

Calculation of the Loudness of Loudspeakers during Listening Tests*

RONALD M. AARTS

Philips Research Laboratories, 5600 JA, Eindhoven, The Netherlands

A computer program is used to calculate loudness levels in order to determine the reproduction levels of loudspeakers during listening tests. The calculation results are compared with the results of listening tests. Pink noise and music, reproduced by various loudspeakers, are used as the stimuli. Subjective listening tests show that the A-weighting can result in misleading conclusions. The ISO loudness models appear to be the best technique for adjusting the interloudness levels of loudspeakers during listening tests.

0 INTRODUCTION

The purpose of this paper is twofold. First, after an introduction to loudness and the history of the scaling of loudness, a brief description is given of the calculation of loudness of stationary signals, written according to ISO standards [1]. A comparison is made between the program results, the traditional A-weighted sum using several noise spectra, and the subjective loudness ratings known from the literature.

Second, an application is discussed concerning the objective measurement of reproduction levels of loudspeakers during listening tests. It is considered to be very important that during listening tests the reproduction levels of the loudspeakers being tested be the same compared with the others; otherwise the test results can be seriously biased. To validate the calculation results, they are compared with several subjective ratings obtained from listening tests, using a varied repertoire and a variety of loudspeakers. From this the conclusion is drawn that the A-weighting method is in general too simple and may produce misleading loudspeaker quality assessments.

0.1 Definitions

For the purpose of this paper, the following definitions are used [2].

Loudness: Attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from soft to loud.

Loudness level: Of a given sound, the sound pressure level of a reference sound, consisting of a sinusoidal plane progressive wave of frequency 1 kHz coming from directly in front of the listener, which is judged by otologically normal persons to be equally loud to the given sound. Loudness level is expressed in phons.

Critical bandwidth: The widest frequency band within which the loudness of a band of continuously distributed random noise of constant-band sound pressure level — (SPL) is independent of its bandwidth.

0.2 Scaling of Loudness

Several experimenters have made contributions to the scaling of loudness. The earliest published work seems to be that credited to Richardson and Ross [3], who required an observer to rate one of two tones of different intensities which he heard as a certain multiple or fraction of the other.

Over the past five decades various methods of evaluating the loudness of complex sounds from objective spectrum analysis have been proposed. The earliest attempt to use algebraic models in psychophysical measurement is probably that of Fletcher and Munson [4]. However, there is still much interest in this subject, and nowadays two procedures are used for calculating loudness levels; both are standardized by the ISO. The first is based on a method developed by Stevens, hereafter referred to as method 532A, and the second one, 532B, on a method by Zwicker [5], [6]. A-weighting is a widely used method, traditionally applied in sound level meters to measure the loudness of signals or to determine the annoyance of noise. It is based on an early 40-phon contour and is a rough approximation of a frequency weighting of the human auditory system.

* Presented at the 88th Convention of the Audio Engineering Society, Montreux, Switzerland, 1990 March 13–16. Manuscript received 1990 February 6; revised 1990 June 6.

However, considerable differences are ascertained between subjective loudness ratings and the A-weighted measurements. A drawback using A-weighting is the way in which the dBA level remains constant when the bandwidth and the loudness of noise increase together. As depicted in Fig. 1, with increasing bandwidth the loudness has increased from 60 to 75 phons (solid curve), but the dBA level (dashed line) has remained constant. The effect has been studied by Brittain [7] and is a striking example that for wideband signals the A-weighted method is generally too simple.

1 CALCULATION OF LOUDNESS

A computer program has been written according to both Stevens' and Zwicker's methods as specified in ISO 532 [1]. The input for both programs is a file containing SPL values measured in one-third-octave bands. The essential parts of both methods is briefly discussed in the following sections.

1.1 Sone Scale

The loudness level is expressed in phons. However, loudness values expressed on this scale do not immediately suggest the actual magnitude of the sensation. Therefore the sone scale, which is a numerical assignment of the strength of a sound, has been established. It has been obtained by subjective magnitude estimation by observers with normal hearing.

As a result of numerous experiments [8], the following equation has evolved to calculate the loudness S of a 1-kHz tone in sones:

$$S = 0.01(p - p_0)^{0.6} \quad (1)$$

where $p_0 = 45 \mu\text{Pa}$ approximates the effective threshold and p is the sound pressure in micropascals. For values $p \gg p_0$, Eq. (1) can be approximated by the well-known equation

$$S = 2^{(P-40)/10} \quad (2a)$$

or

$$P = 40 + 10 \log_2 S \quad (2b)$$

where P is the loudness in phons, leading to the definition of ISO 532: "One sone is the loudness of a sound whose loudness level is 40 phons." It is important to discriminate between the loudness levels that have been computed and those obtained by subjective measurements. In the following, the computed methods will be labeled by the suffix (OD) for Stevens' method, and by (GD) or (GF) for Zwicker's method, where O stands for measurements being made in octave bands, G for group (Frequenzgruppen, or critical bands); D stands for diffuse and F for free field, depending on the measurement environment of the sound source. Method 532A (Stevens) assumes that the measurements are taken in a diffuse sound field, while method 532B (Zwicker)

allows diffuse- and free-field conditions to be set as parameters in the program.

