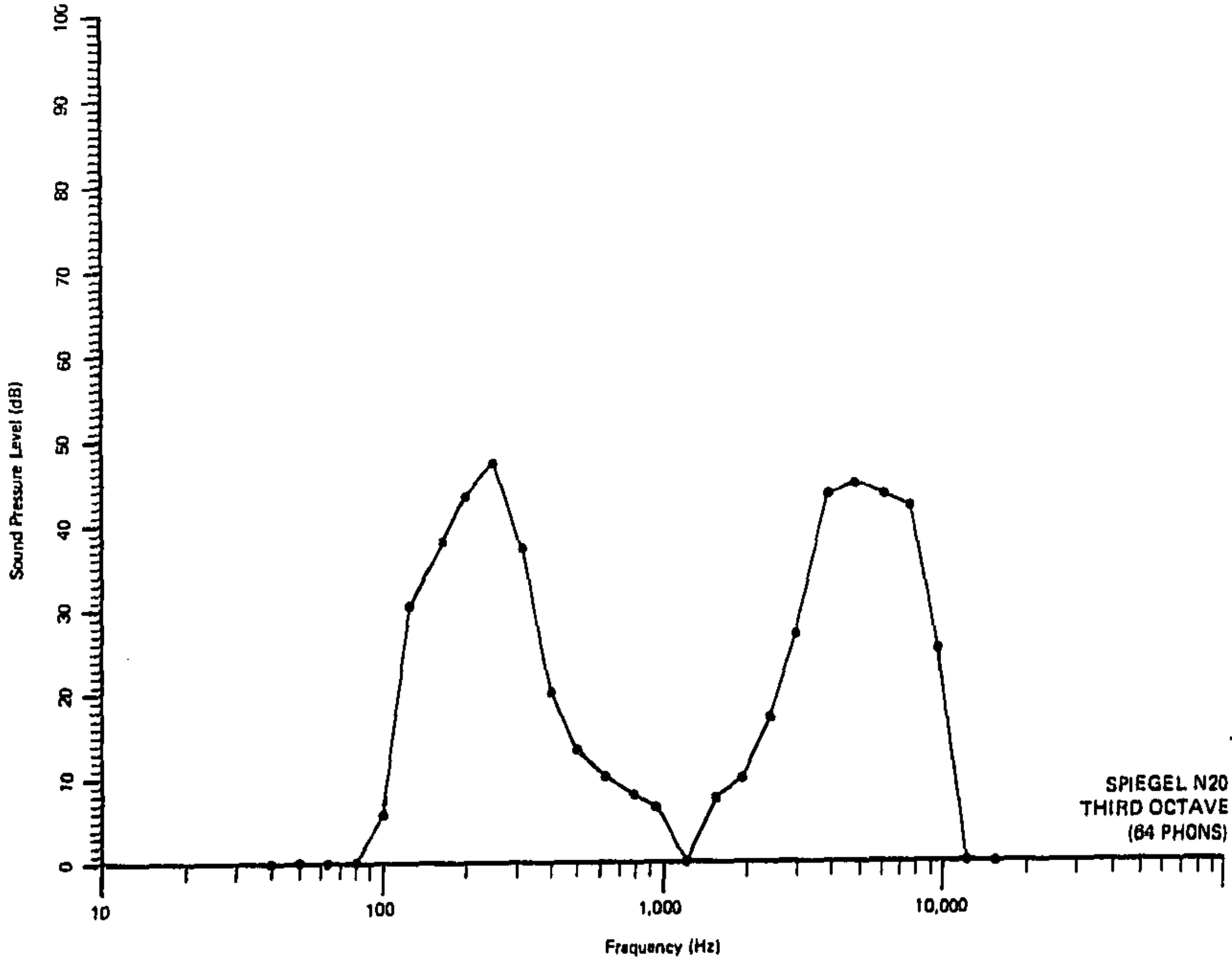


Figure 5. Spectral Category 5 (U-Shaped)

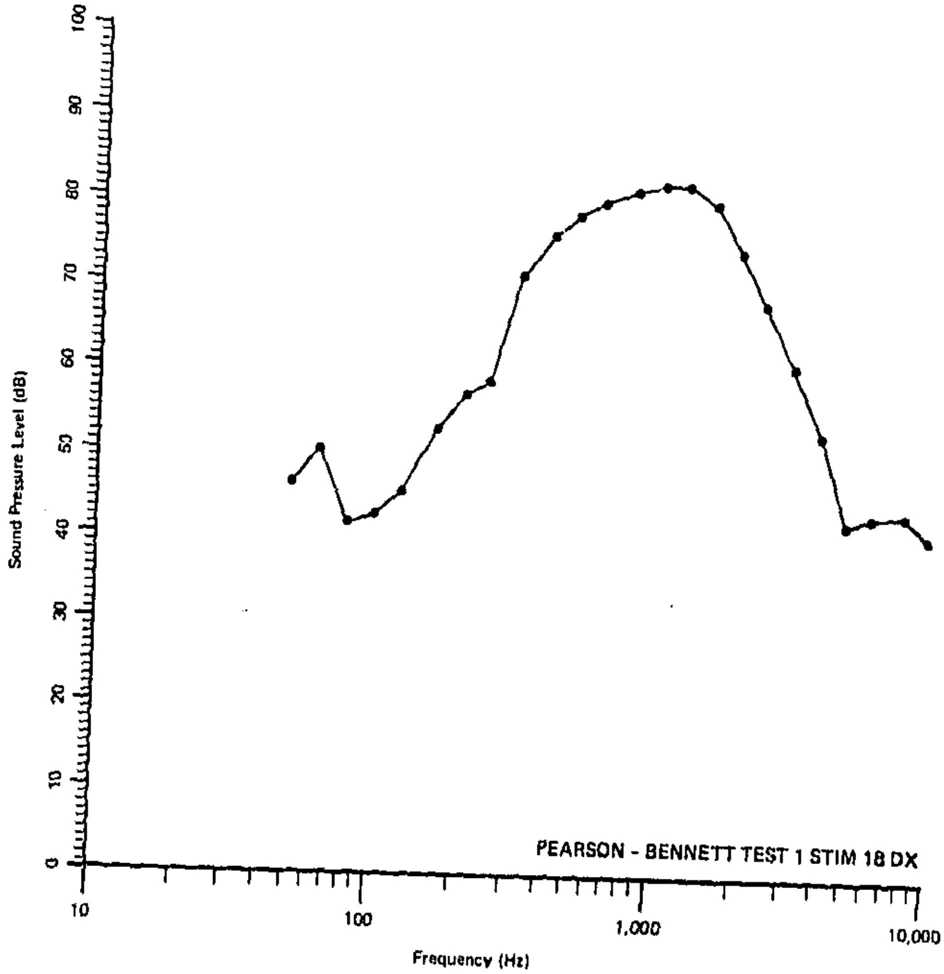


Figure 6. Spectral Category 6 (Inverted U-shaped)

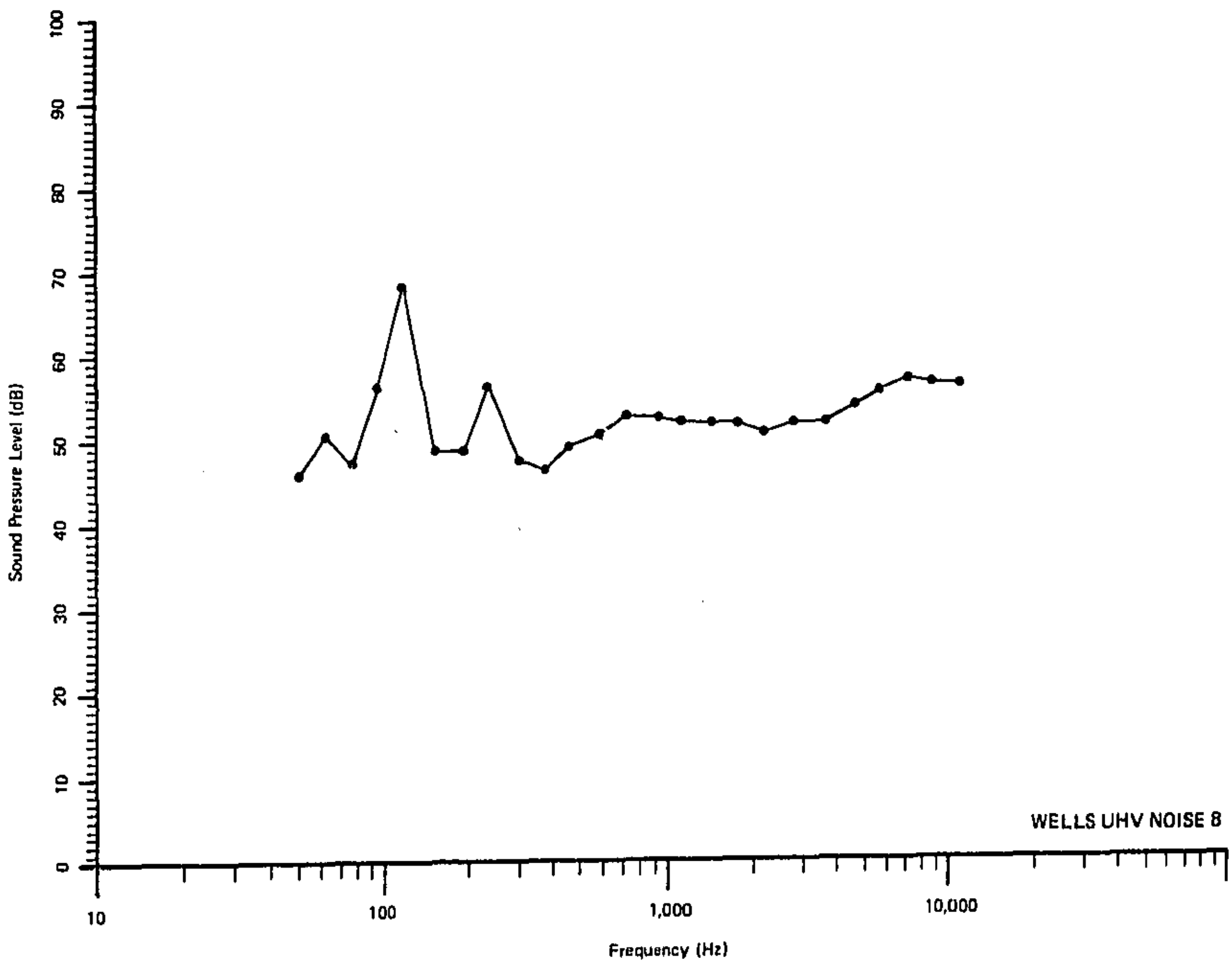


Figure 7. Spectral Category 7 (Low-Frequency Peaks and Valleys)

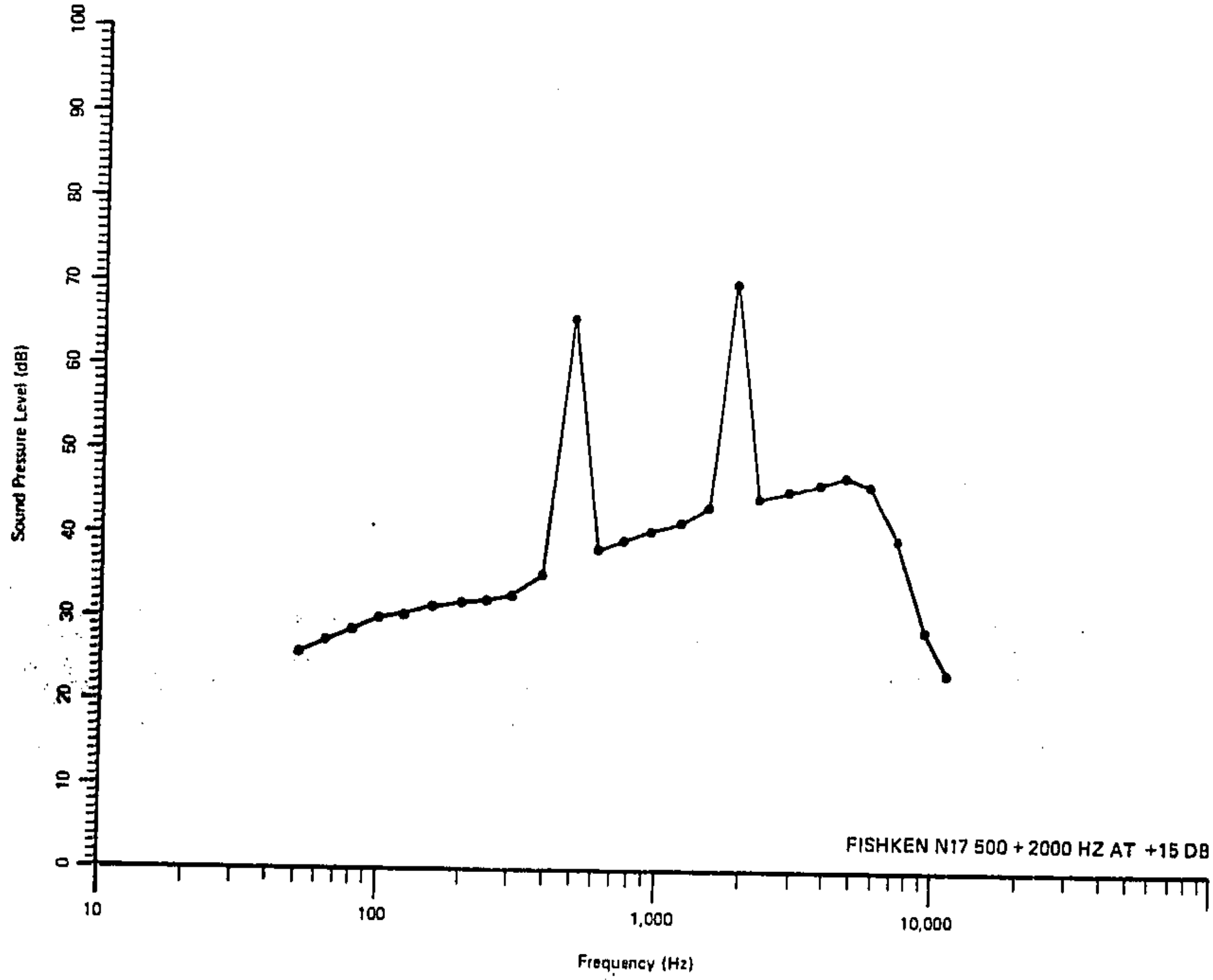


Figure 8. Spectral Category 8 (Mid-to-High Frequency Peaks or Valleys)

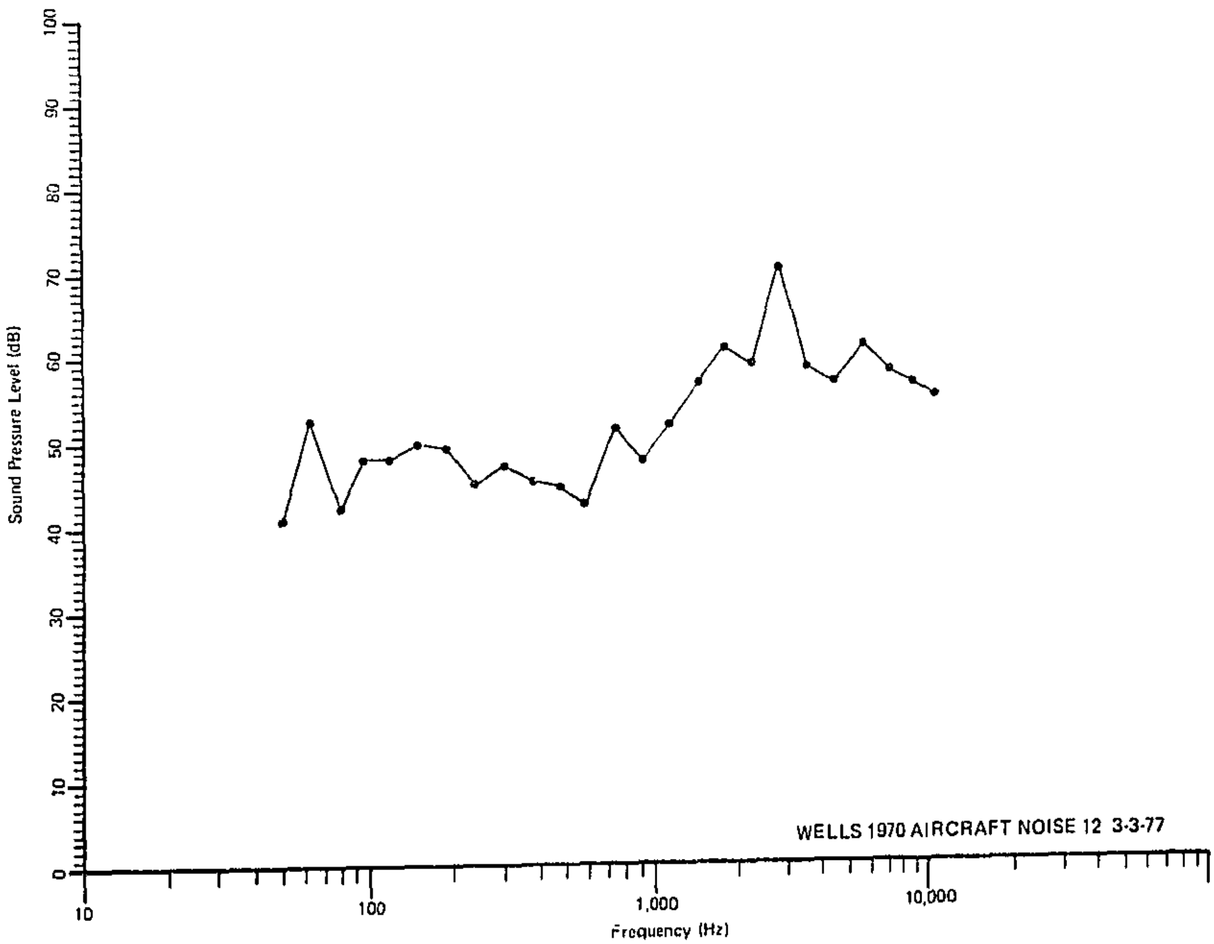


Figure 9. Spectral Category 9 (Mixed Peaks or Valleys)

Table II

Variability of Calculated Levels of Noises
Grouped by Spectral Category or by Study

(Standard deviations in decibels computed either from the calculated levels of a group of sounds judged subjectively equal or from the differences between calculated and judged levels (loudness levels). The smaller the standard deviation, the closer the scheme comes to predicting the measured subjective equality of a set of sounds.)

SOURCE	N/n	A	D1	D2	E	VI	VII	PNL	ZWI
		Descriptors							
Spectral categories (Based on calculated levels)	298/34	2.7	2.2	2.3	2.1	1.9	2.1	2.2	2.7
Spectral categories (loudness levels only)	335/56	2.9	3.0	3.0	3.0	2.4	2.4	3.1	2.3
Spectral categories (total)	633/90	2.8	2.6	2.7	2.6	2.2	2.3	2.7	2.5
Grouped by study (total) from Scharf, et. al. (1977) Table II corrected	763/28	3.1	2.7	2.7	2.6	2.3	2.2	2.6	2.4

LEGEND:

N = number of spectra

n = number of standard deviations

A = standard sound-level meter weighting

D1 = sound-level meter weighting, better known as D, adopted by International Electro-Technical Commission (1975).

D2 = sound-level meter weighting proposed by Kryter, K.D. (1970), Table 2.

E = sound-level meter weighting proposed by Stevens (1972) and circulated as ANSI Draft document S1.XX/104

Mark VI = ANSI S 3.4 (R1972) procedure for the computation of loudness of noise.

Mark VII = proposed by Stevens (1972)

PNL = perceived noise level

ZWI = based on Zwicker (1958). Computer program from Paulus and Zwicker (1972)

Table II also presents the SDs previously calculated across studies (Scharf, et. al., 1977, Table II, with the minor corrections given in Appendix D of this report).

As shown in Table II, most of the sounds represented in row 1 of Table II were judged with respect to some evaluative attribute such as noisiness, unacceptability, etc., whereas the sounds represented in row 2 were judged only with respect to loudness. Thus, the data contained in rows 1 and 2 demonstrate the same tendency noted in Scharf, et. al. (1977, Table V). Those data showed that the studies in which loudness was judged yielded larger SDs than those studies in which an evaluative attribute other than loudness was judged. The difference, however, was not statistically significant. The most probable basis for the difference, described in detail in Appendix B, is the wider range of levels covered by the loudness studies than by those studies in which an evaluative attribute was judged.

The most revealing comparison in Table II is between overall SDs calculated across spectral categories (row 3) and those calculated across studies (row 4). Except for the A-weighting, paired SDs do not differ between rows 3 and 4 by more than 0.1 dB. Thus, classifying the spectra according to shape does not reduce overall variability. Underlying this analysis was the assumption that a descriptor would be less variable if applied to groups of spectra of the same shape than to groups of spectra of different shapes. Although variability is not reduced when calculated across all the spectral categories, it may be smaller for particular descriptors applied to particular spectral categories. The interaction between category and procedure is considered in Table III.

Table III

Summary Table for Partially Hierarchical ANOVA:
Nine Spectral Categories by Ten Descriptors

Source of Variance	Sum of squares	Degrees of Freedom	Mean Square	F	P
Source type	143.18	8	17.90	2.25	<.05
Between groups	563.91	71	7.94		
Descriptor	28.64	7	4.09	12.63	<.001
Descriptor by source	38.70	56	.69	2.13	<.001
Within groups	161.04	497	.22		

Table III presents ANOVA for spectral type or shape by descriptor. This analysis shows that (1) the differences among the nine spectral categories were significant at the .05 level, and (2) the differences among the ten descriptors were statistically significant at the .001 level. Further, the interaction between spectral shape and descriptor was significant at the .001 level. However, despite this significant interaction, a meaningful multiple contrasts test could not be performed due to large variations in numbers of spectra and in numbers of SDs among the nine categories.

Also relevant to the analysis by spectral categories is the question of differences between calculated and observed loudness levels. Table IV, based partly on Table A-10 in Appendix A and partly on Table IV in Scharf, et. al. (1977), gives the overall means of the mean differences for over 300 noises grouped either by spectral category or by study. The corresponding SDs of the mean and total ranges are also shown. Except for Zwicker's loudness calculation procedure and Mark VII after the required addition of an 8-dB constant, all the descriptors are more discrepant for the sounds grouped by spectral category than for the sounds grouped by study. Of more importance, however, is the variability of the mean difference. Both the range and SDs are significantly smaller ($p < .01$ by t-test) for the sounds grouped by spectral category than for those grouped by study, with the sole exception of the SD for the A-weighting. This decreased variability is especially noteworthy since studies that differed with respect to procedures, standards, and instructions were broken up and individual spectra assigned to various spectral categories. These methodological differences would be expected to increase variability. Since the opposite occurred, it is likely

Table IV

Calculated Minus Observed Loudness Levels
(Mean Differences in Decibels)

(Overall means based upon differences for 335 spectra grouped according to spectral type as per Table A-10 in Appendix A. Overall means are also shown for the same spectra when grouped by study as per Table IV of Scharf, et. al., 1977.)

By Spectral Category	A	D1	D2	E	VI	VII	PNL	ZWI
Mean of Mean Differences	-12.1	-5.3	-5.8	-6.8	-1.2	-8.6	-1.4	3.1
S.D. of Means	4.8	4.0	4.3	4.1	3.2	3.2	3.1	3.0
Range	16.0	12.6	13.7	12.6	11.6	11.1	10.8	8.8
By Study								
Mean of Mean Differences	-10.8	-4.5	-5.0	-6.2	-0.1	-6.9	-0.0	5.1
S.D. of Means	4.5	4.7	4.8	4.6	4.5	4.3	4.7	4.2
Range	17.8	18.8	19.3	17.7	17.4	15.3	19.3	14.7

that grouping by spectral shape meaningfully enhances the validity of the descriptors. Moreover, the four calculation procedures with their much greater flexibility showed a larger drop in variability than did the weighting functions. In contrast, the A-weighting with its strong deemphasis of low frequencies revealed an increase in the standard deviation. As can be seen in Table A-10, the A-weighting grossly underestimated the level of sounds with much energy in the low frequencies and less grossly underestimated spectra with little energy in the low frequencies. To a lesser extent, the other frequency weightings also deemphasize low frequencies, and this deemphasis becomes detrimental at high levels (Scharf, et. al., 1977, Figures 6 to 8).

Furthermore, it should be pointed out that categories 7, 8 and 9 which are distinguished by the presence of low-frequency spectral peaks or valleys, mid-to-high-frequency peaks or valleys, and mixed peaks or valleys, respectively, include many sounds with tonal components. Defined as projecting at least 3 dB above their neighboring third-octave bands (see Section III), tonal components were identified in over 80 percent of the sounds in categories 7 and 8, and in 30 percent of those in category 9. The SDs are presented in Tables A-12 and A-13 of Appendix A. No clearcut differences were found between those spectra with tones and those without (except for part of category 9, as discussed in Appendix A). The general problem of tonal components is treated next, in Section III.

III. TONAL COMPONENTS

A number of studies have reported that people react more negatively to noises containing tonal components than to the same or similar noises without tonal components. Tonal components appear to add more to the unpleasantness of a noise than the same amount of acoustical energy would add if spread over a wide band of frequencies. Reports in the literature (Copeland, 1960; Hargest and Pinker, 1967; Kryter and Pearsons, 1965; Little, 1961; Little and Mabry, 1969; Pearsons, 1968; Pearsons and Bennett, 1969, 1971; Pearsons, Bishop and Horonjeff, 1969; Pearsons and Wells, 1968, 1969; Wells, 1967, 1969b) show that tonal components add the equivalent of from 2 to 15 dB or more to the annoyance of a sound than would be expected from the increase in overall energy. Several reports show that loudness or noisiness, as distinct from annoyance or objectionability, is not affected by the presence of tonal components (Fishken, 1971; Kryter and Pearsons, 1963; Rule, 1964; Rule and Little, 1963). One report (Niese, 1965) showed that tonal components affected both loudness and annoyance to the same degree. Another report (Goulet and Northwood, 1972) found no effect of tonal components on either loudness or annoyance. In both these studies stimuli were presented at levels between 45 and 75 dB sound pressure level. On the other hand, the investigations showing that tonal components do contribute unduly to annoyance were conducted mostly at levels of 85 dB and higher.

The present report evaluates a number of the studies cited above. Some studied sounds with tonal components artificially added (Fishken, 1971; Pearsons and Wells, 1969; Wells, 1969b), and others studied natural sounds that contained tonal components (Pearsons and Bennett, 1969, 1971; Wells, 1970, 1972). Although many studies not cited above, but examined in Scharf, et. al. (1977), did include some noises with tonal components, such noises did not usually constitute a large part of a given study. Nevertheless, to provide a preliminary analysis of the effect of pure tones on judgments of loudness and annoyance, 27 of the 28 sets of SDs from Scharf, et. al. (1977) were divided into two groups.* One group of 12 SDs was obtained from subjective judgments of spectra without tonal components, and a second group of 15 SDs was obtained from subjective judgments produced by spectra that contained tonal components. The presence of tonal components was based on the respective authors' definitions. The results are found in Scharf, et. al. (1977), Table V.

A partially hierarchical ANOVA (Lynch and Huntsberger, 1976) based on those data revealed no significant difference between the SDs for 10 of the 11 descriptors (PNL tone corrected in accordance with FAR 36 was omitted). The interaction between the presence or absence of tonal components and descriptors was also not significant. This negative finding, however, may not be meaningful. First, many of the studies included within the group without tonal components had a few spectra with components. Second, other differences (such as attribute judged) among studies could have obscured any effects of tonal components on the variability of the descriptors. Third, and most important, is that if the effect of tonal components is to increase the unpleasantness of a sound, then sounds all or most of which contained tonal components would all be more or less equally affected. Most of the descriptors would then show no change in their variability unless "absolute" levels were measured. Such levels were not measured in most of

*The data by Robinson and Bowsher (1961) were not included in this analysis.

the studies involving sounds with tonal components; sounds were usually all judged equal to a standard, and hence only a measure variability was meaningful.

For the present report, a detailed analysis of more than 600 spectra* from Scharf, et. al., (1977) was undertaken to identify those spectra that contained tonal components. The criterion for identification of a tonal component was that a third-octave band must have a level at least 4.75 dB above that of either of the immediately adjacent third-octave bands. This criterion was adopted to assure that the tone is at least 3 dB above the noise in the band of interest, and is similar to the FAR 36 procedure.** If the 4.75 dB criterion is exceeded, then the tone in the given third-octave band must be at least 3 dB above the level of the noise in the band that contains it. It was felt that, rather than rely on the authors' definition of tones which may vary among authors, a precise identification of the spectra containing tonal components would permit a finer determination of how well the different sound descriptors handle such stimuli. (A partial analysis of this type is presented in Appendix B for individual studies.)

Several procedures specifically designed to "correct" for tonal components will be evaluated in addition to the eight descriptors examined in Section II. These include the FAR 36 (1969) procedure, which was identified as PNLC in Scharf, et. al. (1977); a different correction to Perceived Noise Level proposed by Kryter and Pearsons (1965); and a procedure tentatively proposed by S. S. Stevens (1970) explicitly for use with Mark VII but applicable to any of the other descriptors. To augment the power of these analyses, a large-scale study by Ollerhead (1971, 1973) has been added to the

*The data by Pearsons, et. al. (1968) were not included in this analysis.

**No distinction is made between a "true" tonal component and a sharp increase in level over a restricted range of frequencies.

original group of studies examined in Scharf, et. al. (1977). Not only do these additional 104 spectra include many stimuli with tonal components, but a judged level for each of the stimuli is provided as well.

1. Composition of Studies with Respect to Tonal Components

More than 500 spectra with and without tonal components including 104 spectra from Ollerhead (1971, 1973) underlie the analysis described in this section. Of approximately 300 spectra with tonal components, over one fourth contained more than one tone. Most single components fell between 500 and 2000 Hz; the remainder were nearly evenly divided between those at frequencies below 500 Hz and those above 2000 Hz. With respect to tone-to-noise ratio, over half the components were less than 13 dB above the surrounding third-octave bands, one third were between 14 and 23 dB, and less than one tenth were more than 23 dB above the noise. Approximately half the tonal components were at a sound pressure level between 60 and 80 dB, 30 percent were above 80 dB, and 20 percent at 60 dB or lower.

2. Evidence Demonstrating a Need for a Tone Correction

As noted above, tonal components may contribute unduly to the unpleasantness of noise. If so, then those groups of noises that are a mixture of sounds both with and without tonal components ought to show more variability for a given descriptor than either a group of noises all with tonal components or a group of noises all without. Accordingly, the whole set of noises was first examined for this posited difference in variability without regard to the attribute judged (whether loudness or some evaluative attribute), tone-to-noise ratio, or overall level, parameters which may in fact be relevant to the effect of tonal components on human response.

Table V presents the standard deviations for 542 spectra from 13 studies and subsets listed in column 1, Table VI, that had at least three spectra with tonal components and at least three without. The mean SDs for all the sets of spectra, both with and without tonal components, are given in the first row, followed by the mean SDs for those spectra with tonal components, and then by those without tonal components. The SD of the SDs upon which the mean values are based are also shown. For every descriptor the SD for the overall group is larger than the SD for either subgroup. This result suggests that sounds with tonal components are judged somewhat differently from sounds without; that effect is apparent for this analysis even when studies that contained relatively soft sounds judged with respect to loudness are included.

However, when just those studies are examined that involved evaluative judgments of annoyance, unacceptability, etc. (and studies that involved loudness judgments are excluded), the picture is altered. Table VII shows that in the annoyance studies, those spectra with tonal components produced the largest SDs under all eight descriptors, while those spectra without tonal components produced the smallest SDs. The presence of tonal components made the descriptors more variable without apparently affecting the SDs obtained for the mixture of sounds both with and without tonal components. Had spectra with tonal components been judged differently, on the average, than spectra without tonal components, the SDs for the overall group would have been increased, not decreased slightly as they are in Table VII.

Table V

Standard Deviations (In Decibels) for Spectra Both with and without Tonal Components, for Spectra with Tonal Components, and for Spectra without Tonal Components. (Means were Unweighted. Attribute Judged: Loudness, Annoyance, Noisiness, Etc.)

	Number of Spectra	Number* of SDs	Frequency Weighting A	D1	D2	E	Calculation Procedure VI	VII	PNL	ZWI
Mean SD (in decibels)										
Spectra Both with and without components	542	29	3.1	3.0	3.1	2.9	2.6	2.7	2.8	2.7
Spectra with tonal components	314	29	2.6	2.4	2.4	2.3	2.1	2.1	2.4	2.3
Spectra without tonal components	205	20	2.7	2.4	2.6	2.3	2.1	2.1	2.2	2.4
SD of SDs (in decibels)										
Spectra Both with and without tonal components			1.2	1.2	1.2	1.2	1.1	1.4	1.0	1.2
Spectra with tonal components			1.4	1.2	1.3	1.2	1.0	1.1	1.2	1.4
Spectra without tonal components			1.6	1.3	1.4	1.4	1.4	1.5	1.3	1.5

*The number of SDs varies because some studies do not contain at least 3 spectra required for the computation of a standard deviation.

Table VI

Studies that Contributed Spectra to the Analysis in Table V

Studies that Contributed to Both the 542 and 314 Spectra

Study	Year
Borsky	1974
Fishken	1971
Jahn	1965/66
Lübcke, *et. al.*	1964
Ollerhead	1971, 1973
Pearsons and Bennett	1969
Pearsons and Wells	1969
Spiegel	1960
Wells	1970
Wells 300-400 Series	1969a
Wells (unpublished)	c. 1970
Wells (UHV)	1972
Yaniv	1976

Studies that Contributed to 205 Spectra

Study	Year
Jahn	1965/66
Lübcke, *et. al.*	1964
Ollerhead	1971, 1973
Pearsons and Bennett	1969
Pearsons and Wells	1969
Spiegel	1960
Wells	1970
Wells 300-400 Series	1969a
Wells (Unpublished)	c. 1970
Yaniv	1976

Table VII

Standard Deviations (In Decibels) from Studies Involving Mainly Judgments of Annoyance or Unacceptability, for Spectra Both with and without Tonal Components, for Spectra with Tonal Components, and for Spectra without Tonal Components. (Means were Unweighted.)

	Number of Spectra	Number of SDs	Frequency Weighting A	D1	D2	E	Calculation Procedure VI	VII	PNL	ZWI
Mean SD (in decibels)										
Spectra with and without components	260	13	2.5	2.0	2.1	1.9	1.9	1.9	2.1	2.8
Spectra with tonal components	150	12	2.8	2.2	2.2	2.1	2.1	2.1	2.4	2.9
Spectra without tonal components	106	11	1.9	1.6	1.8	1.4	1.2	1.3	1.4	2.3
SD of SDs (in decibels)										
Spectra with and without tonal components			1.2	0.8	0.9	0.8	0.8	0.9	0.9	1.5
Spectra with tonal components			1.5	1.1	1.3	1.1	1.0	1.1	1.2	1.6
Spectra without tonal components			1.2	0.8	0.9	0.9	0.7	0.7	0.7	1.8

Studies that Contributed to Both the 260 and 150 Spectra

Study	Year
Borsky	1974
Pearsons and Bennett	1969
Pearsons and Wells	1969
Wells	1970
Wells 300-400	1969a
Wells (Unpublished)	1970

Studies that Contributed to 106 Spectra

Study	Year
Pearsons and Bennett	1969
Pearsons and Wells	1969
Wells 300-400	1969a
Wells (Unpublished)	1970

Perhaps combining results from diverse studies that used widely different methods and instructions obscures the possible effect of tonal components on judged annoyance. Moreover, any interpretation of these findings must be limited due to the absence of measurements of "absolute" judged levels of annoyance.

The relevance of the attribute judged is further shown by breaking Table V's 314 spectra with tonal components into two groups, those for studies in which annoyance and noisiness were judged, and those in which loudness was judged. Table VIII shows that five of the eight descriptors are more variable for the annoyance and noisiness judgments than for the loudness judgments; the other three are about the same for both attributes. However, the mean SD for annoyance across the eight descriptors is 2.5 dB compared to 2.2 dB for loudness. Such a small difference, 0.3 dB, is not meaningful.

Earlier studies suggested that tonal components would be a significant factor at high sound pressure levels -- in annoyance judgments -- but not at moderate or low levels. If so, a group of sounds with tonal components judged with respect to annoyance should yield more variable descriptors when a mixture of both low and high level sounds are included than when only low or only high levels are included. Of the 233 spectra with tonal components in Table VIII that were judged for annoyance and noisiness, 121 were at or above an overall sound pressure level of 80 dB. Table IX shows that the SDs for the 233 spectra are, on the average, larger by 0.1 dB than the SDs for the 121 high level sounds. This difference is too small to be meaningful.

Table VIII

Mean Standard Deviations (In Decibels) for Spectra with Tonal Components Based on Annoyance, Noisiness, and Loudness Judgments

Attribute Judged	Number of Spectra	No. of Studies/ No. of SDs	Frequency Weighting A	D1	D2	E	Calculation Procedure VI	VII	PNL	ZWI
Annoyance and Noisiness	233	8/17	2.9	2.4	2.4	2.3	2.2	2.3	2.5	2.7
Loudness	81	5/12	2.2	2.5	2.5	2.3	2.0	1.9	2.3	1.6

Table IX

Standard Deviations in Decibels for 233 Spectra with Tonal Components at Moderate and High Sound Pressure Levels Compared to Standard Deviations in Decibels for 121 Spectra at or above an Overall Sound Pressure Level of 80 dB.

Number of Spectra	No. of Studies/ No. of SDs	Frequency Weighting A	D1	D2	E	Calculation Procedure VI	VII	PNL	ZWI
233	8/17	2.9	2.4	2.4	2.3	2.2	2.3	2.5	2.7
121	4/11	3.2	2.3	2.5	2.3	2.1	2.1	2.2	2.1

Closely allied to overall level of the tone is the tone-to-noise ratio. Only Mark VII and Perceived Noise Level were compared for two ranges of tone-to-noise ratios. Over the range of 3 to 13 dB (relative to the third-octave band level), the mean SD was around 1.6 dB; over the range of 14 to 23 dB, the mean SD increased to around 2.7 dB. Thus, based on the data examined in this report, both Mark VII and Perceived Noise Level, and presumably the other descriptors, may be less accurate in assessing human response to sound when the tone projects out well above the noise i.e., none of the descriptors may adequately assess the subjective annoyance produced by relatively strong tones.

The effect of the frequency of the tonal components could not be adequately evaluated since in the annoyance studies most of the tones were between 500 and 2000 Hz. For 19 spectra with tonal components below 500 Hz, the mean SD was 0.9 dB for Mark VII and 1.5 dB for Perceived Noise Level. For 22 spectra with tonal components above 2000 Hz, the SDs increased to 2.9 dB for Mark VII and to 2.4 dB for Perceived Noise Level. Given the small sample sizes, this finding is highly tentative although it is consistent with the analysis of anomalous studies in Appendix B.

The role of the number of tonal components was also ascertained. Several of the Wells (1969a, 1970, 1972) studies and the Ollerhead (1971, 1973) study contained sounds with multiple tones as well as with single tones. The SDs for Mark VII and Perceived Noise Level were not unusually high for the group of spectra with both single and multiple tones. In the Ollerhead study, as seen in Table X, the SDs produced by the mixture of single and multiple tones is only slightly larger (0.3 dB) than the SDs produced by spectra with multiple tones only. These preliminary findings suggest that the number of components may not affect the variability of the descriptors.

Table X

Analysis of Standard Deviations in Decibels for Mark VII and Perceived Noise Level Produced by Data from Ollerhead (1971) Based on Spectra that Contained Both Single and Multiple Tones and Spectra with Multiple Tones Only.