1.2 Stevens' Method (532A)

The method is equal to the Mark VI version as described in [9]. However, Stevens refined the method resulting in the Mark VII version [10] which is not standardized. In the following, method 532A is discussed briefly. The SPL of each one-third octave is converted to a loudness index using a table based on subjective measurements. The total loudness in sones(OD) S_t is calculated by means of the equation

$$S_t = S_m + F(\sum S_i - S_m) \quad (3)$$

where S_m is the greatest of the loudness indices and $\sum S_i$ is the sum of the loudness indices of all the bands. The value of F is 0.15 for one-third-octave bands, 0.2 for one-half-octave bands, and 0.3 for octave bands. The total loudness may be converted into loudness level in phons(OD) using Eq. (2b).

1.3 Zwicker's Method (532B)

An early version is described in [5], later it is refined (see, for example, [11]–[13]). In the following the essential steps in the procedure are discussed. The spectrum measured in one-third octaves is converted into frequency band having a bandwidth approximating the critical bands of the human ear. This warping of the frequency axis into the critical band rate, with units of bark, is depicted in Fig. 2. The one-third octaves up to 90 Hz are assembled as the first critical band, the next three (90–180 Hz) as the second band, and the bands ranging from 180 up to 280 Hz as the third critical band, leading to the critical band levels L_{c1} , L_{c2} , and L_{c3} . The rule of combination may be understood from the example.

$$L_{c2} = 10_{10} \log(10^{L_{100}/10} + 10^{L_{125}/10} + 10^{L_{160}/10}) \quad (4)$$

where L_{125} etc., is the measured one-third-octave band SPL for the band with center frequency 125 Hz, and L_{c2} is the SPL in the second critical band. The relation

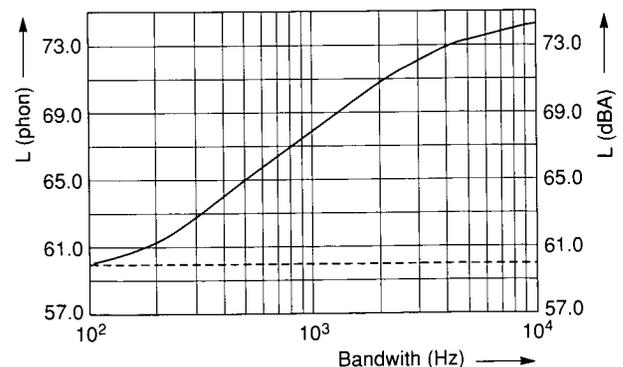


Fig. 1. Loudness (solid curve) and dBA level (dashed line) versus bandwidth of white noise.

between frequency f_i and the third number i is given by

$$f_i = 10^{i/10} \quad (5)$$

Each critical band is subdivided into bands 0.1 bark wide. The SPL in each critical band is converted, by means of a table, into a loudness index for each of its subbands. In order to incorporate masking effects, contributions are also made to higher bands. As an example, the transformation is plotted in Fig. 3 for a 1-kHz tone at 90 dB SPL. Clearly, there is a 'tail' after the actual excitation. When there are more tones in the spectrum it is possible that there is an overlap between the various tails. However, only the highest loudness index in each 0.1-bark band makes a contribution to the overall result, and as a consequence the weaker tones are masked. The total loudness is finally calculated by integrating the loudness indices over all the subbands, resulting in the loudness in sones(GF) or sones(GD).

2 LOUDNESS PROGRAM

The program was written on a PC. However, it can easily be ported to another system, preferably the same system as that used to control the real-time analyzer required for the spectrum measurements. To gain insight into the methods used, some standard noise spectra will be used and a comparison is made between program results, the traditional A-weighted sum, and the subjective loudness ratings known from the literature. The

results are summarized in Table 1. Table 1 has the following entries:

- Type of noise spectrum
 - Noise level in the one-third octave of 400 Hz
 - Total SPL of spectrum, that is, value obtained by simply adding the sound pressure in all one-third octaves
 - Total A-weighted level of noise spectrum, that is, value obtained by using the A-weighting formula [14]
 - Judged loudness level of noise spectrum obtained by comparing noise sources with a narrow-band noise at 1 kHz (these values are taken from [16, table III])
 - Loudness in phons(OD) (method 532A) calculated by assuming diffuse-field measurement conditions
 - Loudness in phons(GD) (method 532B) calculated by assuming diffuse-field measurement conditions
 - Figure numbers referring to corresponding spectra.
- Table 1 contains the following types of noise: M18, K19, M4, and K11 are spectra measured from engines by Lübcke et al. [15]. They recorded 19 different types of noise for subjective loudness ratings. These experiments were repeated and extended by Jahn [16]. The noise sources were all band-limited from 90 to 11 kHz and scaled to a noise power of 74 dB.
- Sine 1k (Fig. 3) is the pure tone of a frequency of 1 kHz at 90 dB SPL.
- Pink noise [Fig. 4(a)] for several power levels.
- White noise [Fig 5(a)] for several power levels.
- Red noise [Fig 6(a)] is noise that falls 6 dB per octave.