	Number of Spectra	Mark VII	PNL
Mean SD (in decibels)			
Spectra with Single and Multiple Tones	60	3.2	3.1
Spectra with Multiple Tones Only	33	2.9	2.8

The preceding analysis of the effect of tonal components on the variability of the descriptors obviously does not lead to clear-cut conclusions. However, the main effect of tonal components on human response appears, from earlier studies, to be an increase in the aversiveness of broadband sounds. Thus, it is essential to examine the mean differences between calculated and observed levels. Only Ollerhead (1971, 1973) provided observed levels based on judgments of an evaluative attribute -- noisiness; the remaining observed levels are based on loudness. Table XI shows the mean differences from a common group of studies listed in Table XII that had some sounds with tonal components and some sounds without. The differences are just about the same for the two sets of spectra; adding tonal components appears to have little effect on the discrepancy between calculated and observed levels. Since none of the eight descriptors makes special provision for tonal components (except, as an integral part of the Zwicker procedure), the lack of any effect of tonal components on the mean differences suggests that adding tones does not increase the subjective magnitude. Moreover, the variability of the mean difference is greater for spectra without tones than for spectra with tones. Taken together, the overall results in Table XI imply that a tone correction procedure may not be needed when the judged attribute is loudness.

The effect of tonal components is different, however, for those sounds that were judged with respect to noisiness in the Ollerhead (1971, 1973) study. Those mean differences, listed separately in Table XI, are more positive for the 44 spectra without tonal components than for the 60 spectra with tonal components. This suggests that the observed levels were higher for the spectra with tones than for those without. The increase in the mean difference is 1.8 dB, averaged

Table XI

Mean Differences in Decibels (Calculated Minus Observed Levels) for Studies Containing Some Sounds with Tonal Components and Some without. (Attribute Judged was Loudness Except in the Ollerhead (1971, 1973) Study which is also Listed Separately.)

	Number of Spectra	Frequency Weighting A	D1	D2	E	Calculation Procedure VI	VII	PNL	ZWI
Mean of Mean Differences									
Spectra without tonal components	99	-7.9	-2.0	-2.6	-3.5	3.0	-4.0	4.0	7.5
Spectra with tonal components	141	-7.7	-1.0	-1.5	-2.9	2.9	-4.6	4.1	7.1
SD of SDs									
Spectra without tonal components		5.6	5.7	5.4	6.1	5.8	6.2	6.6	5.3
Spectra with tonal components		4.7	4.9	5.0	4.6	4.0	4.0	4.8	3.3
Ollerhead only (Noisiness judged)									
Mean Differences									
Spectra without tonal components	44	-3.1	2.7	1.9	1.5	7.2	1.0	9.9	11.9
Spectra with tonal components	60	-5.3	1.0	-0.1	-0.1	5.8	-1.3	7.9	10.2

Table XII

Studies that Contributed Spectra to The Analysis in Table XI

Studies that Contributed to 141 Spectra		Studies that Contributed to 99 Spectra	
Study	Year	Study	Year
Fishken	1971	John	1965/66
John	1965/66	Lübcke. *et. al.*	1964
Lübcke, *et. al.*	1964	Ollerhead	1971
Ollerhead	1971	Spiegel	1960
Spiegel	1960	Yaniv	1976
Yaniv	1976		

over the eight descriptors. The most likely interpretation of this result is that in the Ollerhead (1971, 1973) study, aircraft sounds with tonal components were judged the equivalent of 1.8 dB noisier than sounds without tonal components. Further, it should be noted here that, in contrast to the other studies listed in Table XII, the noise stimuli in the Ollerhead (1971, 1973) study had an overall sound pressure level greater than 80 dB.

In general, the studies examined in this report provide little evidence for the need for a tone correction. This finding only appears to contradict conclusions drawn from some studies cited above. However, the reasons for the apparent disagreement may be found in the specific nature of the studies examined in the present report. (See Section IV below.) Furthermore, Ollerhead's (1971, 1973) data on the aversiveness of sounds with tonal components at high levels do suggest a need for a tone correction, but only of the order of 2 dB. Despite this generally negative result, the following section examines and evaluates several tone-correction procedures.

3. Descriptions of Tone-Correction Procedures

a) PNLC or FAR 36 Tone Corrections

A tone correction is contained within the FAR 36 (1969) aircraft certification regulation. The tone correction was included to increase, in accordance with subjective judgments, the measured Perceived Noise Level of aircraft that produced noise spectra with tonal components. The Perceived Noise Level is calculated in the usual way for a given spectrum (Kryter, 1959). The FAR 36 procedure then smoothes the spectrum and compares the original spectrum to the smoothed spectrum in each third-octave band. If a band level of the original spectrum exceeds the corresponding band level of the smoothed spectrum by 3 dB or more, then a correction in decibels is added to the calculated Perceived Noise Level to account for the presence of discrete tones. Thus, within the FAR 36 procedure, the criterion for a tonal component is that it exceed the noise level in the third-octave band containing it by 3 dB or more. The number of decibels added to the calculated Perceived Noise Level depends on the frequency of the tone and its level relative to the smoothed third-octave band noise level. Tones between 500 and 5000 Hz are penalized twice as much (in decibels) as tones below and above that frequency range. The correction cannot exceed 6.67 dB, which is the penalty for a tone 20 dB or more above the noise level. Between tone-to-noise ratios of 3 dB and 20 dB, the penalty increases linearly with level, more rapidly in the middle frequency range than elsewhere. If more than a single tonal component is identified, only the largest penalty is added to Perceived Noise Level; in essence, multiple tonal components are ignored and a correction is applied only to the strongest tone (taking into account frequency and tone-to-noise ratio). This procedure does not take absolute level into account, presumably because it was designed explicitly for high-level aircraft noise. Figure 10 illustrates how the FAR 36 procedure depends on tone-to-noise ratio and on the frequency of the tone.

b) Kryter and Pearsons's (1965) Tone-Correction Procedure

Like the FAR 36 method, the procedure proposed by Kryter and Pearsons (1965) is designed for use with Perceived Noise Level. It is henceforth referred to in this report as PNLKP. Instead of first calculating Perceived Noise Level and then adding a correction in decibels as in the FAR 36 method, PNLKP first corrects the levels of each third-octave band containing identified pure tones, and then calculates Perceived Noise Level according to

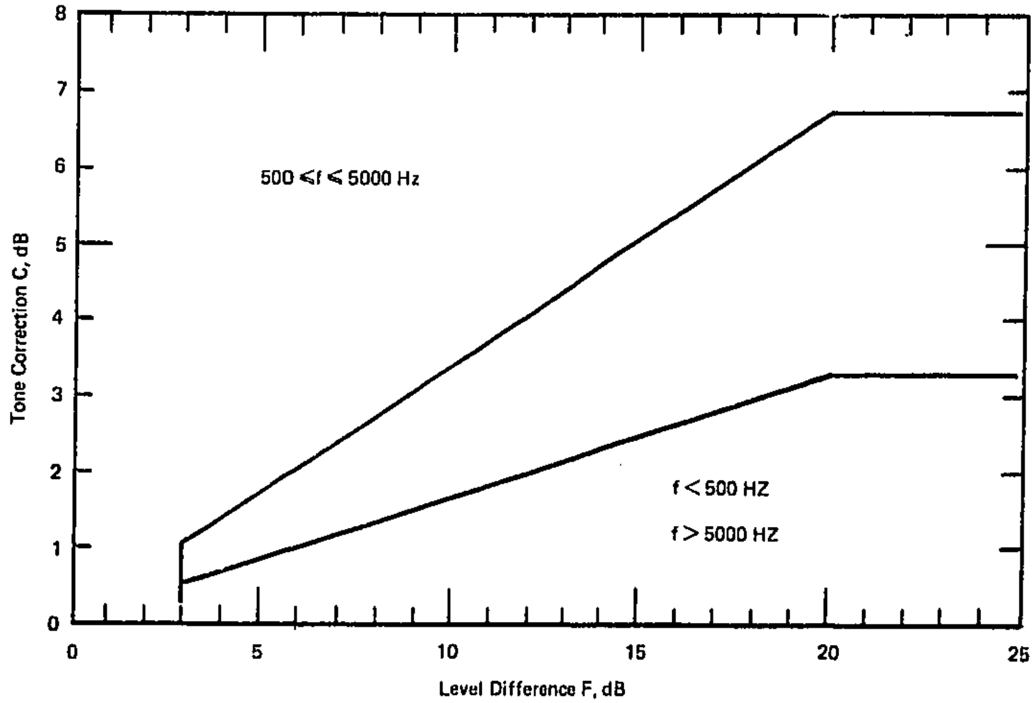


Figure 10. FAR 36 Procedure Tone-Corrections

Kryter (1959) on the basis of the revised spectrum. The result is a tone-corrected Perceived Noise Level. In the current application a correction is made for each band identified as containing a pure tone at third-octave band center frequencies. Only tones 3 dB or more above adjacent third-octave bands have been identified as pure tones in this report although Kryter and Pearsons (1965) suggested a correction for even smaller tone-to-noise ratios. Figure 11 shows that the value of the correction within each band increases with increasing tone-to-noise ratio up to a maximum ratio of 25 dB. The value also varies continuously with frequency with a flat maximum between 3000 and 4000 Hz, depending on tone-to-noise ratio.

c) Stevens's (1970) Preliminary Tone-Correction Procedure

In 1970, S. S. Stevens circulated a tentative proposal for a tone-correction method to be used with his Mark VII or Mark VI computational procedures. His correction was based on the notion that the underestimation of the calculated perceived magnitude of a tone-and-noise complex according to Mark VI or VII arises because the auditory system analyzes components in the complex as distinct sounds and then, in effect, adds them together to obtain a total percept. To develop a procedure that would mimic the auditory system, Stevens turned to data on the masking of a pure tone by broad-band noise. He assumed that the loudness of the partially masked tone would summate with the loudness of the noise when the two are judged as a composite sound. Stevens's procedure takes into account the fact that partial masking depends on the tone-to-noise ratio as well as on the absolute level of the noise.

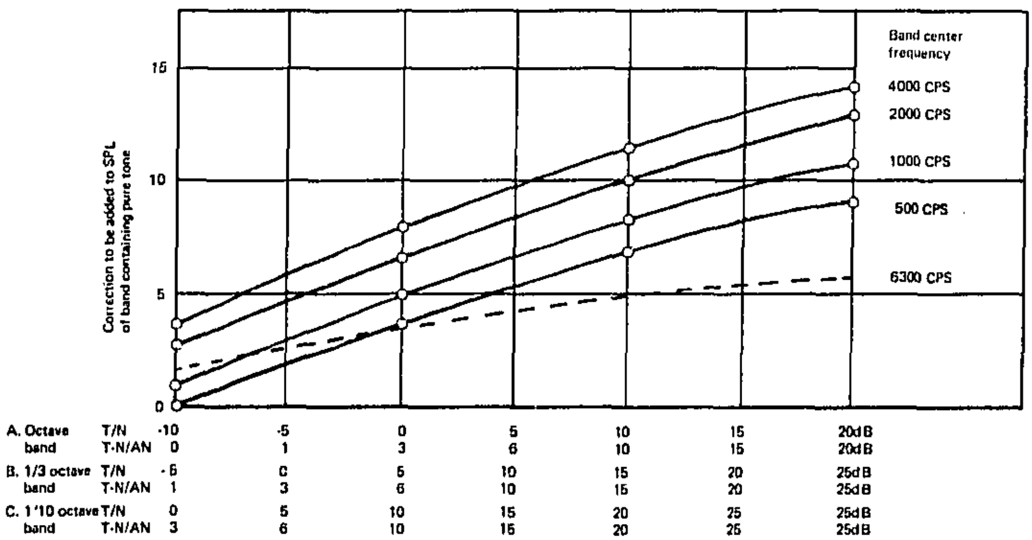


Figure 11. Decibel Correction to be Added to Sound Pressure Level of a Band Containing Pure-Tone Component Prior to Calculation of Perceived-Noise Level. Parameter is Band Center Frequency. Abscissa is Either Ratio, in decibels, Between Tone and Noise Measured Separately within a Band (T/N) or the Ratio Between Level of Band with Tone and Noise Together and Level of Adjacent Bands (T+N/AN) when Measured with Full-, 1/3-, or 1/10-Oct-Band Filters (from Kryter and Pearsons, 1965).

Stevens did not state what criterion to use for identifying the presence of a tonal component. Thus, the same 3-dB criterion described above for PNLC was applied to Stevens' correction procedure. Once identified, the tonal component was removed from the spectrum by averaging the levels of the immediately surrounding third-octave bands. The Perceived Level is then calculated by means of Mark VII for the toneless noise spectrum. The decibel value of the tonal component is read from curves, as shown in Figure 12. Although Mark VII was used in constructing the curves that provided the value of the tonal correction, Stevens' correction can be applied to any one of the descriptors dealt with in the present report. Once the tonal component has been removed, the particular frequency weighting or calculation procedure is used to compute the predicted level of the toneless spectrum. Stevens' tone-correction value in decibels is then added to that computed level.

The Stevens correction procedure differs from the FAR 36 and Kryter and Pearsons (1965) procedures in two main respects. First, it includes the level of the band containing the tone as an important determinant of the value of the tone correction, and second, it omits any dependence of the correction on the frequency of the tonal component. The Stevens procedure also differs in that it is derived from basic psychoacoustic considerations about the interaction between tone and noise and in that it includes an explicit method for handling multiple tonal components.

The correction for multiple tones assumes that the tones may partly mask or inhibit one another, the more so the closer they are in frequency.*

*Unless the tones are in the same critical band, in which case they are treated like a single tonal component. Since a critical band is about as wide as a third-octave band, it would require an analysis finer than the usual third-octave band analysis to identify such closely spaced tones.

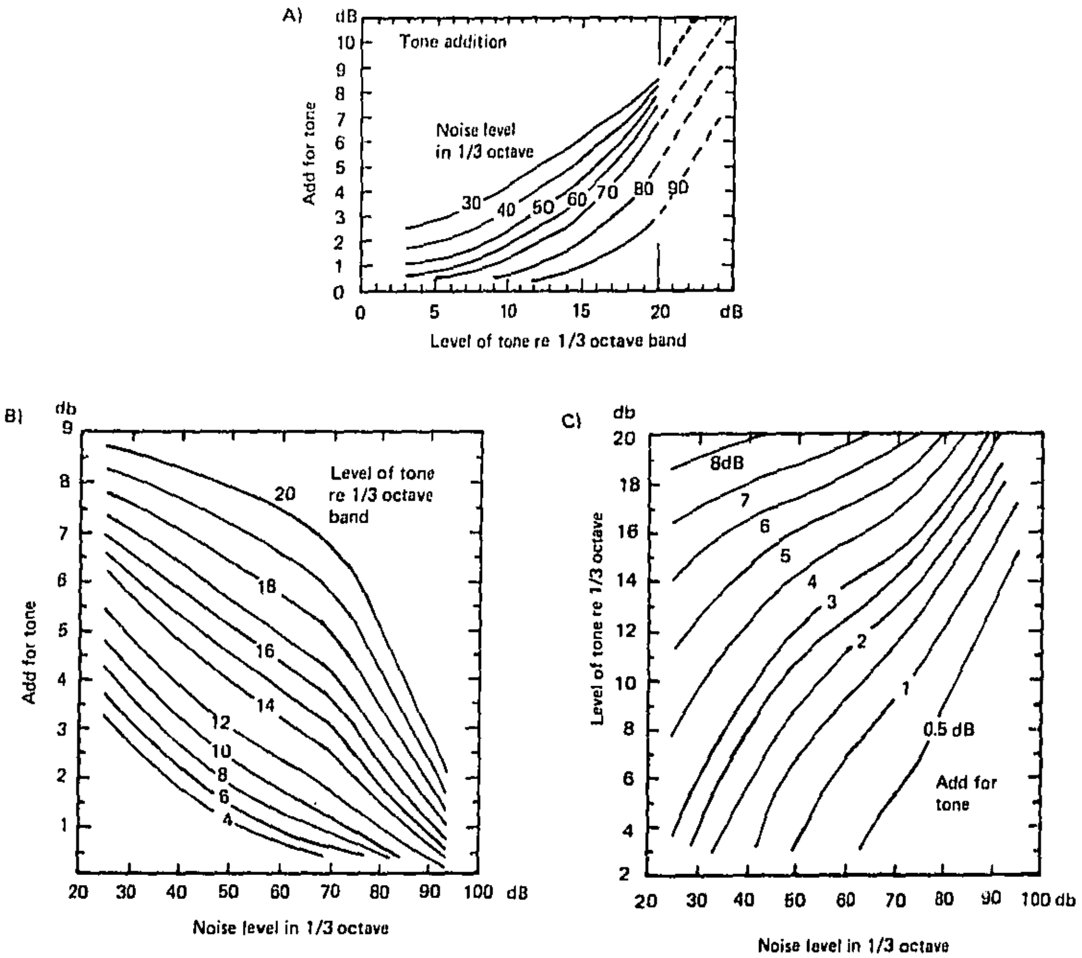


Figure 12. (A) Curves for tone addition showing the number of decibels (ordinate) added by the tone to the perceived level of the broadband noise calculated without tone. The abscissa shows the number of decibels by which the tone projects above the 1/3 octave level of th noise. The 1/3 octave level is found by averaging the band levels above and below the band that contains the tone. For tone projections greater tha 20 dB the tone addition grows linearly with a slope of 1.0 (dashed lines). (B) Same as (A), but with the parameter being the level of the tone as it projects above the level of the noise in the 1/3 octave band. (C) Another alternative to (A). Here the parameter is the number of decibels to be added to the calculated level of the noise for the tone projections given by the ordinate (From Stevens, unpublished, 1970.)

Mark VII is used in conjunction with the mel scale of pitch to compute the amount of inhibition from lower to higher frequencies. Within each band, the calculated inhibition, expressed in sones, is subtracted from the perceived magnitude of a given tonal component, as determined by Mark VII. This value is then converted to Perceived Level in decibels which is used in calculating the correction to be applied for that component. Finally, the corrections computed for each component, after inhibition is taken into account, are all added to the Perceived Level of the toneless noise. As with single tones, this procedure can be applied to any of the descriptors examined in the present report. Accordingly, the Stevens procedure will be applied to eight descriptors, with special attention to Mark VII for which it was primarily intended.

It must be emphasized that Stevens did not publish this tone-correction procedure, developed in 1969 and 1970, and in all likelihood intended to modify it before publication. Therefore, it is to be considered a tentative model that may yield insights into just how and when to apply a tone correction.

4. Other Tone-Correction Procedures

A tone-correction procedure not evaluated in this report was proposed by Wells (1969b) for use with his general annoyance-level (ANL) procedure for assessing negative effects of noise. This report does not deal with his procedure primarily because it has not been as widely discussed in the literature as other descriptors have.

Zwicker's (1958) procedure may be considered a tone-correction procedure in that it is designed to handle pure tones and combinations of tones and noise, with respect to loudness. Only in so far as noisiness or aversiveness differs from loudness, would his procedure require a tone correction. Tables V and VII do not contradict such a possibility; Table VIII lends some support since, applied to spectra with tonal components, Zwicker's procedure was the most variable (procedure) when annoyance or noisiness was judged, and was the least variable when loudness was judged. Zwicker's procedure handles tones on the same basic principles of mutual inhibition that inspired Stevens's correction procedure.

5. Evaluation of Tone-Correction Procedures

Similar to the analysis by spectral shape in Section II, evaluation of the relative effectiveness of the three tone-correction procedures described above consists of two parts. First, the effect of the procedures on the variability of predicting subjective magnitude is assessed. Next, their effect on differences between calculated and judged levels is examined.

a) Variability

Table XIII shows the SDs for 260 spectra, some with and some without tonal components, from six studies and subsets listed in column 1, Table VII, in which listeners judged an evaluative attribute (e.g., annoyance, unacceptability). According to Table XIII, the mean SD for Mark VII corrected by Stevens's preliminary tone-correction procedure is larger than for Mark VII uncorrected. The outcome appears the same when the Stevens correction is applied to the other seven descriptors, as shown in Appendix C. Similarly, the FAR 36 (PNLC) and Kryter and Pearsons (1965) tone-correction procedures

Table XIII. Effect on Standard Deviations of Three Tone-Correction Procedures. (SDs are given for Mark VII with and without the preliminary tone-correction procedure of S.S. Stevens. SDs are also given for PNL with and without the FAR 36 correction, listed under PNLC, and the proposed correction by Kryter and Pearsons, listed under PNLKP. Tonal components were present in all the spectra or only in some.)