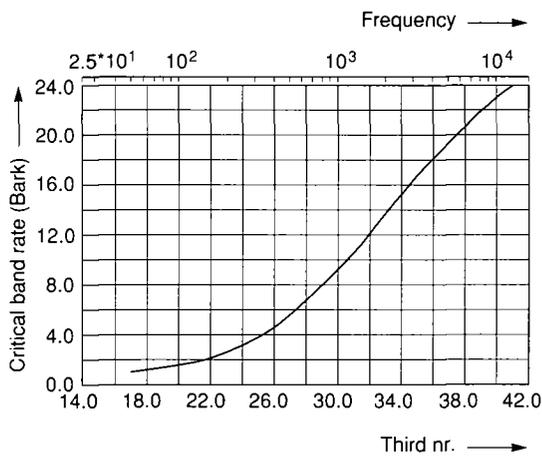


Fig. 2. Relationship between frequency, one-third-octave number, and critical band rate.

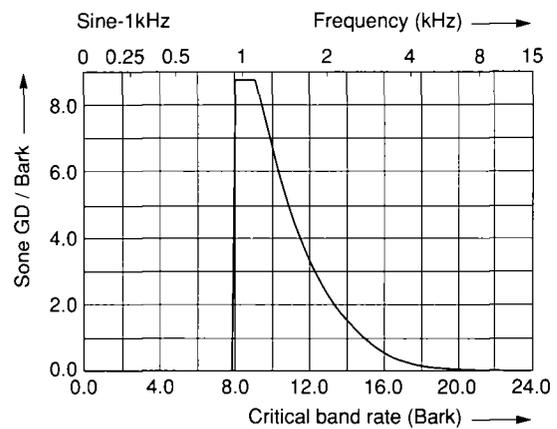


Fig. 3. Transformation from SPL into loudness for 1-kHz tone at 90 dB.

Table 1.

Noise	400-Hz level (dB)	SPL (dB)	Weighted level (dBA)	Judged level (dB)	Stevens [Phons(OD)]	Zwicker [Phons(GD)]	Fig.
M18	65.5	74.0	69.8	81.5	81.0	87.7	—
K19	59.5	74.0	73.5	85.0	85.0	90.7	—
M4	59.0	74.0	71.4	83.5	83.2	88.4	—
K11	55.0	74.0	70.7	84.0	83.0	89.4	—
Sine 1k	—	90.0	90.0	—	90.4	91.0	3
Pink	96.0	110.2	107.9	—	121.8	122.1	4
White	60.0	82.9	81.2	—	94.3	96.9	5
Red	62.0	84.3	67.4	—	81.2	86.4	6

Parts (a) of Figs. 4–6 represent the spectra of the various noise sources measured with a one-third-octave analyzer in a diffuse sound field. They serve as inputs for the actual loudness calculations. The corresponding parts (b) of these figures represent the loudness versus critical band rate according method 532B, assuming a

diffuse sound field, parts (c) show the loudness according method 532A.

2.1 Comparison of Methods 532A and 532B

At first glance there is quite a difference when Figs. 4(b)–6(b) are compared with their counterparts, Figs.

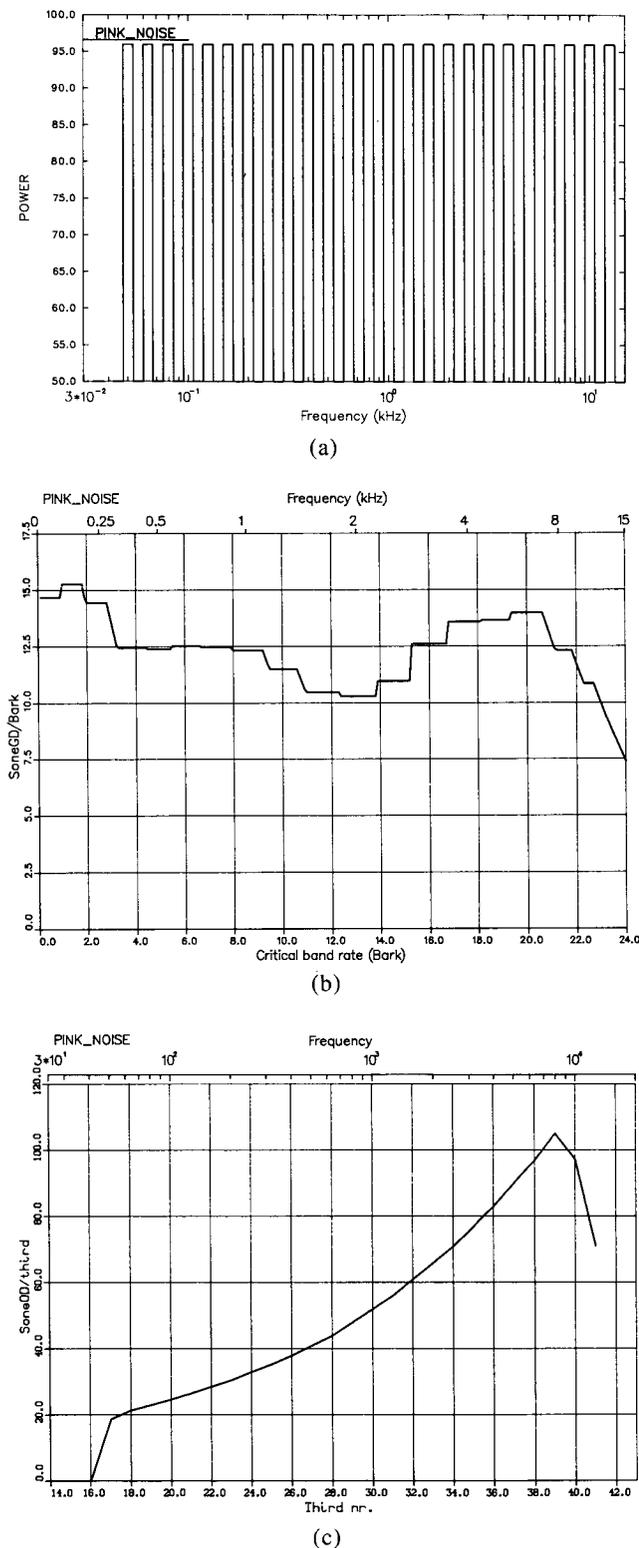


Fig. 4. (a) Pink noise spectrum at 96 dB SPL. (b) Pink noise loudness spectrum according 532B [sones(GD)/bark]. (c) Pink noise loudness spectrum according 532A [sones(OD)/one-third octave].

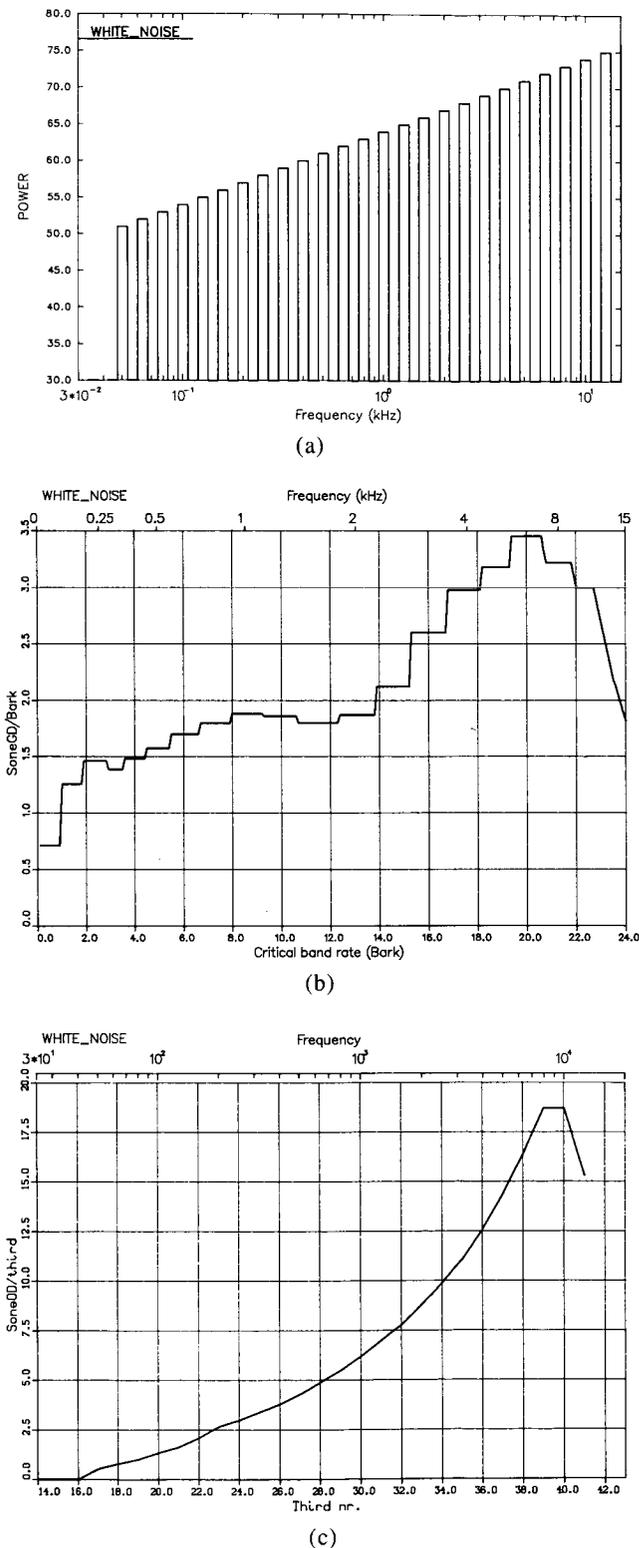


Fig. 5. (a) White noise spectrum at 81.2 dBA level. (b) White noise loudness spectrum according 532B [sones(GD)/bark]. (c) White noise loudness spectrum according 532A [sones(OD)/one-third octave].

4(c)–6(c), but one has to bear in mind that the frequency axes are different (as discussed in Sec. 1.3). Another apparent difference is the low-frequency gain for 532B, which may be due to the addition of the first three groups of thirds into critical bands, as discussed in Sec. 1.3. Finally there remain discrepancies due to the

different approaches of the two methods. Zwicker's method is elegant because of its compatibility with the accepted methods of the human ear, but Stevens' method is based on a heuristic approach and fits better to the subjective ratings. The tendency for the Zwicker procedure to give values systematically larger than Stevens' has also been noted by others. This is not a drawback however, since, as discussed in Sec. 3.2, we were interested in relative loudness levels only.

3 CALCULATION OF LOUDNESS

As is necessary for loudness calculations, the responses of the loudspeakers have to be determined first. Six loudspeakers were used for this purpose. The loudspeakers, labeled LS₁–LS₆, were of different brands and covered wide ranges of price and quality. The frequency responses of the loudspeakers were measured both in an anechoic chamber and in the listening room where the listening tests were performed. The results of the free-field measurements are shown in Fig. 7. Clearly, the loudspeakers have different qualities and exhibit dissimilar frequency responses and, consequently, all sound very different.