				Calculation Procedures				
				Mark VII				
Attribute Judged	Number of Spectra	Tonal Components		Uncorrected	Corrected	PNL	PNLC	PNLKP
Evaluative only	260	Present in some	Mean SD (dB)	1.9	2.2	2.1	2.2	3.2
			SD of SDs (dB)	0.9	1.1	0.9	0.9	1.3
Evaluative and Loudness	314	Present in all	Mean SD (dB)	2.1	3.0	2.4	2.6	3.5
			SD of SDs (dB)	1.1	1.4	1.2	1.2	1.8

inflate the SDs. Thus it appears that the tone corrections do not improve the descriptors' predictability of the negative reactions to noises that contain tonal components. If the correction procedures had worked, differences between noises with tonal components and those without should be reduced, and the SD of a mix of both kinds of noise should become smaller after correcting for the presence of tones. The failure of the three correction procedures to decrease the SDs may be due to the inclusion of many noises below 80 dB (although none below 70 dB) where tonal components may be subjectively less important than at high levels. Such a level effect would be especially detrimental for Stevens's correction procedure which adds larger corrections at low than at high noise levels. Nevertheless, in a separate analysis, Ollerhead's (1971, 1973) 104 aircraft noises judged for noisiness were almost all above 90 dB sound pressure level, and yet variability for those noises also increased from about 3.9 dB to about 4.4 dB when either Stevens's correction was applied to Mark VII, or the FAR 36 (PNLC) procedure was applied to Perceived Noise Level.

Table XIII also shows that the correction procedures increase the variability for 314 spectra from 13 studies and subsets listed in column 1, Table VI, all of which contained tonal components. In some studies, an evaluative attribute was judged, in others loudness was judged. Mixing the two types of judgments together may be part of the reason for the increase in variability when a correction procedure is introduced. Further, the tone-correction procedures did not decrease the variability when applied only to tone-noise complexes (a) with multiple tones, (b) with tones at high tone-to-noise ratios (14 to 23 dB), (c) with tones at lower ratios (3 to 13 dB), (d) at or above an overall sound pressure level of 80 dB, or (e) with tones below 500 Hz, or above 2000 Hz. These results are indicated in Table XIV.

b) Mean Differences Between Calculated and Observed Levels

If the tone-correction procedures do not reduce variability of the descriptors, do they at least reduce the discrepancy between calculated and judged levels? Table XV shows the mean differences between calculated and observed levels for 141 spectra with identified tonal components, from six studies and subsets listed in column 1, Table XII. Observed levels were calculated according to Mark VII, with and without a tone correction, and according to the Perceived Noise Level (PNL), FAR 36 (PNLC), and Kryter and Pearsons (1965, PNLKP) procedures. If the required 8-dB constant is added to the Mark VII values to make them 3.4 and 6.7 dB, respectively, then all three tone-correction procedures increase the over-estimation of the measured level. More important, the corrections also increase the SDs of the mean differences, thus indicating that calculated values vary more around their means when corrected than when uncorrected. These 141 spectra included 81 for which loudness judgments were made and for which a tone-correction procedure may not be needed (see Table XI). A separate analysis was also made in Table XI of the other 60 spectra, all from Ollerhead (1971, 1973). Those data did suggest that a tone correction of about 2 dB may be necessary for judged noisiness.

Table XIV. Effect of Multiple Tones, Tone-to-Noise Ratio, Sound Pressure Level of Tone-Noise Complexes Above 80 dB, and Tone Frequency on Mean Standard Deviations in Decibels Produced by Three Tone-Correction Procedures.

Parameter Assessed	Number of Spectra	Calculation Procedures Mark VII Uncorrected	Mark VII Corrected	PNL	PNLC	PNLKP
Evaluation of Multiple Tones	62	2.7	2.9	2.8	3.0	3.4
Effect of Tone-to-Noise Ratio						
3-13 dB	47	1.6	2.2	1.7	1.9	1.6
14-23 dB	68	2.7	3.5	2.8	2.8	3.0
Effect of Overall Sound Pressure Level at or above 80 dB	121	2.1	2.7	2.2	2.3	2.7
Effect of Tone Frequency (Hz)						
Tones at or below 500 Hz	19	0.9	1.4	1.5	1.5	1.7
Tones at or above 2000 Hz	22	2.9	3.5	2.4	2.4	3.1

Table XV. Mean Differences (in Decibels) (Calculated Minus Observed Levels for 141 Spectra with Tonal Components. Attributes Judged: Loudness and Noisiness.)

	Mark VII		PNL	PNLC	PNLKP
	uncorrected	corrected			
Means (dB)	-4.6	-1.3	4.1	7.8	7.8
SD of Means (dB)	4.0	6.9	4.8	5.8	6.7

The results in Tables XIII, XIV, and XV do not necessarily mean that the three proposed tone-correction procedures are basically inadequate. Most of the data used for an assessment of the descriptors, particularly those used for Table XIV, are based on subjective judgments produced by spectra from either Wells (1969a, 1970, 1972) or Ollerhead (1971, 1973). As pointed out in part 2 of this section, as well as in Appendix A, the inconclusive findings with respect to the need for a tone-correction are most likely due to the dearth of relevant data. Before the tone-correction procedures can be properly assessed, a need for a correction must be clearly demonstrated.

Ollerhead's (1971, 1973) study was the only one to provide differences in level (relative to a specified standard) for a large group of noises for which an aversive quality, and not loudness, had been judged. The variability of the differences between calculated and observed levels, like the combined results in Table XV, did not decrease for Ollerhead's (1971, 1973) noises when the three tone-correction procedures were applied. The absolute discrepancies did go down for Mark VII (and also Mark VI) corrected by the Stevens-correction procedure, but by less than 1 dB, from an average overestimation of nearly 7 dB to around 6 dB. A reduction in the overestimation was unexpected since the correction procedure was designed to increase the calculated values. Stevens's procedure, however, often results in a decrease in the calculated level, especially when the tonal components are at low levels relative to a high-level noise as was the case with Ollerhead's (1971, 1973) sounds. On the other hand, PNLC and PNLKP overcorrected and increased the discrepancy between calculated and measured levels.

6. Summary of Findings Relative to Tonal Components

The examination of large numbers of spectra with and without tonal components lends only tentative support to the trends noted in the literature. When the judged attribute is either loudness or noisiness, tonal components did not seem to add to the subjective magnitude of broad-band noise for stimuli below 80 dB sound pressure level. Only when the noise was at a high level (above about 80 dB overall sound pressure level), did the introduction of tonal components appear to add to the aversiveness of the noise. Above 80 dB sound pressure level, the increase in noisiness ascribed to the presence of tonal components is about 2 dB. No data seem to be available to adequately assess the contribution of tonal components to the "absolute" magnitude of judged annoyance or unacceptability.

Procedures in use or proposed to correct for the presence of tonal components did not decrease the variability of Mark VII and Perceived Noise Level to which they were applied. The corrections also did not bring the calculated levels closer to the measured levels. Although a small correction may be necessary for the presence of tonal components at high levels, the procedures now available cannot be properly assessed until more data demonstrating the need for a tone correction become available.

IV. CONCLUSIONS AND RECOMMENDATIONS

The present report is a continuation of an earlier report by Scharf, et. al. (1977). The present survey sought (1) to discover whether particular noise descriptors (sound-level frequency weightings and various calculation procedures) are more appropriate for certain types of spectral shapes than for others, and (2) to determine just how important tonal components are in human response to noise and how best to take tonal components into account.

The analysis of data by spectral shape provided a mixed outcome. Results revealed little change in the standard deviations (SDs) of eight descriptors (frequency weightings A, D1, D2, and E, and calculation procedures Mark VI, Mark VII, Perceived Noise Level, and Zwicker's Loudness Computation) when more than 600 sounds were grouped according to spectral shape instead of according to experimental study. Thus no overall advantage would accrue from regrouping sets of data across studies on the basis of similar spectral shapes.
The relative efficacy of the eight descriptors in terms of variability was the same as in Scharf, et. al. (1977) whether the sounds were grouped by spectrum or by study. Mark VI, Mark VII, and Zwicker's procedures were the least variable and the A-weighting was the most variable (C- and B-weightings having been excluded)--but the difference between the largest SD (2.8 dB for the A-weighting) and the smallest SD (2.2 dB for Mark VI) was only 0.6 dB. However, although variability was not reduced when considered across all the nine spectral categories, it was smaller across the eight descriptors for some categories than for others. An interaction between descriptor and spectral shape was found to be statistically significant at the .001 level. Despite this significant interaction, the present data do not reveal which descriptors are more suited than which others for

specific spectral categories. More judgments of sound annoyance and noisiness are needed, particularly for categories 1 (negative slope), 5 (U-shaped), 7 (low-frequency peaks and valleys), 8 (mid-to-high frequency peaks and valleys), and 9 (mixed peaks and valleys) using a known calibrated standard before this question can be answered.

Results obtained with a known, calibrated standard would provide additional information that permits the computation of mean differences as well as the standard deviations of the mean differences. Table IV showed that regrouping data by spectral shape rather than by study resulted in a larger reduction in both the SD and range of the mean differences for the calculation systems than for the frequency-weighting functions. In fact, such a regrouping of data enlarged the variability produced by the A-weighting. These results are in line with earlier findings (Scharf, et. al. 1977) showing that the SDs produced by the frequency-weighting functions, particularly the A-weighting, strongly depend on level whereas the calculation systems are less sensitive to level effects. Taken together, the results from this report and Scharf, et. al. (1977) argue for the use of a calculation procedure such as Mark VI to achieve a significant improvement in predicting subjective magnitude from physical measurements. Further, the greater flexibility provided by the calculation procedures offers a distinct advantage should such factors as tonal components and duration need to be incorporated into these systems.

A detailed analysis of over 500 spectra with and without tonal components provided little evidence for the need for a tone correction. This outcome would appear to be at variance with previous conclusions in the literature. However, the nature of the studies evaluated was such as to reduce the likelihood of showing any effect of tonal components. Many of the studies required loudness judgments or evaluative judgments at levels below 80 dB. Even those studies such as Ollerhead's (1971, 1973), which required evaluative judgments at high levels, stressed noisiness as opposed to annoyance. Studies by Berglund, et. al. (1975, 1976) suggest that at high levels, noisiness and loudness are essentially indistinguishable, whereas annoyance may remain considerably greater than both noisiness and loudness. Subjects identify noisiness more as a characteristic of the sound and annoyance more as a description of their own general reaction to noise. The presence of tonal components at high levels may affect judgments of annoyance more than they affect either noisiness or loudness. However, no measurements seem to be available of "absolute" magnitude of annoyance caused by sound with tonal components. Thus Ollerhead's subjects would probably have given higher estimates of annoyance, had they been asked, than they did of noisiness when exposed to high-level noise containing tonal components.

Given the small effect of tonal components in the present group of studies, the evaluation of three different tone-correction procedures, FAR 36 (1969), Kryter and Pearson's (1965) and Stevens's (1970) could not lead to definitive conclusions about their relative merits. Nevertheless, none of the three improved the effectiveness of the descriptors to which they were applied; the variability and the discrepancy between calculated and judged level either remained the same or increased. This disappointing outcome should reinforce the realization that data are needed on a large enough set of sounds with and without tonal components to permit adequate evaluation of tone-correction procedures. Special attention must be paid to the instructions given to the subjects. The present report has tended to distinguish studies on the basis of a simple dichotomy, between loudness and evaluative judgments such as noisiness, unacceptability, and annoyance. This dichotomy was necessitated by the nature of the studies investigated which usually mixed together a number of adjectives when giving instructions other than loudness. The reports by Berglund, et. al. (1975, 1976) suggest that a careful distinction should be made among loudness, noisiness, and annoyance in instructions. A further important point is that most of the studies heretofore have used psychophysical procedures that emphasize the overall level. Thus, observers are asked to adjust one sound to be subjectively equal to another sound or to report when one sound, presented at various levels, is subjectively greater (or less) than a standard sound. Such a procedure is usually appropriate for investigations of loudness but may inadvertently focus the subject on loudness in a study that aims to investigate annoyance or even noisiness. Magnitude estimation has been used successfully for judgments of sound annoyance by Berglund, et. al. (1975, 1976), Bishop (1966), Galanter (1978), Hiramatsu, et. al. (1976), and Scharf and Horton (1978). By presenting sounds with tonal components at different tone-to-noise ratios, frequencies, and configurations, an experimenter can obtain a fine grain scale of the relative annoyance of various sounds. Such experiments would yield the kind of data needed to determine when tonal corrections are needed and how best to implement them.

APPENDIX A

CATEGORICAL ANALYSIS ACCORDING TO SPECTRAL TYPE

Introduction

Categorical analysis involved the identification and classification of more than 600 spectra that were evaluated in a previous study (Scharf, Hellman, and Bauer, 1977). The spectra were obtained from 23 studies that encompassed a wide variety of natural and simulated noises. In addition to the identification and classification of spectra, a statistical analysis of subjective measurements produced by the noises in each spectral category and across spectral categories was made based on four frequency weighting functions (A, D1, D2, E) and four calculation procedures (Mark VI, Mark VII, Perceived Noise Level, Zwicker).

Procedure

The spectra were subdivided into the following nine primary categories: (1) negative slope, (2) positive slope, (3) broadband flat, (4) narrow band, (5) U-shaped, (6) inverted U-shaped, (7) low-frequency peaks, (8) mid-to-high-frequency peaks, and (9) mixed peaks. In order to obtain a finer analysis of the data, category 1 (negative slope) was further divided into two parts, and category 4 (narrow band) was divided into three parts.

Each of the nine categories and subcategories is defined as follows:

(1) Negative slope - maximum energy located at low frequencies.

- (A) Strong negative slope - greater than 5 dB per octave fall-off of energy above approximately 500 Hz, but fall-off often begins between 100 and 1000 Hz.

(B) Slight negative slope - noise energy falls off from 3 to 5 dB per octave above 500 Hz, but fall-off often begins between 100 and 1000 Hz.

(2) Positive slope - maximum energy located at high frequencies. Noise energy falls off rapidly below 500 Hz, but often the fall-off begins at higher frequencies.

(3) Broadband flat - spectral distribution of energy remains about the same (±5 dB) across a band of frequencies at least two octaves wide.

(4) Narrow band - octave band or narrower.

(A) Noise band centered at frequencies below 500 Hz.

(B) Noise band centered at frequencies between 500 and 2000 Hz.

(C) Noise band centered at frequencies above 2000 Hz.

(5) U-shaped - noise energy reaches a maximum at low and at high frequencies, i.e., the noise has a mid-frequency notch.

(6) Inverted U-shaped - noise energy falls off at low and at high frequencies, i.e., the noise energy peaks over a broad range of mid-frequencies.

(7) Low-frequency peaks - peaks and valleys in spectra (±5 dB) located below 500 Hz.

(8) Mid-to-high-frequency peaks - peaks and valleys in spectra (±5 dB) located above 500 Hz.

(9) Mixed peaks - peaks and valleys in spectra (±5 dB) located at frequencies both below and above 500 Hz.

Table A-1 shows the distribution of noises according to spectral category and type of noise. It is evident that the number of spectra are very unevenly

TABLE A-1 IDENTIFICATION OF SPECTRA ACCORDING TO SPECTRAL CATEGORY AND TYPE OF NOISE

Category	----------------- Number and Type of Noise				----------------------------		
	Aircraft	Industrial	Vehicle	Household	Artificial	Miscellaneous	Total Numbe of Spectra
1A - strong, negative	22	18	10		8	6	64
1B - slight, negative		4				4	8
2 - positive					17	5	22
3 - broadband, flat		3				5	8
4A - narrow band, centered below 500 Hz					18		18
4B - narrow band, centered between 500 and 2000 Hz					22		22
4C - narrow band, centered above 2000 Hz					18		18
5 - U-shaped					6		6
6 - inverted U-shaped	15	14	14		15	10	68
7 - low-frequency peaks and valleys	5			18	25	25	73
8 - mid-to-high frequency peaks and valleys	32			6	180	4	222
9 - mixed peaks and valleys	38	16		9	14	27	104
TOTAL	112	55	24	33	323	86	633

divided according to categories. Categories 1B (slight negative slope) and 3 (broadband flat) contain only eight spectra each and category 5 (U-shaped) contains only six spectra. On the other hand, category 8 (mid-to-high-frequency peaks and valleys) has 222 spectra, and category 9 (mixed peaks and valleys) contains 104 spectra. Together, categories 8 and 9 contain over half the total number of spectra. Table A-1 also provides a breakdown by type of sound (aircraft, industrial, etc.). The most striking concentration of spectral shapes is in category 8 (mid-to-high-frequency peaks and valleys) which contains 56% of the artificial spectra.

Evaluation of Subjective Measurements

A. Overall Evaluation

Within each spectral category noises were grouped according to whether or not judged loudness levels were provided in the original study. In those studies where loudness levels were available, it was possible to calculate for each spectral category both mean differences between predicted and observed loudness levels as well as standard deviations (variability measure). Whenever loudness levels were not available, only standard deviations computed from calculated levels could be obtained. For every category of spectra eight overall values based on four different frequency-weighting functions (A, D1, D2, E) and four different calculation schemes (Mark VI, Mark VII, Perceived Noise Level, Zwicker) were computed. The eight functions and schemes are described in greater detail in Table II of Scharf, et. al. (1977).

Table A-2 shows the computations of standard deviations determined by those studies that did not provide calibrated loudness levels. A total of 298 spectra from 11 studies listed in Table A-3 contributed to the values

TABLE A-2 STANDARD DEVIATIONS (IN DECIBELS) COMPUTED FROM CALCULATED LEVELS
(Values were weighted within each category according to the number of spectra per study.)

Category	Frequency Weighting				Calculation Procedure				Number of Spectra	Number of SDs
	A	D1	D2	E	VI	VII	PNL	ZWI		
1A - strong, neg. slope	2.8	2.7	2.6	2.9	2.4	2.8	2.7	1.9	27	7
1B - slight, neg. slope	-	-	-	-	-	-	-	-	0	0
2 - positive slope	4.1	3.3	3.3	3.3	3.1	3.3	3.0	4.0	12	1
3 - broadband, flat	-	-	-	-	-	-	-	-	0	0
4A - narrow-band, low frequency	3.6	1.1	2.3	1.7	1.2	2.2	1.4	2.5	13	1
4B - narrow-band, mid frequency	2.4	2.2	2.3	1.7	2.0	2.3	2.2	3.0	10	1
4C - narrow band, high frequency	2.8	3.2	3.0	2.2	1.8	1.9	3.0	5.0	12	1
5 - U-shaped	-	-	-	-	-	-	-	-	0	0
6 - inverted U-shaped	2.8	2.3	2.5	2.2	1.8	1.6	2.0	1.6	22	5
7 - low-frequency peaks and valleys	1.5	1.5	1.5	1.6	1.2	1.0	1.3	1.1	53	6
8 - mid-to-high frequency peaks and valleys	2.4	1.8	1.7	1.7	1.9	1.9	2.0	2.5	106	6
9 - mixed peaks and valleys	2.1	2.0	1.8	1.8	2.0	2.1	1.9	2.6	43	6
MEAN (unweighted) (in decibels)	2.7	2.2	2.3	2.1	1.9	2.1	2.2	2.7	298	34

Table A-3 List of Studies that Contributed to Table A-2

No. SDs/study	Author	Year	No. Spectra/study
2	Borsky	1974	10
3	Kryter	1959	13
1	Pearsons and Bennett, part 1	1969	30
1	Pearsons and Bennett, part 3	1969	20
2	Peasons and Wells	1969	38
2	Robinson and Bowsher	1961	3*
1	Wells (Aircraft)	1970	29
1	Wells (Unpublished)	c.1970	30
1	Wells 300	1969	39
2	Wells 400	1969	58
1	Wells UHV	1972	25

*Same spectra judged twice, once for loudness and once for annoyance.

shown in Table A-2. None of the 11 studies contained spectra for categories 1B (slight negative slope), 3 (broadband flat), and 5 (U-shaped). Further, with the exception of the study by Robinson and Bowsher (1961) which provided both equal-loudness and equal-annoyance judgments, only equal-annoyance judgments could be obtained from this group of studies.

Each value in Table A-2 is a weighted mean standard deviation. That is, within each category the standard deviations for an individual study were weighted according to the number of spectra per study. The criterion established for inclusion of a group of sounds was a minimum of three spectra per study per category. Moreover, whenever a study consisted of more than one experiment, standard, or group of sounds, the standard deviation for each part was determined separately before computing the weighted average for that particular study. This added consideration is reflected in the column in Table A-2 labeled "number of standard deviations". Therefore, except for categories 2 (positive slope) and 4 (narrow-band noises), this number is larger than the number of studies that contributed to the standard deviation for a given category.

The method of averaging the results of a particular study may have an important effect on the outcome. When a single overall standard deviation is calculated across diverse portions of a study, such as differences in procedures, standards, or type of spectra (artificial versus natural sounds), the standard deviation is inflated. In addition to parts 1 and 3 of Pearsons and Bennett (1969) that were kept separate in the analysis undertaken by Scharf, et.al.(1977), the following studies were also divided into parts: Borsky (1974), two parts; Kryter (1959), three parts; Pearsons and Wells (1969), two parts; Wells 400 (1969), two parts.

Finally, for each weighting or calculation procedure a mean standard deviation (unweighted) determined across categories is shown in Table A-2. The mean standard deviations (unweighted) and standard deviation of standard deviations across spectral categories and across studies are presented in Table A-4. The values calculated across studies were obtained from those values in Scharf, et. al. (1977) that contributed to the spectral analysis shown in Table A-2.

Table A-4 shows that regrouping the noises according to similar spectral categories increases the mean standard deviations across weighting and calculation procedures by an average of about 0.1 dB and decreases the standard deviation of standard deviations by an average of 0.2 dB. Thus, the overall variability of these data is about the same whether they are grouped according to study or according to spectral shape. However, regardless of how these data are grouped, the A-weighting and Zwicker's scheme produce the largest SDs and Mark VI produces the smallest SDs.

More information from Table A-2 can be obtained by evaluating the results category by category. The outcome of such an evaluation is summarized in Table A-5. According to Table A-5, of the nine categories for which spectra are available, the A-weighting as well as Zwicker's scheme produce the largest standard deviation for five out of nine categories. On the other hand, Mark VI or Mark VII yield the smallest standard deviations for seven of nine categories. (See Section II, Table III for a discussion of the statistical analysis.)

TABLE A-4

Mean Standard Deviations (in decibels) and Standard Deviation of Standard Deviations (in decibels) across Spectral Categories and across Studies (for which loudness levels were not available)

	Frequency Weighting				Calculation Procedure				
	A	D1	D2	E	VI	VII	PNL	ZWI	N*
Mean SD (Unweighted)									
1) across spectral categories	2.7	2.2	2.3	2.1	1.9	2.1	2.2	2.7	12
2) across studies	2.5	2.1	2.1	2.2	1.9	2.0	2.1	2.4	11
SD of SDs									
1) across categories	.78	.75	.60	.62	.58	.64	.63	1.2	12
2) across studies	1.0	.92	.88	.95	.78	.79	.90	1.4	11

*Number of studies and parts or, number of spectral categories.

TABLE A-5

Analysis of Standard Deviations (in Decibels) According to Categories
(Based on Calculated Levels)

Category	No. Spectra/No SDs	Largest SD	Smallest SD
1A	27/7	E	Mark VI, Zwicker
2	12/1	A, Zwicker	Mark VI, PNL
4A	13/1	A	D1, Mark VI
4B	10/1	Zwicker	E, Mark VI
4C	12/1	Zwicker	Mark VI, Mark VII
6	22/5	A, D2	Mark VII,Zwicker
7	53/6	A, D2, E	Mark VII,Zwicker
8	106/6	A, Zwicker	D2, E
9	43/6	Zwicker	D2, E

Similarly, standard deviations were computed for each of the nine spectral categories using data from the 10 studies, listed in Table A-6, that provided loudness levels.

The results are shown in Table A-7. A total of 335 noises are distributed across the nine spectral categories. More than one-third of the noises are found in category 8 (mid-to-high-frequency peaks) and the lowest number are found in category 4A (narrow band low-frequency noises).

Initially, standard deviations within a specific category were computed across studies representing a wide range of loudness levels and mean differences. No attempt was made to group studies or spectra or to obtain an actual weighted SD according to the number of spectra per study. This procedure led to standard deviations as large as 6 dB for categories 1A (strong negative slope) and 9 (mixed peaks) and as large as 5 dB for categories 4B (narrow band mid-frequencies) and 8 (mid-to-high-frequencies). By combining data across studies and computing a single standard deviation the possible effect of spectral distribution of energy on standard deviations is obscured by the very large variation among studies. A better assessment of variability within categories is achieved by first calculating the standard deviation for each individual study or group of individual spectra, averaging these standard deviations for all studies within a given category and then calculating a weighted or unweighted mean across spectral categories. This revised procedure was used for determining the within category standard deviations shown in Table A-7. For the same reasons, i.e., to reduce to a minimum the within and between study variations, it closely followed the procedure used to describe the standard deviations indicated in Table A-2. The initial and

TABLE A-6 List of Studies that Contributed to Table A-7

No. Spectra/Study	Study	Year
18	Berglund, et. al.	1976
105	Fishken	1971
10	Jahn	1965/66
8	Kryter and Pearsons	1963
31	Lübcke, et. al.	1964
30	Molino	1976
37	Quietzsch	1955
24	Rademacher	1959
39	Spiegel	1960
33	Yaniv	1976

TABLE A-7 STANDARD DEVIATIONS (in decibels) COMPUTED FROM DIFFERENCES BETWEEN COMPUTED AND OBSERVED LOUDNESS LEVELS (Standard deviations were first computed within each study and then weighted within a category according to the number of spectra.)

	Weighting Scheme				Calculation Procedure					No.
Category	A	D1	D2	E	VI	VII	PNL	ZWI	Number	SDs
1A - strong, neg. slope	3.8	3.5	3.6	3.5	3.1	2.6	4.1	3.1	37	5
1B - slight, neg. slope	3.7	2.3	3.4	2.4	2.4	2.2	2.6	1.7	8	2
2 - positive slope	3.95	4.2	4.2	4.4	2.6	2.7	4.0	3.0	10	2
3 - broadband flat	3.5	3.2	3.2	3.3	2.0	2.1	3.9	2.2	8	2
4A - narrow-band, low freq. noises	3.4	3.3	3.3	3.4	3.8	3.7	3.5	3.6	5	1
4B - narrow-band, mid freq. noises	1.5	1.6	1.6	1.5	1.3	1.6	1.5	1.4	12	3
4C - narrow-band, high freq. noises	1.9	2.2	2.1	2.8	2.1	1.8	1.7	1.4	6	1
5 - U-shaped	3.7	4.2	4.1	4.0	1.7	2.0	3.5	2.5	6	2
6 - -shaped	3.2	3.2	3.3	3.2	2.9	2.9	3.2	2.8	46	7
7 - low-frequency peaks	1.4	1.2	1.3	1.1	1.6	1.5	2.4	1.4	20	4
8 - mid-to-high frequency peaks	3.0	4.0	4.0	3.7	3.2	3.4	3.7	2.7	116	12
9 - mixed peaks	1.8	2.5	2.4	2.2	2.3	2.4	2.7	2.2	61	15
Means (Unweighted) (N = 12) (in decibels)	2.9	3.0	3.0	3.0	2.4	2.4	3.1	2.3	Total N=335 noises	
Means (Weighted according to number of spectra) (N = 335) (in decibels)	2.8	3.2	3.25	3.1	2.7	2.75	3.3	2.5		
Means in Scharf, et. al. Table II, corrected re: Appendix D, Table D-1 (Based on 20 studies) (in decibels)	3.05	2.65	2.73	2.63	2.26	2.22	2.6	2.		

revised procedures were the same for calculating the standard deviations across categories.

Each value in Table A-7 for a specific spectral category and weighting or calculation scheme is a weighted mean standard deviation. Within each spectral category individual standard deviations were weighted according to the number of spectra per study, provided the study had at least three spectra. In some studies, spectra fell naturally into groups according to such variables as signal-to-noise ratio or overall sound pressure level, e.g., Lubcke, et. al. (1964), Spiegel (1960), Fishken (1971), and Yaniv (1976). Standard deviations were then computed for each grouping within a study. On the other hand, whenever the number of spectra per study fell below the minimum number of three, the results of more than one study or overall sound pressure level were combined to produce a single estimate of the standard deviation. Hence, as in Table A-2, the numbers in Table A-7 in the column labeled "number of standard deviations" do not necessarily reflect the number of studies that contributed to the standard deviations for a given category. For this analysis, the number of standard deviations is sometimes less than the number of contributing studies.

Compared to the initial procedure (i.e. computing a single estimate of the standard deviation across studies within a spectral category), the revised procedure (i.e., taking into account standard deviations for individual studies or groups of spectra within a category before averaging) reduced substantially the standard deviations computed both within and across categories. Only for categories 4A (narrow band low frequency) and 4C (narrow band high frequency) that are based on one standard deviation do the initial and revised procedures yield the same result. Within categories 1A (strong negative slope), 4B (narrow band mid-frequency), 8 (mid-to-high-frequency peaks), and 9 (mixed peaks), the revised procedure reduced the maximum standard deviation to 4.0 dB. The mean standard deviations (unweighted) and standard deviation of standard deviations calculated across categories according to both the initial and revised within category procedures are indicated in Table A-8. Also shown are the mean standard deviations (unweighted) and standard deviation of standard deviations calculated across studies. Those values were obtained from the standard deviations in Scharf, et. al. (1977) that contributed to the spectral analysis shown in Table A-7.

According to Table A-8, the revised procedure reduces the mean standard deviation across categories by an average of 0.8 dB for the four frequency-weighting procedures and by an average of 1.1 dB for the four calculation schemes. The mean standard deviation determined by the revised procedure is about the same as the mean SD calculated across studies, but the SD of SDs is about 0.2 dB smaller. Further, regardless of how these data are grouped, the calculation schemes, with the exception of PNL, produce mean SDs about 0.5 dB smaller than the four frequency weightings.

A finer analysis of Table A-7 can be accomplished by examining the results, spectral category by spectral category. The results are summarized in Table A-9. In contrast to the results in Table A-5, the A-weighting fares much better when loudness levels are provided. Only for categories 1A (strong negative slope) and 1B (slight negative slope) does the A-weighting yield the largest SDs. For category 9, involving 61 spectra with mixed peaks, the A-weighting produces the smallest SDs. On the other hand, the D1-, D2-, and E-weightings produce the largest SDs for 6 out of 12 spectral categories. Mark VI, Mark VII, and Zwicker calculation procedures perform about equally

TABLE A-8

Comparison of Mean Standard Deviations (in Decibels) and Standard Deviations of Standard Deviations (in Decibels) Across Spectral Categories and Across Studies (Loudness Levels Provided)

	Frequency Weighting				Calculation Procedure				
	A	D1	D2	E	VI	VII	PNL	ZWI	N*
Mean SD (Unweighted)									
1) Across Categories									
Initial Procedure	3.7	3.9	3.9	3.8	3.5	3.5	4.0	3.6	12
Revised Procedure	2.9	3.0	3.0	3.0	2.4	2.4	3.1	2.3	12
2) Across Studies (Based on 10 studies)	3.2	3.0	3.1	2.9	2.5	2.4	2.9	2.3	16
SD of SDs									
1) Across Categories									
Initial Procedure	0.9	1.1	1.0	0.9	1.4	1.5	1.2	1.5	12
Revised Procedure	1.0	1.0	1.0	1.0	0.7	0.7	0.9	0.7	12
2) Across Studies (Based on 10 studies)	1.3	1.2	1.0	1.2	0.85	1.1	1.2	1.0	16

*N = Number of studies and parts, or number of spectral categories.

LEGEND:

Initial Procedure: Within a specific category a single estimate of the standard deviation was computed across studies.

Revised Procedure: Obtained first the standard deviation for each individual study or group of individual spectra and then obtained a weighted mean within a category.

Table A-9

Analysis of Standard Deviations (in Decibels) According to Categories (Loudness Levels Provided)

Category	No. Spectra/ No. SDs	Largest SD	Smallest SD
1A	37/5	A, PNL	Mark VII
1B	8/2	A	Zwicker
2	10/2	D1, D2, E	Mark VI, Mark VII
3	8/2	PNL	Mark VI, Mark VI, Zwicker
4A	5/1	Mark VI, Mark VIII	D1, D2
4B	12/3	D1, D2, Mark VII	Mark VI
4C	6/1	E	Zwicker
5	6/2	D1	Mark VI
6	46/7	D2	Zwicker
7	20/4	PNL	D1, E
8	116/12	D1, D2	Zwicker
9	61/15	PNL	A

well producing the smallest SDs for 5 out of 12 categories. Mark VI and Mark VII calculation procedures perform well for both groups of studies - those that provided judged loudness levels and those that did not. (See Section II, Table III on the spectral categories for a discussion of the statistical analysis of these data).

In addition to an analysis of standard deviations within spectral categories, for those 10 studies that provided loudness levels it was also possible to perform a within category analysis of mean differences between calculated and observed levels. Table A-10 shows the results for each of the eight frequency-weighting and calculation procedures and for the same 335 noises upon which the standard deviations in Table A-7 are based. Each value for a specific category and weighting or calculation scheme represents a weighted mean difference. Within a category the mean differences for an individual study were weighted according to the number of spectra per study. The mean differences calculated across categories, the standard deviations of the means, and the range of mean differences for each weighting and calculation scheme are also indicated in Table A-10*.

Table A-10 suggests that the A-weighting produces the largest mean difference, the largest standard deviation, and the largest range of mean differences. The smallest overall mean difference is produced by Mark VI and Perceived Noise Level calculation procedures. Zwicker's procedure produces the smallest standard deviation as well as the smallest range of mean differences. The differences between Zwicker's procedure and Mark VI

*Note that, whereas the means within categories are weighted values, the means calculated across categories are unweighted.

Table A-10

CALCULATED MINUS OBSERVED LOUDNESS LEVEL (in decibels)

Category	Frequency Weighting A	D1	D2	E	Calculation Procedure VI	VII	PNL	ZWI	Number of Spectra
1A - strong, neg. slope	-15.4	-8.2	-10.0	-9.8	-2.5	-10.1	-2.6	+3.7	37
1B - slight, neg. slope	-18.1	-10.5	-11.7	-12.6	-6.5	-12.9	-5.1	-0.15	8
2 - positive slope	-14.0	-5.8	-5.4	-6.8	-0.82	-9.3	-2.2	+4.0	10
3 - broadband flat	-16.7	-9.9	-10.0	-11.9	-4.0	-11.4	-3.9	+1.1	8
4A - narrow-band, low freq. noises	-11.5	-3.2	-5.6	-4.3	-2.9	-10.7	-2.1	-0.02	5
4B - narrow-band, mid freq. noises	-1.96	-0.83	-0.56	-1.46	+3.56	-4.65	+2.0	+6.7	12
4C - narrow-band, high freq. noises	-9.9	-0.63	-0.47	-2.07	-0.27	-7.97	+1.2	-0.8	6
5 - U-shaped	-16.4	-7.95	-7.9	-9.25	-2.8	-10.6	-4.0	+0.1	6
6 - inverted U-shaped	-13.6	-8.0	-8.3	-9.8	-2.9	-10.2	-2.6	+2.6	46
7 - low-frequency peaks	-10.4	-5.0	-6.0	-6.3	-0.8	-6.2	-2.7	+5.8	20
8 - mid to-high frequency peaks	-4.8	+2.1	+2.1	-0.04	+5.1	-1.9	+5.7	+8.0	116
9 - mixed peaks	-12.3	-5.9	-6.2	-7.8	+0.62	-6.7	-0.11	+5.9	61
Mean (unweighted)	-12.1	-5.3	-5.8	-6.8	-1.2	-8.6	-1.4	+3.1	TOTAL 335
SD	4.8	4.0	4.3	4.1	3.2	3.2	3.1	3.0	
Range	16	12.6	13.7	12.6	11.6	11.1	10.8	8.8	

and Perceived Noise Level are so small that they are probably not statistically significant. Moreover, in contrast to Table IV of Scharf, et. al. (1977) which suggests that the standard deviation of mean differences across the same eight frequency-weighting and calculation procedures varies between 4 and 5 dB, regrouping the data on the basis of spectral categories reduces the standard deviation of means for the calculation procedures to an average value of 3.1 dB. This value is about 1.2 dB smaller than the standard deviation of means computed for the four frequency weightings. Due to the large differences in number of spectra that contributed to the weighted mean differences among the nine spectral categories, a meaningful, statistical analysis of these data could not be accomplished. Nevertheless, they do suggest that regrouping the data into similar spectral categories produces an advantage to the four calculation procedures but not to the four frequency-weighting functions.

A more detailed analysis of Table A-10 can be obtained by evaluating the results category by category, as summarized in Table A-11. Table A-11 shows that the A-weighting consistently underestimates the subjective magnitude of noise for most categories of spectra, i.e., it produces the largest mean difference for 10 out of 12 categories. For the remaining two categories (4B, narrow band mid-frequency; 8, mid-to-high-frequency peaks), Zwicker's procedure produces the largest mean difference. On the other hand, Mark VI produces the smallest mean difference for six out of the 12 categories. The results in Tables A-10 and A-11, together with the analysis of Tables A-2 and A-7, indicate that the current ANSI standard (1972), Mark VI, is probably most generally suitable for predicting the loudness or noisiness of noise, despite the small differences between Mark VI and the other descriptors evaluated.

Table A-11

Analysis of Mean Differences According to Categories
(Loudness Levels Provided)

Category	No. Spectra/ No. Mean Diffs.	Largest Mean Diffs.	Smallest Mean Diffs.
1A	37/14	A	Mark VI, PNL
1B	8/3	A	Zwicker
2	10/4	A	Mark VI
3	8/3	A	Zwicker
4A	5/3	A	Zwicker
4B	12/8	Zwicker	D1, D2
4C	6/3	A	Mark VI
5	6/2	A	Zwicker
6	46/7	A	Mark VI, PNL, Zwicker
7	20/5	A	Mark VI
8	116/15	Zwicker	D1, D2, E
9	61/20	A	Mark VI, PNL

B. Effect of Tonal Components on Analysis of Categories 7, 8, and 9

The categorical analysis described in Section II also permits a preliminary assessment of the need for a tone correction procedure to be applied to the existing weighting and calculation procedures. Table A-1 indicates that about two-thirds of the stimuli evaluated are contained within categories 7, 8, and 9 (low-frequency, mid-to-high-frequency, and mixed peaks, respectively). A more detailed analysis of these spectra was performed to determine (1) what proportion of spectra in each spectral category included tonal components, (2) whether within a category those spectra with tonal components produce a larger variability and larger mean differences than do those spectra without tonal components, and (3) whether a specific frequency-weighting or calculation procedure was more suited than another for predicting the perceived magnitude of noise-tone complexes. Within each category, noises were grouped according to whether or not loudness levels were provided in the original study.

Table A-12 and A-13 provide two sets of weighted standard deviations for categories 7, 8, and 9; within each category one set of SDs is for spectra that contained peaks and valleys both with and without tones, and the other set for such spectra without tones. Category 9 in Table A-12 includes an additional set of standard deviations with tones. The presence of tonal components was based on criterion developed for the tone-correction procedures described in section III.

Table A-12 and A-13 show that most of the sounds in categories 7 and 8 contained tonal components. Owing to the large differences in the number of spectra in the total groups and the groups without tones, a comparison of SDs is inappropriate. (The larger the n, the larger the SD tends to become.)

Table A-12

Standard Deviations (in Decibels) Computed from Calculated Levels for Categories 7, 8, and 9. (Loudness Levels not Measured in Original Studies. Values were Weighted within Each Category According to the Number of Spectra Per Study)

		Frequency Weighting				Calculation Procedure			
Category	Number Spectra	A	D1	D2	E	VI	VII	PNL	ZWI
7	53 (with and without tones)	1.5	1.4	1.5	1.5	1.2	1.1	1.3	1.1
	7 (without Tones)	1.65	1.1	1.4	1.2	0.6	0.6	0.5	1.1
8	106 (with and without tones)	2.4	1.8	1.7	1.7	1.9	1.9	2.0	2.5
	20 (without tones)	1.7	1.8	1.8	1.6	1.3	1.5	1.8	1.4
9	43 (with and without tones)	2.1	2.0	1.8	1.8	2.0	2.1	1.9	2.6
*	21 (without tones)	1.3	1.1	1.2	0.85	0.83	0.95	1.0	1.4
*	18 (with tones)	3.5	2.6	2.25	2.4	2.4	2.5	2.5	2.7

*Two spectra were not included in this analysis because they did not satisfy the criterion of at least three spectra per study required for computation of a standard deviation.