The loudspeaker measurements in the listening room were made with a real-time one-third-octave analyzer

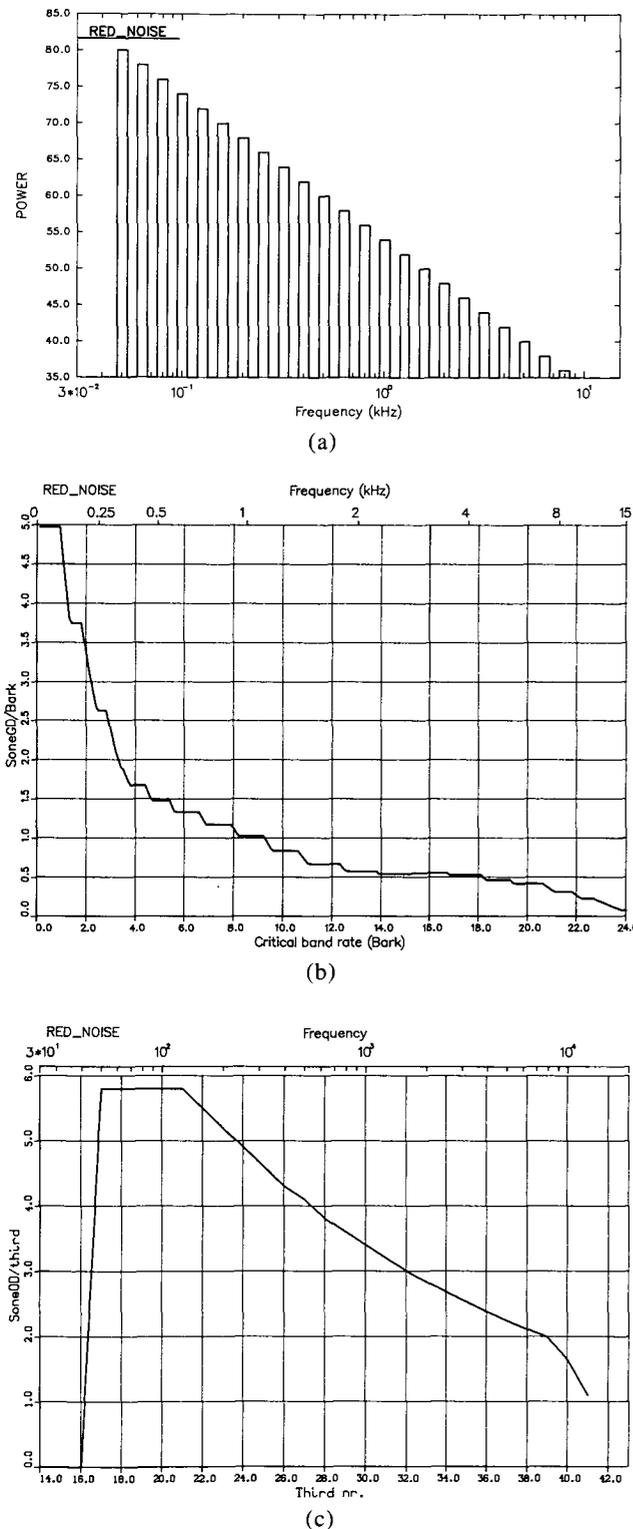


Fig. 6. (a) Red noise spectrum at 67.4 dBA level. (b) Red noise loudness spectrum according 532B [sones(GD)/bark]. (c) Red noise loudness spectrum according 532A [sones(OD)/one-third octave].

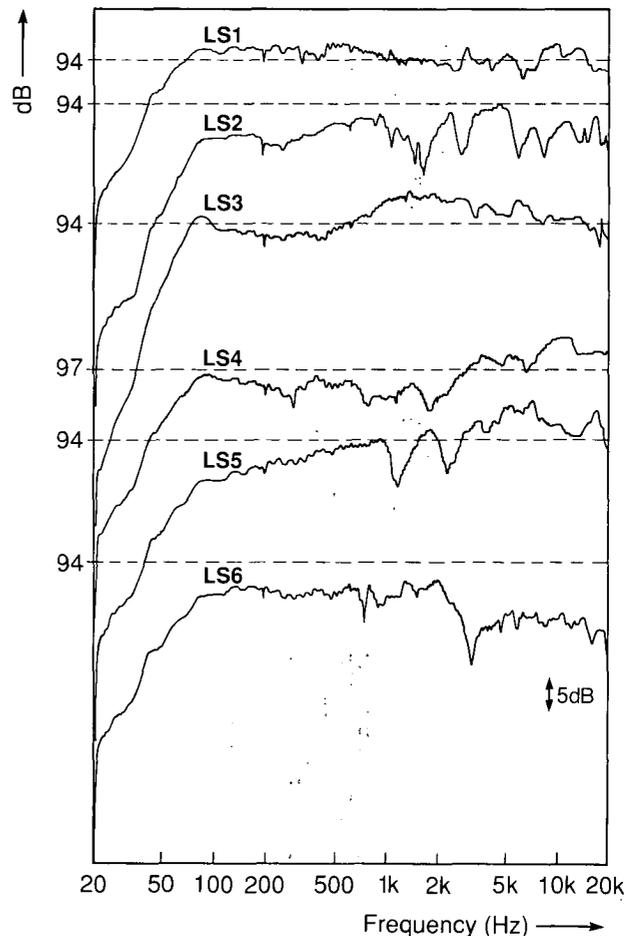


Fig. 7. Free-field on-axis response of six loudspeakers (3 V across terminals).