Table A-13

Standard Deviations (in Decibels) Computed from Differences Between Computed and Observed Loudness Levels for Categories 7, 8, and 9 (Values were Weighted within Each Category According to the Number of Spectra Per Study)

		Frequency Weighting				Calculation Procedure			
Category	Number Spectra	A	D1	D2	E	VI	VII	PNL	ZWI
7	20 (with and without tones)	1.4	1.2	1.3	1.1	1.6	1.5	2.4	1.4
	14 (without tones)	1.3	1.4	1.4	1.2	1.3	1.6	2.4	1.7
8	116 (with and without tones)	3.0	4.0	4.0	3.7	3.2	3.4	3.7	2.7
	10 (without tones)	2.7	2.9	2.8	2.5	2.2	2.8	2.5	2.1
9	61 (with and without tones)	1.8	2.5	2.4	2.2	2.3	2.4	2.7	2.2
	51 (without tones)	21	2.8	2.7	2.6	2.5	2.8	2.8	2.4

Only in category 7 for loudness levels (Table A-13) can a comparison be made, and there the 14 spectra without tones tended to give slightly larger SDs for 5 of the 8 descriptors than did the 20 spectra with and without tones. However, the overall difference of 0.2 dB is not meaningful. Category 9 provides a more even distribution for those sounds judged with respect to an evaluative attribute (Table A-12). There the SDs are larger by about 1.5 dB for the 18 sounds with tones than for the 21 sounds without. This finding suggests, quite tentatively, that in judgments of noisiness, unacceptability, etc., tonal components may increase the variability of the descriptors for spectra that contain mixed peaks and valleys.

APPENDIX B

"ANOMALOUS" DATA

The term anomalous data is used as short hand for the six studies in Scharf, et. al. (1977) that produced the largest SDs (see Scharf et. al.,1977, Table II). A closer examination of those studies reveals characteristics that distinguish them from the average of the 20 studies and especially from those studies that yielded the smallest SDs.

Table B-1 shows the standard deviations produced by the six anomalous studies and those produced by all 20 studies. For every descriptor, the average standard deviation from the six studies is not only larger than from the entire group of studies, but the disparity is larger by about 0.5 dB for the six weighting functions than for the five calculation procedures.

Table B-2 provides a comparison between the six anomalous studies and the six studies that produced below average standard deviations. Values are given for eight descriptors; B, C, and PNLC are omitted. The mean standard deviation from the anomalous studies is about 2 dB larger than from the least variable studies. The average standard deviation for the weighting functions is about 0.5 dB larger than that for the calculation procedures. An examination of the less variable studies show that they share the following characteristics: a) the spectra tend to be fairly homogeneous; b) the stimuli are exclusively natural, as opposed to artificial, sounds; c) the range of sound pressure levels in a study is less than 25 dB; d) the standard deviations are based on a single set of measurements or experimental conditions.

On the other hand, those studies that produced unusually large SDs differed from the least variable studies in at least one of the characteristics indicated below.

Table B-1

Standard Deviations (in Decibels) from
Six Studies Yielding the Greatest Variability

Study	Frequency Weighting------						Calculation Procedure------				
	A	B	C	D1	D2	E	VI	VII	PNL	PNLC	ZWI
Fishken (1971)											
84/12*	2.7	2.9	3.0	3.9	3.9	3.6	2.9	2.8	3.8	3.6	2.5
21/3*	4.5	4.6	4.6	4.4	4.4	4.4	4.4	5.4	3.4	3.5	3.7
Pearsons and Bennett (part 1, 1969)											
30/30	4.3	4.5	4.7	3.5	3.7	3.3	2.8	2.8	2.9	2.2	3.7
Pearsons, et.al. (1968)											
108/54*	6.5	5.1	5.3	2.5	2.8	3.0	2.2	2.2	3.0	2.6	2.1
Spiegel (1960)											
20/20	4.7	6.2	6.8	4.2	4.0	4.2	2.4	1.9	3.2	3.7	2.4
20/20	5.3	4.9	5.1	3.5	4.1	3.6	2.6	2.6	2.9	3.2	3.0
Quietzsch (1955)											
27/27	4.2	4.4	5.7	4.0	4.3	4.2	3.1	3.2	4.0	4.2	3.3
10/10	3.8	6.3	7.0	3.3	2.9	3.8	2.5	2.5	2.6	2.8	2.5
Wells 300-400 (1969a)											
300- 42/42	3.7	5.2	6.6	2.4	2.7	2.1	2.1	2.2	2.3	2.4	5.3
400- 60/60	2.5	4.2	4.9	2.5	2.0	2.6	2.5	2.6	2.5	1.8	3.1
$\bar{X}$ 6 Studies (N = 10)	4.2	4.8	5.4	3.4	3.5	3.5	2.8	2.8	3.1	3.0	3.2
$\bar{X}$ 20 Studies (N = 28) Scharf, et.al.1977 Table II, corrected re: Appendix D, Table D-1	3.1	3.6	4.2	2.7	2.7	2.6	2.3	2.2	2.6	2.7	2.4

* The number in front of the slash is the number of conditions (e.g. different sound pressure levels, instructions, etc.)

** N = number of standard deviations

Table B-2

Standard Deviations (in Decibels) from
Six Studies that Produced the Smallest Variability

Study	Frequency Weighting A	D1	D2	E	Calculation Procedure VI	VII	PNL	ZWI
Jahn (1965/66) 10/10	1.3	1.2	1.3	1.2	0.9	0.9	1.0	0.8
Pearsons and Bennett (1969, part 3) 20/20	1.7	1.4	1.4	1.7	1.3	1.5	1.3	1.8
Robinson and Bowsher (1961)	1.9	1.4	1.5	1.9	1.2	1.6	1.1	0.9
Wells (1970)	1.6	1.2	1.3	0.9	1.2	1.2	1.3	2.2
Wells (Unpublished) (1970)	1.1	1.3	1.3	1.1	0.9	0.9	1.2	1.1
Wells (UHV) (1972)	1.5	1.3	1.5	1.3	1.1	1.0	1.3	0.9
X 6 studies (N = 6)	1.5	1.3	1.4	1.4	1.1	1.2	1.2	1.3
X 6 anomalous studies (N = 10)	4.2	3.4	3.5	3.5	2.8	2.8	3.1	3.2

Characteristics of Anomalous Studies

a) The spectra tend to be heterogeneous.

Pearsons, et. al. (1968, 1969)

Quietzsch (1955)

Spiegel (1960)

b) The spectra include only artificial sounds.

Fishken (1971)

Pearsons and Bennett, Part I, (1969); Pearsons, et. al. (1968, 1969)

Spiegel (1960)

Wells 300-400 series (1969a)

c) The range of levels is large

Fishken (1971)

Quietzsch (1955)

d) The standard deviations are based on more than one set of experimental conditions or measurements.

Fishken (1971)

Pearsons, et. al. (1968, 1969)

These characteristics suggest under what conditions a group of sounds is likely to be less well assessed by the descriptors.

A detailed analysis of five of the six anomalous studies follows.

Spiegel (1960)

Spiegel's (1960) study illustrates how averaging data produced by heterogenous spectra inflates standard deviations. Spiegel studied 20 noises distributed across six spectral categories (2 (positive slope), 4A (narrow band low-frequency), 4B (narrow band mid-frequency), 4C (narrow band high-frequency), 5 (U-shaped) and 6 (inverted U-shaped)). Measurements were made at two loudness levels, 64 phons and 85.5 phons. The standard deviations computed separately for each spectral category produce an average value smaller than the single standard deviation computed for all 20 noises, as in Table B-1. Table B-3 presents a re-evaluation of Spiegel's study. Both sets of mean standard deviations, weighted and unweighted, are considerably smaller than the overall mean standard deviations shown in Table B-1.

Recomputing the mean differences between observed and calculated loudness levels by first calculating the means for a given category and then computing the overall means only slightly reduces the overall mean differences.

Wells 300 series (1969a)

The measurements by Wells provide another example of the way in which homogeneous grouping of spectra can reduce SDs. Wells's 300 series comprised mainly octave-band noises both with and without tones. The large SDs in Table B-1 can be ascribed to heterogeneity of spectra both across and within categories.

The first source of variability can be reduced by computing the SDs separately for each spectral category before obtaining the overall average standard deviation. Table B-4 shows that the mean standard deviations (weighted or unweighted) computed across categories are smaller by about 0.5 dB than the previously determined average values. The A-weighting and Zwicker's procedure show the largest reductions. The A-weighting does least well for narrow-band, low-frequency noises, and Zwicker's procedure is poorest for narrow-band, high-frequency noises. Further, with the exceptions of the A-weighting, Mark VI, and Mark VII, the SDs tend to be larger for category 4C which consists of high-frequency noises than for categories 4A and 4B.

Table B-3

Spiegel Study (Standard Deviations in Decibels)

			Frequency Weighting				Calculation Procedure			
			A	D1	D2	E	VI	VII	PNL	ZWI
CAT.	LL	N								
2	64	2	0.31	0.51	0.47	0.46	0.54	0.01	0.25	0.45
2	85	2	1.4	1.2	1.3	1.3	1.4	1.3	1.4	1.0
4a	64	2	0.58	0.41	0.53	0.44	0.60	0.70	0.25	0.20
4a	85	2	0.84	1.0	0.88	0.97	0.89	1.06	1.34	1.49
4b	64	3	1.9	2.0	2.0	1.8	1.4	1.5	1.65	1.60
4b	85	3	2.1	2.3	2.4	2.2	1.6	2.0	2.1	2.2
4c	64	2	3.7	3.8	3.9	4.0	3.1	2.2	3.4	1.4
4c	85	2	0.21	0.25	0.31	0.43	0.22	0.55	0.33	2.0
5	64	3	4.3	4.7	4.7	4.6	2.6	2.5	4.5	3.4
5	85	3	3.1	3.6	3.5	3.3	0.72	1.4	2.5	1.5
6	64	4	3.5	2.6	2.3	2.8	2.7	1.9	2.4	2.7
6	85	4	6.1	4.0	5.1	3.8	3.7	3.7	4.1	4.0
	(N=32)									
X, unweighted across categories										
	64	16	2.4	2.3	2.3	2.4	1.8	1.5	2.1	1.6
	85	16	2.3	2.1	2.2	2.0	1.4	1.7	2.0	2.0
X, weighted across categories										
	64	16	2.6	2.5	2.0	2.5	2.0	1.6	2.2	1.9
	85	16	2.8	2.4	2.1	2.3	1.7	1.9	2.3	2.3
X, from Table B-1										
	64	20	4.7	4.2	4.0	4.2	2.4	1.9	3.2	2.4
	85	20	5.3	3.5	4.1	3.6	2.6	2.6	2.9	3.0

Table B-4

Wells 300 Series (Standard Deviations in Decibels)

	N	Frequency Weighting A	D1	D2	E	Calculation Procedure VI	VII	PNL	ZWI
Category 4a	13	3.6	1.1	2.3	1.7	1.2	2.2	1.4	2.5
4b	10	2.4	2.2	2.3	1.7	2.0	2.3	2.2	3.0
4c	12	2.8	3.2	3.0	2.2	1.8	1.9	3.0	5.0
$\overline{X}$ SD (unweighted)		2.9	2.2	2.5	1.9	1.7	2.1	2.2	3.5
$\overline{X}$ SD (weighted)		3.0	2.1	2.5	1.9	1.6	2.1	2.2	3.5
$\overline{X}$ from Table B-1	42	3.7	2.4	2.7	2.1	2.1	2.2	2.3	5.3

A further reduction in standard deviations can be obtained by subdividing spectra within a category into two groups, one with tones and another without tones. The results of this analysis for an arbitrarily chosen subgroup of 20 of Wells's spectra are presented in Table B-5 together with an analysis of results for categories 4A and 4C. The overall analysis shows that, for seven out of eight descriptors, spectra with pure tones produce larger standard deviations than spectra without tones. Similarly, for categories 4A and 4C the presence of pure tones enlarges standard deviations for six of the eight descriptors.

Wells 400 series (1969a)

The Wells 400 series can be analyzed in the same way as the Wells 300 series. Wells's 400 series consisted mainly of broadband noises that contained either single or multiple pure tones. Of the 60 noises, 57 fell into categories 7, 8, and 9 (low-frequency, mid-to-high-frequency, and mixed peaks, respectively). The relatively large SDs in Table B-1 can be ascribed to the presence of multiple and single pure tones as well as to the heterogeneity of spectra across categories. It is possible, for example, to subdivide the spectra of the Wells' 400 series into two groups, one that consisted of noise-tone complexes with single tones and another that consisted of noise-tone complexes with multiple tones. The results of this analysis for six spectra that contained multiple pure tones and for 12 arbitrarily chosen spectra that contained single tones are shown in Table B-6. With the exception of Zwicker's procedure, the presence of multiple pure tones produces larger calculated SDs than the presence of single tones.

Table B-5

Wells 3001 Series Comparison of SDs (In Decibels) Produced by Octave-Band Noises With and Without Tones

		Frequency Weighting				Calculation Procedure			
	N	A	D1	D2	E	VI	VII	PNL	ZWI
With Tones	9	5.2	2.9	3.8	2.5	1.3	2.8	3.2	6.2
Without Tones	11	3.5	1.6	2.5	0.9	2.2	1.8	1.9	3.9
Cat. 4A									
With Tones	4	5.5	0.7	3.1	2.4	0.24	2.9	0.5	2.8
Without Tones	3	3.4	0.3	1.7	0.9	0.7	0.9	0.7	1.7
Cat 4C									
With Tones	3	2.8	4.8	5.0	3.5	1.4	3.4	4.9	7.0
Without Tones	3	3.4	2.1	1.9	.94	1.6	1.0	1.6	3.3

Table B-6

Wells 400 Series. Comparison of SDs (in Decibels) Produced by Wide-Band Noises with Single and with Multiple Pure Tones

	N	Frequency Weighting A	D1	D2	E	Calculation Procedure VI	VII	PNL	ZWI
Single Tones	12	1.6	1.4	1.2	1.2	1.6	1.6	1.5	2.0
Multiple Tones	6	2.7	3.6	3.0	3.5	2.0	2.3	2.9	1.8
$\overline{X}$ (Weighted)	18	2.0	2.1	1.8	2.0	1.7	1.8	2.0	1.9

The variability across categories can be reduced by computing the standard deviation for each category separately before computing the overall SD. Table B-7 shows the results of this analysis as well as the previously calculated estimate of the standard deviation for the Wells 400 series. Two features of Table B-7 are of interest. First, the calculation of a single estimate of the standard deviation across diverse portions of a study enlarges slightly the standard deviation. Second, as suggested by the Wells 300 series, those spectra that only contain low-frequency tonal spikes produce smaller standard deviations than do those spectra that contain tonal spikes above 500 Hz. The A-weighting and Zwicker's procedure produce the largest standard deviations for category 8 (mid-to-high-frequency peaks), and Zwicker's procedure also produces the largest standard deviation for category 9 (mixed peaks).

Quietzsch (1955)

Quietzsch's results cannot be analyzed in the same straightforward manner as those of Spiegel or Wells because his noises varied widely in spectral shape as well as in amplitude. The 37 noises varied from 47 to 98 dB overall sound pressure level and 49 to 106 phons loudness level. Thus, categorizing the sounds according to spectral shape and computing the standard deviation separately for each category has only a small effect on the overall mean standard deviation. In order to evaluate Quietzsch's results it is necessary to determine more exactly the effect of sound pressure level on standard deviations. However, his results have too few noises at each level to make this determination. The specific effect of sound pressure level on SDs will be demonstrated in the section on Fishken's measurements.

Table B-7

Wells 400 Series (Standard Deviations in Decibels)

	N	Frequency Weighting A	D1	D2	E	Calculation Procedure VI	VII	PNL	ZWI
Category 7	18	1.8	2.2	1.8	2.2	1.4	1.3	1.6	1.3
8	36	2.8	2.1	1.8	2.3	2.4	2.6	2.0	3.2
9	3	1.9	2.8	1.8	3.3	3.4	4.2	3.4	4.7
$\overline{X}$ SD (Unweighted)		2.2	2.4	1.8	2.6	2.4	2.7	2.3	3.1
$\overline{X}$ SD (Weighted)		2.4	2.2	1.8	2.3	2.1	2.3	1.9	2.7
$\overline{X}$ from Table B-1	60	2.5	2.5	2.0	2.6	2.5	2.6	2.5	3.1

Pearsons and Bennett (1969)

Pearsons' and Bennett's data, part 1, produced above average standard deviations whereas part 3 produced below average standard deviations. The range of noise levels in parts 1 and 3 is nearly the same, and the noises are distributed among the same number of spectral categories. Parts 1 and 3 appear to differ only in that Part 1 consists exclusively of artificial noises whereas part 3 consists exclusively of natural noises.

Table B-8 shows that artificial noises produced larger SDs than did natural noises. In addition, consistent with the results of the Wells 300-400 series, those spectra that contain only low-frequency tonal spikes produced the lowest SDs.

Further evidence tha' .w-frequency tonal spikes produce smaller standard deviations than mixed low- and high-frequency spikes can be obtained from a within-category analysis of spectra from Pearsons and Bennett part 1, category 2. The 12 spectra in this category of noises that produced a positive slope were divided into two equal subgroups. One group consisted of six noises that contained a low-frequency tonal spike while another group consisted of six noises that contained both a low- and a high-frequency tonal spike. The standard deviations were computed separately for each group. The results are indicated in Table B-9.

Table B-9 shows that, except for the A-weighting, those noises with low-frequency spikes produce smaller standard deviations than those noises with low- and high-frequency spikes. The largest difference in standard deviation between the two groups is produced by the Perceived Noise Level procedure. Table B-9 also shows that the computation of a single standard deviation across diverse portions of a study within a category enlarges the mean standard deviation by about 1.5 dB.

Table B-8

Pearsons and Bennett (Standard Deviations in Decibels)

		Frequency Weighting				Calculation Procedure			
	N	A	D1	D2	E	VI	VII	PNL	ZWI
Category									
6, part 1	7	3.8	2.8	3.2	2.5	1.8	1.6	2.2	1.5
6, part 3	5	1.2	0.9	1.0	0.7	0.9	0.8	1.1	0.9
7, part 1	5	1.8	2.1	2.2	2.0	0.9	0.7	0.9	1.5
7, part 3	2	1.5	0.05	0.6	0.4	0.3	0.4	0.09	0.7
9, part 1	6	3.5	3.4	3.5	3.3	3.1	2.9	3.5	3.1
9, part 3	6	1.5	1.1	1.1	1.0	1.5	1.6	1.3	2.5

Table B-9

Pearsons and Bennett, Part 1 (Standard Deviations in Decibels)

		Frequency Weighting				Calculation Procedure			
		A	D1	D2	E	VI	VII	PNL	ZWI
	N								
Category 2									
Group I Low-frequency spikes	6	2.2	2.0	1.9	1.8	1.5	1.5	1.6	1.7
Group II High-frequency spikes	6	2.1	2.1	2.1	2.1	1.9	2.0	2.2	2.1
$\bar{X}$		2.15	2.05	2.0	1.95	1.7	1.75	1.9	1.9
Single SD computed across category		4.1	3.3	3.3	3.1	3.1	3.3	3.0	4.0

The standard deviations across weighting and calculation procedures for Pearsons and Bennett, part 1, may be reduced about 1.1 dB by computing the standard deviation for each of the four spectral categories separately before computing a mean SD across categories. This procedure also reduces the overall SDs produced for Pearsons and Bennett, part 3, but the decrease in standard deviations is smaller than for part 1. Nevertheless, the discrepancy between the standard deviations produced by parts 1 and 3 remains about 1.2 dB, suggesting that the difference in SDs produced by artificial and natural noises is not easily eliminated.

Fishken (1971)

Fishken measured the overall loudness of broadband noise with tonal spikes in two separate series of experiments. In the first series, the overall SPL of the tone and noise was held constant at one of seven overall sound pressure levels between 30 and 90 dB. At a given overall sound pressure level, both the frequency of the tone or tonal complex and the tone-to-noise ratio were varied. Four different tones or tonal complexes were combined with three different tone-to-noise ratios so that a given experimental session consisted of 12 different tone and noise combinations. The second series of experiments by Fishken consisted of three parts. In each part the tone-to-noise ratio and the frequency of the tonal complexes were held constant but the overall sound pressure level of the tone and noise was varied in 10-dB steps over a range of seven levels between 30 and 90 dB. The frequency evaluations for tonal complexes concern those measurements of a pair of 500-Hz and 2000-Hz tones added to a broadband noise. The results of both series are evaluated with respect to each of the following variables: a) frequency of tone, b) tone-to-noise ratio, c) overall sound pressure level of the tone and noise complex.

a) Frequency of Tone

The analysis of results by Wells (1969a) and by Pearsons and Bennett (1969) which were based on annoyance judgments showed that the presence of low-frequency tonal spikes produced smaller SDs than the presence of high-frequency spikes. A reevaluation of the first series of experiments by Fishken (1971) indicates that loudness judgments produce a similar outcome. Two sets of standard deviations were obtained, one that omitted the 500-Hz data and another that omitted the 4000-Hz data. These results are shown in Table B-10 together with the standard deviations previously calculated (see Table II, Scharf, et. al. 1977 and Appendix D, Table D-1, this report) for the entire group of 84 stimuli. To minimize the possible effect of sound pressure level on SDs, each value was obtained by first computing the SD at each level and then averaging the results across levels.

Table B-10 suggests that, unlike the results of the Wells (1969a) 400 series based on annoyance judgments, the SDs produced by the A-weighting and Zwicker's procedure do not depend on sound frequency when results are based on loudness judgments. Five procedures (D1, D2, E, Mark VI, and Perceived Noise Level), however, do appear sensitive to a high-frequency spike, i.e., the SDs are reduced when the 4000-Hz data are omitted suggesting that the presence of a 4000-Hz tone inflates the SDs. The opposite occurs for a tone at 500 Hz. With the exception of the A-weighting and Zwicker's procedure, the SDs produced by the remaining six procedures are larger when the 500-Hz tone is ommitted. The results at 500 Hz suggest, in agreement with the outcome for annoyance, that the presence of a 500-Hz tone decreases the overall standard deviations.

Table B-10

Fishken (First Experimental Series). Effect of Frequency of Tone on Standard Deviations (in Decibels)

		Frequency Weighting				Calculation Procedure			
	N	A	D1	D2	E	VI	VII	PNL	ZWI
SD Scharf, et.al., 1977 (Corrected) Table II	84	2.7	3.9	3.9	3.6	2.9	2.8	3.8	2.5
SD without 4000-Hz Tone	63	2.8	3.3	3.2	3.0	2.8	2.9	3.5	2.9
SD without 500-Hz Tone	63	2.8	4.4	4.4	4.0	3.2	3.2	4.3	2.7

The variability of loudness judgments produced by tone and noise does not show directly how the presence of a tone may alter the overall judgment of loudness. To answer this question, it is necessary to examine mean differences between predicted and measured loudness levels. Therefore, mean differences, computed for the same series of measurements that contributed to Table B-10, are shown in Table B-11.

According to Table B-11, the mean differences calculated by Zwicker's procedure are independent of frequency. Moreover, a tone at 500 Hz heard together with noise has very little effect on the calculated mean differences, whereas the removal of a 4000 Hz tone has a more noticeable effect. When the 4000 Hz tone is omitted, the mean differences approach zero more closely for the D1 and D2 frequency-weighting functions and for the Mark VI and Perceived Noise Level calculation procedures. Taken together, SDs and calculated mean differences show that, except for Zwicker's procedure, the descriptors predict results less well when a 4000 Hz tone is added to broadband noise than when a 500 Hz tone is added.

b) Tone-to-Noise Ratio

The available evidence (Little, 1961; Pearsons, et. al., 1968) suggests that, when single and multiple tones are introduced into bands of noise at tone-to-noise ratios of +15 dB and greater, the sounds become more annoying than the perceived level predicted by any frequency-weighting or calculation scheme. The same 84 stimuli of Fishken (1971) were regrouped to determine whether this effect of tone-to-noise ratio on annoyance also obtains for loudness. Table B-12 shows the effect of tone-to-noise ratio on calculated SDs and Table B-13 shows the effect on mean differences.

Table B-11

Fishken (First Experimental Series)
Effect of Frequency of Tone on Mean Differences
(Calculated Minus Observed Loudness Levels, in Decibels)

		Frequency Weighting				Calculation Procedure			
	N	A	D1	D2	E	VI	VII	PNL	ZWI
Mean Differences Scharf, et.al., 1977 (Table IV)	84	-4.8	+2.1	+2.0	+0.3	+4.9	-1.9	+5.5	+7.8
Mean Differences without 4000-Hz Tone	63	-5.2	+0.7	+0.6	-0.9	+4.1	-2.6	+4.2	+7.8
Mean Differences without 500-Hz Tone	63	-4.6	+2.5	+2.4	+0.5	+5.4	-1.8	+5.9	+7.8

Table B-12

Fishken (First Experimental Series)
Effect of Tone-to-Noise Ratio on Standard Deviations (in Decibels)

		Frequency Weighting				Calculation Procedure			
		A	D1	D2	E	VI	VII	PNL	ZWI
	N								
SDs Scharf, et.al.,1977, (Corrected) Table II	84	2.7	3.9	3.9	3.6	2.9	2.8	3.8	2.5
SDs T/N ratios of -5 and +5 dB	56	2.0	2.9	2.8	2.6	2.3	2.0	3.0	1.9
SDs T/N ratios of +15 dB	28	2.0	5.5	5.5	4.4	3.7	4.0	5.2	4.0

Table B-13

Fishken (First Experimental Series)
Effect of Tone-to-Noise Ratio on Mean Differences
(Calculated Minus Observed Loudness Levels, in Decibels)

		Frequency Weighting				Calculation Procedure			
	N	A	D1	D2	E	VI	VII	PNL	ZWI
Mean Diffs. Scharf, et.al., 1977, Table IV	84	-4.8	+2.1	+2.0	+0.3	+4.9	-1.9	+5.5	+7.8
Mean Diffs. T/N Ratios of -5 and +5 dB	56	-6.0	+1.3	+1.1	-0.9	+4.0	-2.5	+4.7	+7.8
Mean Diffs. T/N ratio of +15 dB	28	-2.1	+3.9	+3.9	+2.5	+6.5	-0.80	+6.9	+7.7

The overall SDs are reduced by about 1.0 dB by omitting those stimuli that produce tone-to-noise ratios of +15 dB. A tone-to-noise ratio of +15 dB inflates the SDs for all descriptors except the A-weighting.

In contrast to the data for annoyance, Table B-13 shows that a tone-to-noise ratio of +15 dB produces an overestimation of perceived loudness level (not an underestimation). Only Zwicker's procedure does not overestimate at +15 dB more than at lower tone-to-noise ratios.

c) Overall Sound Pressure Level of Tone and Noise Complex

A striking reduction in standard deviations is obtained by omitting those data for noises at 30 and 40 dB sound pressure level. With the exception of Mark VII and Zwicker's procedure, none of the weighting or calculation procedures was designed to assess loudness below 40 dB sound pressure level. Therefore, the use of these procedures for calculating the loudness of noises at low sound pressure levels is not entirely justified.

Table B-14 gives the mean SDs calculated on the basis of sounds at all sound pressure levels from 30 to 90 dB (from Table IV of Scharf, et. al., 1977) and on the basis of only those sounds between 50 and 90 dB. The SDs for the four frequency weightings go down dramatically from an average of 2.5 dB to 0.21 dB. The SDs for the four calculation procedures go down from 2.0 dB to 0.9 dB. Variability also went down quite a bit for Fishken's second experimental series when the two lowest levels were ommitted.

The overall sound pressure level of the tone and noise complex also modifies the difference between predicted and measured loudness levels. Table B-15 provides an example from the second series of measurements by Fishken for a constant tone-to-noise ratio of +15 dB. The discrepancy

Table B-14

Fishken: Standard Deviations (in Decibels) with and without Low Sound Pressure Levels. First Experimental Series

		Frequency Weighting				Calculation Procedure			
	N	A	D1	D2	E	VI	VII	PNL	ZWI
Mean SDs from Scharf, et.al., 1977 Table IV (30 to 90 dB SPL)	84	2.5	2.5	2.5	2.5	2.0	3.3	1.1	1.4
SD Means without 30 and 40 dB overall sound pressure level	60	.21	.21	.20	.21	1.1	1.2	0.35	1.0

Table B-15

Mean Differences (in Decibels) from Fishken's Second Experimental Series as a Function of Loudness Level (Tone-to-Noise Ratio +15 dB)

	Frequency Weighting---				Calculation Procedure---			
	A	D1	D2	E	VI	VII*	PNL	ZWI
OASPL								
90	+0.5	+7.0	+7.2	+3.4	+7.4	-0.8	+10.4	+8.0
80	-0.9	+5.5	+5.6	+2.0	+5.9	-2.3	+9.1	+8.1
70	-2.0	+4.4	+4.5	+0.9	+5.5	-3.0	+8.0	+8.0
60	-0.3	+6.0	+6.1	+2.6	+8.2	-0.4	+9.5	+10.1
50	+2.2	+8.5	+8.6	+5.2	+11.1	+3.1	+11.6	+12.1
40	+5.5	+11.8	+11.9	+8.4	+14.0	+6.5	+14.1	+14.2
30	+10.3	+16.6	+16.7	+13.2	+17.7	+9.6	+17.8	+17.0

*Note that Mark VII are unadjusted values.

between predicted and measured loudness levels decreases somewhat from 90 to 70 dB overall sound pressure level and then grows progressively larger as sound pressure level becomes smaller.

APPENDIX C

STEVENS'S TONE CORRECTION — 1970 PRELIMINARY PROPOSAL

Stevens's tone correction may be added to any of the descriptors examined in this report. However, the spectrum must be smoothed; that is, the tonal component or components removed, before the descriptor is calculated for the noise. Then the tone correction in decibels calculated according to Stevens's procedure is added to the descriptor's value. Since the correction worked poorly when used with Mark VII for which it was intended, it is not surprising that it fares no better with the seven other descriptors as shown in Tables C-1 and C-2 for SDs, and in Table C-3 for mean differences.

Table C-1

Standard Deviations (in Decibels) for 314 Spectra from 13 Studies with Tonal Components Listed in Column 1, Table VI. (SDs are Given for Each Descriptor with and without a Correction, Based on Preliminary Tonal-Correction Procedure of S.S. Stevens, Added to the Raw Descriptor Value. Means were not Weighted According to the Number of Contributing Values.)

	Frequency Weighting				Calculation Procedure			
	A	D1	D2	E	VI	VII	PNL	ZWI
Mean SD								
Uncorrected	2.6	2.4	2.4	2.3	2.1	2.1	2.4	2.3
Corrected	3.3	3.0	3.1	3.0	3.0	3.0	3.0	3.0
SD of SDs								
Uncorrected	1.4	1.2	1.3	1.2	1.0	1.1	1.2	1.4
Corrected	1.8	1.4	1.5	1.4	1.4	1.4	1.4	1.4

Table C-2

Standard Deviations (in Decibels) for 260 Spectra from 6 studies with and without Tonal Components Listed in Column 1, Table VII. (SDs are Given for Each Descriptor with and without a Correction Based on Preliminary Tonal-Correction Procedure of S.S. Stevens, Added to the Raw Descriptor Value. Means were not Weighted According to the Number of Contributing Values.) (Attribute Judged: Annoyance, Unacceptability, etc.)

	Frequency Weighting				Calculation Procedure			
	A	D1	D2	E	VI	VII	PNL	ZWI
Mean SD								
Uncorrected	2.5	2.0	2.1	1.9	1.9	1.9	2.1	2.8
Corrected	2.7	2.3	2.4	2.3	2.1	2.2	2.4	2.6
SD of SDs								
Uncorrected	1.2	0.8	0.9	0.8	0.8	0.9	0.9	1.5
Corrected	1.4	1.1	1.2	1.1	1.0	1.1	1.3	1.5

Table C-3

Mean Differences (in Decibels) (Calculated Minus Observed Levels)
Differences are Given for Levels Calculated Without and With
a Tonal Correction Proposed on a Preliminary Basis by
S.S. Stevens. Based Upon 141 Spectra from 6 Studies with
Tonal Components Listed in Table XII, Column 1

	Frequency weightings A	D1	D2	E	VI	VII	PNL	ZWI
Mean of Mean Differences								
Uncorrected	-7.7	-1.0	-1.5	-2.9	2.9	-4.6	4.1	7.1
Corrected	-5.6	1.6	1.3	-0.2	5.9	-1.3	6.9	10.7
SD of Means								
Uncorrected	4.7	4.9	5.0	4.6	4.0	4.0	4.8	3.3
Corrected	6.6	7.4	7.6	7.2	7.0	6.9	6.9	6.6

APPENDIX D

ERRATA AND ADDENDA TO SCHARF, ET. AL. (1977)

1. Errata

Several computational errors were noted in four Tables shown in Scharf, et. al. (1977). Although these corrections do not change the overall interpretation of results in that report, the revised Tables are included herein.

Table D-1 (Corrected Table II of Scharf, et. al.): A computational error was noted in line 3 based on some of Fishken's data. This correction produced a small change in the mean SDs and in the SD of SDs across the 11 descriptors.

Table D-2 (Corrected Table IV): Computational errors were noted in line 1, based on the Berglund, et. al. data, in line 2, based on some of Fishken's data, and in line 8, based on the data by Molino. These corrections produced small changes in the calculation of the mean of the mean differences and in the SD of the means for the C- and D-weightings and for Mark VII, PNL, and PNLC.

Table D-3 (Corrected Table V): The computational changes made in Table D-1 resulted in small changes in the values of the SDs in lines 1, 7, 9, and 13.

Table D-4 (Corrected Table VI): The computational changes made in Table D-2 resulted in small corrections to the values of mean differences in lines 1 and 9.

2. Addenda

A repeated-measures analysis of variance (ANOVA), treating studies like subjects, was performed on the data in Table II of Scharf, et. al. (1977)

Table D-1. Variability of Calculated Levels of Noise by Study (Standard Deviations in Decibels Computed Either from the Calculated Levels of a Group of Sounds Judged Subjectively Equal or from the Differences Between Calculated and Judged Levels. The Smaller the Standard Deviation, the Closer the Scheme Comes to Predicting the Subjective Equality of a set of Sounds)

STUDY	N/n	NUMBER OBSERVERS	A	B	C	D1	D2	E	MARK VI	MARK VII	PNL	PNLC	ZWI
Berglund, *et.al.*	18/3	30	4.6	4.6	4.6	4.6	4.6	4.6	3.8	3.9	5.6	5.6	3.7
Borsky	13/13*	319	3.6	3.0	2.8	3.3	3.5	3.3	3.0	3.0	3.8	4.2	3.4
Fisken	84/12*	12	2.7	2.9	3.0	3.9	3.9	3.6	2.9	2.8	3.8	3.6	2.5
	21/3*	8	4.5	4.6	4.6	4.4	4.4	4.4	4.4	5.4	3.4	3.5	3.7
Jahn	10/10	28	1.3	1.3	1.4	1.2	1.3	1.2	0.9	0.9	1.0	1.5	0.8
Kryter	17/17*	4-100	2.4	5.3	6.5	3.4	2.6	3.7	2.5	2.9	2.8	2.6	1.7
Kryter and Pearsons	9/9	13-19	3.5	4.8	5.4	2.8	3.1	2.8	2.1	1.9	2.1	2.2	3.7
Lubcke, *et. al.*	11/11	12	2.0	2.2	2.3	1.6	1.7	1.5	2.5	1.8	1.8	1.4	1.5
	20/20	12	2.3	2.1	2.2	2.6	2.8	2.6	2.1	2.0	2.2	2.3	1.6
Molino	30/5*	7	4.4	4.6	5.6	2.9	2.9	2.9	2.4	1.8	2.5	2.6	2.6
Pearsons and Bennett	30/30	20	4.3	4.5	4.7	3.5	3.7	3.3	2.8	2.8	2.9	2.2	3.7
	20/20	20	1.7	4.0	4.8	1.4	1.4	1.7	1.3	1.5	1.3	1.3	1.8
Pearsons, *et. al.*	103/54*	20	6.5	5.1	5.3	2.5	2.8	3.0	2.2	2.2	3.0	2.6	2.1
Pearsons and Wells	19/19*	20,20	2.8	3.4	3.6	1.8	1.8	1.9	2.4	2.3	2.5	2.7	2.6
Quietzsch	27/27	20	4.2	4.4	5.7	4.0	4.3	4.2	3.1	3.2	4.0	4.2	3.3
	10/10	20	3.8	6.3	7.0	3.3	2.9	3.8	2.5	2.5	2.6	2.8	2.5
Rademacher	24/24	20-25	2.2	2.6	3.2	1.8	2.0	1.9	1.6	1.7	1.6	1.7	1.6
Robinson and Bowsher	10/5*	558	1.9	2.8	3.1	1.4	1.5	1.9	1.2	1.6	1.1	1.4	0.9
Spiegel	20/20	10	4.7	6.2	6.8	4.2	4.0	4.2	2.4	1.9	3.2	3.7	2.4
	20/20	10	5.3	4.9	5.1	3.5	4.1	3.6	2.6	2.6	2.9	3.2	3.0
Wells (aircraft)	30/30	35	1.6	2.4	3.5	1.2	1.3	0.9	1.2	1.2	1.3	1.7	2.2
Wells (unpubl.)	33/33*	30	1.1	1.7	2.1	1.3	1.3	1.1	0.9	0.9	1.2	1.6	1.1
Wells 300	42/42	30	3.7	5.2	6.6	2.4	2.7	2.1	2.1	2.2	2.3	2.4	5.3
Wells 400	60/60	30	2.5	4.2	4.9	2.5	2.0	2.6	2.5	2.6	2.5	1.8	3.1
Wells UHV	25/25	31	1.5	0.9	1.4	1.3	1.5	1.3	1.1	1.0	1.3	1.4	0.9
Yaniv	11/11	10	1.6	2.2	4.2	2.2	2.3	1.6	---	2.6	4.6	4.9	2.0
	11/11	10	2.0	1.7	3.4	2.4	2.7	1.7	2.7	1.5	2.7	3.2	0.9
	11/11	10	2.6	1.2	2.8	2.9	3.3	2.1	1.7	1.4	2.7	3.1	1.4
Mean SD			3.05	3.55	4.16	2.65	2.73	2.63	2.26	2.22	2.60	2.69	2.36
Sd of SDs			1.4	1.6	1.6	1.1	1.1	1.1	0.8	1.0	1.1	1.1	1.1

LEGEND:

N	= number of conditions (e.g. different sound pressure levels, instructions, tone-to-noise ratios)
n	= number of different spectra
*	= standard deviation based on average of two or more distinct sets of measurements
A, B, C	= standard sound-level meter weightings
D1	= meter weighting adopted by IEC
D2	= weighting values suggested by K. Kryter
E	= weighting values proposed for trial and study by ANSI
Mark VI	= ANSI S 3.4 (R1972) procedure for the computation of noise
Mark VII	= based on modification of Mark VI (S. S. Stevens, *JASA*, 1972, *51*
PNL	= perceived noise level
PNLC	= PNL with tone correction as per FAR 36
ZWI	= based on Zwicker's loudness calculation system. Program from E. Paulus and E. Zwicker, Acustica, 1972, 27. Free-field (FF) and diffuse-field (DF) values used as appropriate. For earphone listening, FF values used.

Table D-2

MEAN DIFFERENCES (in decibels) (CALCULATED MINUS OBSERVED LEVELS)
See Legend for Table II.