and a Brüel & Kjær 4134 microphone pointed toward the ceiling, while the loudspeaker being tested was reproducing pink noise. The microphone was placed in the middle of the listening area at ear height and the loudspeakers were placed next to each other. The results are depicted in Fig. 8. From these measurement results the weighted and unweighted sound pressure levels and the loudness levels were calculated. However, both direct sound as well as reflections due to the room were measured by the microphone and thus contributed to the measured SPL. The loudness models ignored non-simultaneous masking of the human ear, such as backward and forward masking [26]. The calculated loudness versus the critical band rate, using the Zwicker procedure, is given in Fig. 9, showing the prediction of the model as what one would perceive it. As can be seen, there is quite a difference between the various loudspeakers; for example, LS₅ exhibits a poor low-frequency response.

3.1 Subjective Loudness Measurements

The perceived sound quality of a loudspeaker and its relation to the various physical properties of the loudspeaker have been the subject of discussion and research for a long time; see, for example [17]–[21]. An important parameter during listening tests is the

setting of the sound level, both for the different programs and for the relative levels between the different loudspeakers. The latter is especially important since it is well known that a higher reproduction level, or loudness level, of a loudspeaker can lead to a higher appreciation score than another loudspeaker of the same quality, or even the same loudspeaker. The importance of equal loudness levels of the sounds being compared is shown by a striking investigation of Illényi and Korpássy [22]. They found that the rank order of the loudspeakers according to subjective quality judgments was in good agreement with the rank order obtained by the corresponding calculated loudness.

3.2 First Experiment

3.2.1 Aim

The aim of this experiment was to investigate the capability of subjects to match the loudness of noise sounds, each having a different timbre. The aim was also to inquire into the variability among subjects for such a task.

3.2.2 Method

Usually loudness is determined directly by using a pure tone at 1 kHz, and is obtained at other frequencies indirectly by means of loudness matching. However,