STUDY	N/n	A	B	C	D1	D2	E	MARK VI	MARK VII	PNL	PNLC	ZWI
Berglund, et al.	18/3*	-12.9	-4.7	0.4	-4.1	-6.7	-6.3	2.5	-6.0	1.9	2.6	8.6
Fishken	84/12*	-4.8	-5.1	-5.1	2.1	2.0	0.3	4.9	-1.9	5.5	10.7	7.8
	21/3*	-1.0	-1.7	-1.7	5.8	6.0	2.9	8.8	1.3	9.9	15.6	11.2
Jahn	10/10	-11.9	-10.8	-10.3	-5.1	-5.3	-7.7	-0.3	-7.8	1.1	2.3	5.1
Kryter and Pearsons	9/9	-8.9	-7.6	-7.0	-2.3	-2.2	-3.7	0.3	-7.3	2.1	5.3	4.1
Lubcke, et al	11/11	-18.8	-16.9	-16.0	-13.0	-13.3	-14.8	-8.6	-13.3	-9.4	-8.2	-2.4
	20/20	-17.3	-15.7	-14.9	-11.7	-11.8	-13.3	-6.2	-14.0	-5.2	-3.7	-0.7
Molino	30/5*	-6.5	-4.8	-3.1	-0.8	-1.0	-2.4	6.4	-1.0	5.1	6.3	12.3
Quietzsch	27/27	-14.6	-13.0	-11.6	-8.3	-8.6	-10.3	-3.5	-11.0	-2.5	0.5	1.8
	10/10	-13.0	-9.4	-7.8	-6.9	-7.5	-7.9	-1.4	-7.6	-2.9	-1.0	4.7
Rademacher	24/24	-8.8	-4.2	-2.4	-2.1	-3.0	-3.7	1.7	-5.3	3.9	6.3	8.0
Spiegel	20/20	-12.8	-10.9	-10.0	-6.7	-7.0	-7.6	-1.5	-9.4	-3.9	-1.8	1.0
	20/20	-11.9	-10.0	-9.0	-5.8	-6.2	-6.8	-2.8	-10.9	-2.3	1.0	0.7
Yaniv	11/11	-7.3	-3.9	-1.3	-1.7	-2.4	-3.6	-	-4.3	-1.5	-0.1	6.3
	11/11	-10.3	-6.9	-4.3	-4.7	-5.4	-6.6	0.2	-4.9	-0.9	0.5	6.5
	11/11	-11.8	-8.4	-5.8	-6.2	-6.8	-8.1	0.9	-6.3	-0.9	0.4	6.2
Mean of Mean diffs.		-10.8	-8.4	-6.9	-4.5	-5.0	-6.2	-0.13	-6.9	-0.0	2.3	5.1
SD of Means		4.53	4.36	4.85	4.74	4.84	4.55	4.54	4.27	4.71	5.66	4.16

Table D-3

EFFECT ON STANDARD DEVIATION OF FOUR PARAMETERS (Standard Deviations in Decibels)
See Legend for Table II.

VARIABLE	No. of STUDIES/ SDs	A	B	C	D1	D2	E	VI	VII	PNL	PNLC	ZWI
1. Attribute Judged												
Loudness	9/15	3.2	3.5	4.1	3.0	3.1	2.9	2.5*	2.4	3.0	3.2	2.2
Acceptability	10/12	2.9	3.7	4.3	2.3	2.3	2.3	2.0	2.0	2.3	2.2	2.6
2. Type of Noise												
Aircraft	7/8	2.0	3.0	3.5	1.9	1.8	2.0	1.5	1.6	1.9	2.0	1.6
Industrial	3/4	2.7	2.7	2.8	2.7	2.8	2.7	2.5	2.4	2.9	1.7	2.1
Vehicle	1/1	2.2	2.6	3.2	1.8	2.0	1.9	1.6	1.7	1.6	1.7	1.6
Household	1/3	2.1	1.8	3.5	2.5	2.8	1.8	2.2*	1.8	3.3	3.7	1.4
Artificial	7/10	4.1	4.6	5.0	3.2	3.3	3.2	2.6	2.7	2.9	2.8	3.2
Miscel.	3/4	3.5	4.1	4.9	2.9	2.9	3.1	2.3	2.1	2.6	2.8	2.3
3. Tonal Components												
Present	9/12	3.0	3.5	3.9	2.5	2.5	2.4	2.2	2.3	2.4	2.4	2.7
Absent	10/15	3.2	3.7	4.5	2.9	3.0	2.9	2.4*	2.2	2.8	3.0	2.2
4. Mode of Sound Presentation												
Free Field	11/14	2.7	3.1	3.7	2.1	2.2	2.1	1.9	1.9	2.1	2.1	2.3
Diffuse Field	7/8	3.7	4.7	5.3	3.1	3.1	3.3	2.3	2.3	2.6	2.8	2.5
Earphones	3/6	3.0	2.9	3.8	3.4	3.5	3.0	3.1*	2.9	3.8	4.0	2.4

* 1 SD less

Table D-4

EFFECT ON MEAN DIFFERENCES OF TWO PARAMETERS (Calculated minus observed levels in decibels)

See Legend for Table II.

VARIABLE	No. of STUDIES/ MEANS	A	B	C	D1	D2	E	VI	VII	PNL	PNLC	ZWI
1. Type of Noise												
Aircraft	1/1	-12.8	-10.2	-8.7	-5.1	-5.5	-7.3	0.9	-6.6	1.4	-	6.1
Industrial	3/4	-15.0	-11.6	-9.6	-8.4	-9.4	-10.5	-3.1	-10.2	-3.6	-3.2	2.7
Vehicle	1/1	-8.8	-4.2	-2.4	-2.1	-3.0	-3.7	1.7	-5.3	3.9	6.3	8.0
Household	1/3	-9.8	-6.4	-3.8	-4.2	-4.9	-6.1	-0.6**	-5.2	-1.1	0.3	6.3
Artificial	3/5	-7.9	-7.1	-6.6	-1.4	-1.5	-3.0	1.9	-5.6	2.3	5.8	9.5
Miscel.	2/3	-11.9	-9.1	-7.5	-5.3	-5.7	-6.9	0.5	-6.5	-0.1	2.0	6.3
2. Mode of Stimulus Presentation												
Free Field	4/5	-14.3	-12.1	-11.0	-8.0	-8.4	-10.0	-3.4	-10.3	-2.4	-0.6	2.4
Diffuse Field	4/5	-10.6	-8.5	-7.4	-4.5	-4.8	-5.7	0.2	-7.2	-0.4	2.0	4.6
Earphones	3/6	-8.0	-5.1	-3.0	-1.5	-2.2	-3.6	3.5	-3.7	2.3	5.0	7.8

** 2 means

(after being corrected as explained above). Table II gave the SDs for 28 sets of spectra for eleven descriptors (six sound-level meter frequency weightings and five calculation procedures). The results of the ANOVA are given in Table D-5. Although the differences among the mean SDs for the eleven descriptors were small, they were highly significant, as were the differences among studies and subsets. However, the interaction between procedure (descriptor) and study was not significant.

To determine which mean SDs differed from each other significantly, a Duncan's multiple-range test (Lynch and Huntsberger, 1976) was performed on the matrix of differences between descriptors given in Table D-6. The number of asterisks indicates the level of significance. Generally, differences greater than 0.45 dB were significantly different at the .05 level or better. Thus the A-weighting had significantly larger SDs than four of the five calculation procedures. With the exclusion of B- and C-, among the four frequency weightings only A- and D1-weightings differed significantly. Except for PNLC, none of the calculation procedures differed significantly from one another. (N.B. Table D-6 supercedes Table VII in Scharf, et. al. (1977). Table VII was based on t-tests and was presented as a preliminary analysis pending an ANOVA and a more appropriate multiple-range test.

Table D-5

REPEATED MEASURES ANOVA

(Based on 28 standard deviations from 20 studies.)

Source Variance	Sum of Squares	Degrees of Freedom	Mean	F	P
Weighting or calculation procedure	95.28	10	9.53	17.08	<<.001
Study	274.04	27	10.14	18.20	<<.001
Procedure x Study	150.56	270	.56		
Total	530.14	307			

Table D-6

DIFFERENCES [1] IN DECIBELS BETWEEN MEAN STANDARD DEVIATIONS IN TABLE II.

	B	C	D1	D2	E	VI	VII	PNL	PNLC	ZWI
A	.50*	1.11***	-.40*	-.32	-.42	-.79***	-.83***	-.45*	-.36	-.69
B		.61**	-.90***	-.82**	-.92***	-1.29***	-1.33***	-.95***	-.86***	-1.19***
C			-1.5.***	-1.43***	-1.53***	-1.90***	-1.94***	-1.56***	-1.47***	-1.80***
D1				.08	-.02	-.39	-.43	-.05	.04	-.29
D2	Results of Duncan's Multiple Range Test				-.10	-.47*	-.51*	-.13	-.04	-.37
E	N=28					-.37	-.41	-.03	.06	-.27
VI	blank = Not Significant						-.04	.34	.43	.10
VII	* = Significant at .05 or better							.38	.47*	.14
PNL	** = Significant at .01								.09	-.24
PNLC	*** = Significant at .001									-.33

[1]Standard deviation for a given calculation scheme listed in the column of this matrix is subtracted from the deviation for the calculation scheme, with which it is paired, listed in the row. Thus B minus A =.50, D1 minus A =-.40, etc.

Legend:

A, B, C	standard sound-level meter weightings
D1	meter weighting adopted by IEC
D2	weighting values suggested by K Kryter
E	weighting values proposed for trial and study by ANSI
Mark VI	ANSI S3.4 (R1972) procedure for the computation of the loudness of noise
Mark VII	based on modification of Mark VI (S.S. Stevens, JASA, 1972, 51)
PNL	perceived noise level
PNLC	PNL with tone correction as per FAR 36
ZWI	based on Zwicker's loudness calculation system. Program from E. Paulus and E. Zwicker, Acustica, 1972, 27. Free-field (FF) and diffuse-field (DF) values used as appropriate

REFERENCES

ANSI S 3.4 (R 1972). USA Standard Procedure for the computation of loudness of noise.

Berglund, B., Berglund, U., and Lindvall, T. Scaling loudness, noisiness, and annoyance of aircraft noise. J. Acoust. Soc. Am. 1975, 57, 930-934.

Berglund, B., Berglund, U., and Lindvall, T. Scaling loudness, noisiness and annoyance of community noises. J. Acoust. Soc. Am. 1976, 60, 1119-1125.

Bishop, D. E. Judgments of the relative and absolute acceptability of aircraft noise. J. Acoust. Soc. Am. 1966, 40, 108-122.

Borsky, P. Annoyance and acceptability judgments of noise produced by three types of aircraft by residents living near JFK Airport. NASA report, 1974, 1-50.

Copeland, W.C.T. A controlled experiment on the subjective effects of jet engine noise. Journal of the Royal Aeronautical Society, 1960, 64, 33-36.

FAR 36, Federal Aviation Administration, 1969. Aircraft Certification Procedure, Federal Aviation Regulation Part 36.

Fishken, D. Loudness summation between tones and noise. Unpublished doctoral dissertation. Northeastern University, Boston, Massachusetts, 1971.

Galanter, E. The annoyance of sound and of other life events. J. Acoust. Soc. Am. 1978, 63, (S1) S17.

Goulet, P. and Northwood, T. D. Subjective rating of broad-band noises containing pure tones. Paper presented at the 84th Meeting of the Acoustical Society of America, Miami Beach, Florida. December 1, 1972.

Hargest, T. J. and Pinker, R. A. The influence of added narrow band noises and tones on the subjective response to shaped "white" noise. Journal of the Royal Aeronautical Society, 1967, 71, 428-430.

Hiramatsu, K., Takagi, K., and Yamamoto, T. The effect of sound duration on annoyance. J. Acoust. Soc. Japan, 1976, 32, 739-750.

International Electrotechnical Commission. Frequency weighting for the measurement of aircraft noise (D-weighting). Document 29C (Central Office) 26, 1975.

Jahn, M. Subjektive and objective berwertung von Maschinengerauschen. Acustica, 1965/1966, 16, 175-185.

Kryter, K. Scaling human reactions to the sound from aircraft. J. Acoust. Soc. Am. 1959, 31, 1415-1429.

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

Kryter, K. D., and Pearsons, K. S. Some effects of spectral content and duration on perceived noise level. J. Acoust. Soc. Am., 1963, 35, 866-883.

Kryter, K. D., and Pearsons, K. S. Judged noisiness of a band of random noise containing an audible pure tone. J. Acoust. Soc. Am., 1965, 38, 106-112.

Little, J. W. Human response to jet engine noises. Noise Control, 1961, 7, 11-13.

Little, J. W. and Mabry, J. E. Empirical comparisons of calculation procedures for estimating annoyance of jet aircraft flyovers. J. Sound Vib., 1969, 10, 59-68.

Lübcke, E., Mittag, G., and Port, E. Subjective und objective Bewertung von Maschinengerauschen. Acustica, 1964, 14, 105-114.

Lynch, M. D., and Huntsberger, D. V. Elements of Statistical Inference for Education and Psychology. Allyn and Bacon, Inc.: Boston, 1976.

Molino, J. National Bureau of Standards. Unpublished Study, 1976.

Niese, H. Beitrag zur Relation zwischen Lautstarke und Lastigkeit von Gerauschen. Acustica, 1965, 15, 236-243.

Ollerhead, J. B. An evaluation of methods for scaling aircraft noise perception. NASA CR-1883, 1971.

Ollerhead, J. B. Scaling aircraft noise perception. J. Sound Vib., 1973, 26(3), 361-388.

Paulus, E. and Zwicker, E. Program zur automatischen Bestimmung der Lautheit aus Terzpegeln oder Frequenzgruppenpegeln. Acustica, 1972, 27, 253-266.

Pearsons, K. S. Assessment of the validity of pure tone corrections to perceived noise level. Conference on progress of NASA research relating to noise alleviation of large subsonic aircraft, NASA SP-189, 1968.

Pearsons, K. S. and Bennett, R. L. The effects of temporal and spectral combinations on the judged noisiness of aircraft sounds. FAA report, 1969.

Pearsons, K. S. and Bennett, R. L. Effects of temporal and spectral combinations on the judged noisiness of aircraft sounds. J. Acoust. Soc. Am., 1971, 49, 1076-1082.

Pearsons, K. S. Bishop, D. E., and Horonjeff, R. D. Judged noisiness of modulated and multiple tones in broad-band noise. J. Acoust. Soc. Am., 1969, 45, 742-750. Also NASA report, 1968.

Pearsons, K. S. and Wells, R. J. Comparison of objective measures of noisiness for pure tone and varying bandwidth noise. Paper presented at the 6th International Congress on Acoustics, Tokyo, Japan, August 21-28, 1968.

Pearsons, K. S. and Wells, J. R. Judged noisiness of sounds containing multiple pure tones. Paper presented at the 78th Meeting of the Acoustical Society of America, San Diego, California, November 4-7, 1969.

Quietzsch, G. Objektive und subjektive Lautstarkemessungen. Acustica, 1955, Akust. Beih. 1, 49-66.

Rademacher, H. Die Lautstarke von Kraftfahrzeuggerauschen. Acustica, 1959, 9, 91-108.

Robinson, D. W. and Bowsher, J. M. A subjective experiment with helicopter noises. J. of the Royal Aeronautical Society, 1961, 65, 635-637.

Rule, S. J. Effect of instructional set on responses to complex sounds. J. Exp. Psych. 1964, 67, 215-220.

Rule, S. J. and Little, J. W. The design and construction of an annoyance scale for jet engine noise. Technical report. Commercial Airplane Division, The Boeing Company, Seattle, Washington, January 23, 1963.

Scharf, B., Hellman, R. and Bauer, J. Comparison of various methods for predicting the loudness and acceptability of noise. U.S. Environmental Protection Agency Report. 550/9-77-101, August 1977.

Scharf, B. and Horton, T. J. Scaling loudness and annoyance as a function of duration. J. Acoust. Soc. Am. 1978, 63 (S1) S16.

Spiegel, M. Prüfung verschiedener Lautstarkeberechnungsmethoden bei diffuser Beschallung. Unpublished, 1960.

Stevens, S. S. Procedure for calculating loudness: Mark VI, J. Acoust. Soc. Am. 1961, 33, 1577-1585.

Stevens, S. S. Perceived level of noise plus tones. Unpublished report for the Laboratory of Psychophysics, Harvard University, 1970.

Stevens, S. S. Perceived level of noise by Mark VII and decibels (E). J. Acoust. Soc. Am., 1972, 51, 575-601.

Wells, R. J. Recent research relative to perceived noise level. Paper presented at the 74th Meeting of the Acoustical Society of America, Miami Beach, Florida, November 14-17, 1967.

Wells, R. J. A new method for computing the annoyance of steady state noise versus perceived noise level and other subjective measures. Paper presented at the 77th Meeting of the Acoustical Society of America, Philadelphia, Pennsylvania, April 8-11, 1969a.

Wells, R. J. Calculation of annoyance level for sounds containing multiple pure tones. Paper presented at the 78th meeting of the Acoustical Society of America, San Diego, California, November 4-7, 1969b.

Wells, R. J. Jury ratings of complex aircraft noise spectra versus calculated ratings. Paper presented at the 80th Meeting of the Acoustical Society of America, Houston, Texas, November 3-6, 1970.

Wells, R. J. Unpublished (about 1970).

Wells, R. J. A subjective study of ultra high voltage transmission line noise. Paper presented at the Second Arden House Workshop on Noise Control Engineering, January, 1972.

Winer, B. J. Statistical Principles in Experimental Design, McGraw-Hill: New York, 1962.

Yaniv, S. Equal loudness contours for household appliances. National Bureau of Standards. Unpublished Study, 1976.

Zwicker, E. Uber psychologische und methodische Grundlagen der Lautheit. Acustica, 1958, 8 (Beiheft 1), 237-258.

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA 550/9-79-102	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE Part II: Comparison of Various Methods for Predicting the Loudness and Effects of Spectral Pattern Acceptability of Noise and Tonal Components		5. REPORT DATE Nov 1979
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) B. Scharf and R. Hellman		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Auditory Perception Laboratory, Northeastern University, Boston, Massachusetts 02115		10. PROGRAM ELEMENT NO.
		11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS U.S. Environmental Protection Agency, Office of Noise Abatement and Control (ANR-471), Washington, D.C. 20460		13. TYPE OF REPORT AND PERIOD COVERED
		14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

The present report is a continuation of an earlier report by Scharf, Hellman and Bauer (1977). The objectives are (1) to determine whether subjective judgments of particular types of noise, categorized by spectral shape, are better approximated by some descriptors (frequency weightings and calculation procedures) than by others, and (2) to investigate the role of tonal components in these studies and to assess the adequacy of several tone-correction procedures. The analysis of data by spectral shape produced a mixed outcome. Results showed that no overall advantage would accrue from regrouping sets of data across studies on the basis of similar spectral shapes. However, although variability was not reduced when considered across nine spectral categories, the interaction between spectral shape and descriptor was highly significant ($p < .001$). The examination of over 500 spectra with and without tonal components provided only tentative support for the trends noted in the literature. When the judged attribute is either loudness or noisiness, tonal components do not seem to add to the subjective magnitude of broad-band noise below 80 dB sound pressure level. At higher levels, according to one large-scale study, tonal components seemed to add the equivalent of 2 dB to the judged noisiness. No data could be located that would permit adequate assessment of the contribution of tonal components to the "absolute" magnitude.......(continued on page 2, attached)

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Loudness, Spectral Patterns, Tonal Components, Calculation Procedures, Frequency Weightings		

18. DISTRIBUTION STATEMENT Available at EPA/ONAC and Research Triangle Park, North Carolina	19. SECURITY CLASS *(This Report)* Unclassified	21. NO. OF PAGES
	20. SECURITY CLASS *(This page)* Unclassified	22. PRICE

EPA Form 2220-1 (9-73)

...of judged annoyance or unacceptability (as distinct from noisiness or loudness). Given the small effect of tonal components in the present group of studies, the evaluation of three different tone-correction procedures (FAR 36, 1969; Kryter and Pearson's, 1965; and Steven's, 1970) could not lead to definitive conclusions about their relative merits. Although a small correction may be necessary for the presence of tonal components at high levels, the tone-correction procedures now available cannot be properly evaluated until more appropriate data that demonstrate the need for a tone correction are obtained.

United States
Environmental Protection
Agency
Washington DC 20460
ANR-471

Official Business
Penalty for Private Use $300

Postage and Fees Paid
Environmental Protection Agency
EPA-335

Special
Fourth-Class
Rate
Book