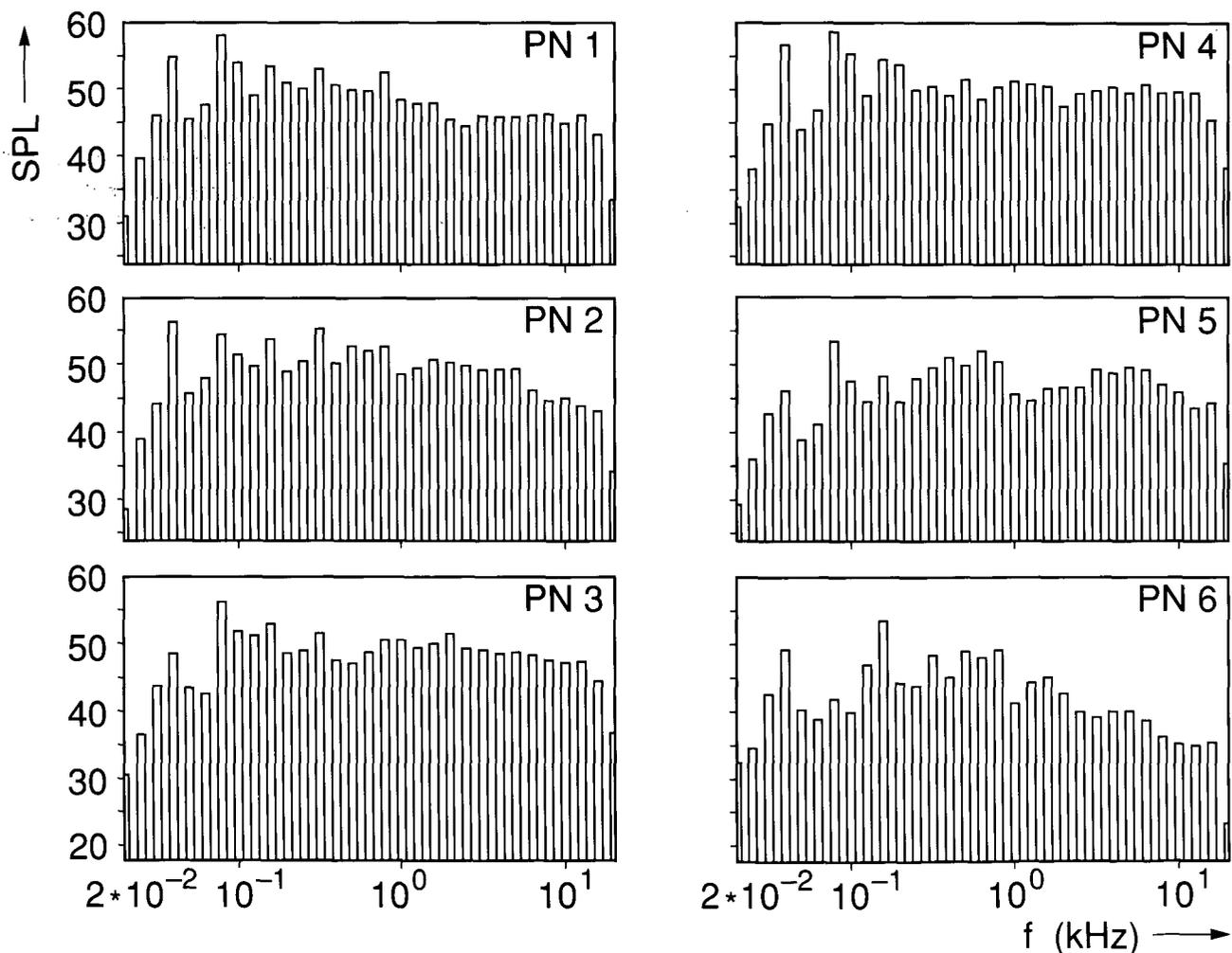


Fig. 8. Response of loudspeakers in listening room.

with this experiment only the relative loudness of several loudspeakers is considered, while eases the subjects' task considerably.

3.2.2.1 Listening Conditions. The subjects were seated in front of the loudspeakers, at a distance of 3.5 m (see Fig. 10). The listening room was a sound-proof room of 8.35-m length, 4.50-m width, and 2.62-m height and fulfilled the requirements of IEC 268-13 [23]. The room was arranged and equipped as a normal living room with chairs and furniture. Diffusion of the sound field was enhanced by a glass window, bookcases, and framed pictures. The room had a reverberation time of about 500 ms at 125 Hz, which gradually decreased to about 300 ms at 4 kHz.

3.2.2.2 Technical Equipment. We used six different loudspeakers, including the standard LS₁. An ostensible seventh loudspeaker LS₇ was used, but physically it was the same as LS₁. The loudspeakers were not seen by the subjects, due to an acoustically transparent but visually opaque screen. The loudspeakers were connected to a switching facility, which contained a set of high-quality relays, remotely controlled by the subject. Variable attenuators were placed in the signal path from the Compact Disc player to the power amplifier. Each loudspeaker could be attenuated by the

experimenter by adjusting the knob corresponding to the loudspeaker that was playing. To eliminate possible cues, the knobs could not be seen by the subjects.

3.2.2.3 Subjects. There were five male subjects involved, ranging in age from 22 to 46 years, some of whom were naive listeners and some highly experienced. All subjects had been recently tested for normal hearing, and they could all be considered as otologically normal



Fig. 10. Experimental setup.

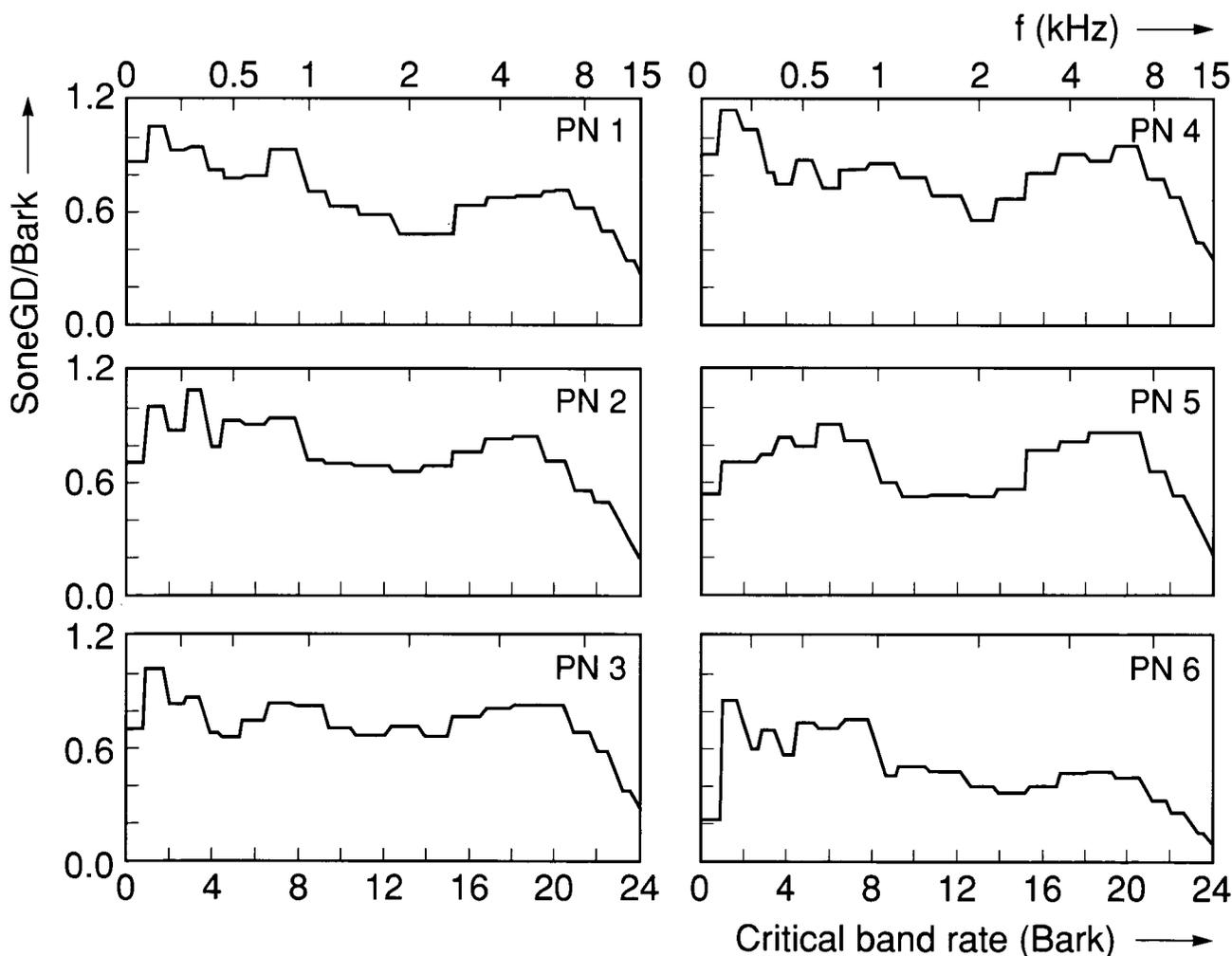


Fig. 9. Loudness of loudspeakers in listening room according to 532B [sones(GD)/bark].

persons.

3.2.2.4 Procedure. The stimuli were presented by reproducing pink noise via six different loudspeakers, LS₁–LS₆. The subjects could compare loudspeakers LS₂–LS₇ with the reference LS₁ as often as they desired. The reference loudspeaker was used as an anchor or standard, its volume setting remaining constant during all the tests, resulting in an SPL of 60 dBA for pink noise. The other loudspeakers were to be matched by the subjects so that they perceived an equal loudness level in comparison with the standard. The subjects gave a signal to the experimenter to lower or raise the volume of the loudspeaker being tested. The test was accomplished within two runs for each subject. In the first run the subject balanced loudspeakers LS₂–LS₇ to the reference one, which took approximately 10 min. The result of the test was the volume setting of each attenuator. After a short break of about 2 min the second run began, following the same procedure. Both runs used the same pink noise source for the programs. However, the programs, hereafter referred to as PN₁ and PN₂, were considered as factors when the ability of the subjects to reproduce the same loudness level as during the first run was studied.

3.2.3 Results

Analysis of variance (ANOVA) ([23]–[25], see Appendix) was performed to analyze the results of the listening tests. ANOVA essentially means that the total variance in the data is split up into different components due to the different sources of variation. These sources can be the program, the loudspeakers, the subjects, or other causes, or the interaction between them. The statistical tests make it possible to decide whether the differences between the sources of variation are real, with a certain probability, or whether they are caused by other (random) errors. The six loudspeakers and the program were considered as factors, being the possible sources of variation. The subjects were not considered as factors, however, this variation was taken in account also. Formally this is called a 6×2 design with repeated measures on the same elements. All the responses of the subjects are in decibels relative to the absolute level of LS₁. The individual responses are averaged and listed in Table 2. The results of the ANOVA computations are recorded in Table 3. The ratio of the resulting variance quantities will possess, under the ordinary normality assumptions, an F distribution [24]. When the F values exceed a certain critical value F_c ,

one may conclude that the corresponding factor is of statistical significance. The value of F_c depends on the value of p , which gives the probability that the variance is caused by change and not by the studied factor. For the computation of F_c , p was chosen to be equal to 0.0001.

The entries SS are the sums of squares and df are the number of degrees of freedom for the particular factors. (The definitions are given in the Appendix.) It appears from Table 3 that the loudspeakers are the only statistically significant factor ($p < 0.0001$). Furthermore, it was very surprising, at least to the author, that most subjects were very good at reproducible loudness balancing. They could adjust the relative loudness of a loudspeaker, say LS₄, to that of LS₁ and reproduced that level after a retention period of, say, 10–15 min with a good accuracy. One has to bear in mind that the timbres of all the loudspeakers were very different.

3.3 Second Experiment

3.3.1 Aim

The aim of this experiment was to investigate the capability of subjects to match the loudness of several loudspeakers for different programs, consisting of pink noise and music excerpts. A further aim was to study the influence of the program, or the existence of a possible interaction between program and loudspeakers.

3.3.2 Method

The same method as in the first experiment was used, but the program was extended by including music and the set of subjects was increased.

3.3.2.1 Stimuli. The program material consisted of excerpts from Compact Discs which, besides the quality of reproduction, has the advantage of allowing repetition of a fragment as often as required with only a brief interruption.

1) Pink noise, from the audio frequency test sample 3, Philips no. 410 055-2, which was also the noise source used in the first experiment.

2) Pop1, excerpt from “Early in the mornin’” (0:00–0:25), from the album *Step by Step* by Eddie Rabbit, Mercury no. 800 046-2. This fragment was used because of its wide spectrum and its rather large dynamic range in a short interval.

3) Pop2, excerpt from “It’s only love” (0:24–0:41), from the album *Let’s Stick Together* by Bryan Ferry, E’G no. 821 521-2.

Table 2. Table of means; values in decibels.

	PN ₁	PN ₂	Average
LS ₂	-0.05	-0.10	-0.07
LS ₃	-0.40	-0.50	-0.45
LS ₄	-1.00	-0.80	-0.90
LS ₅	1.15	0.35	0.75
LS ₆	4.70	5.05	4.88
LS ₇	-0.05	0.10	0.02
Average	0.72	0.68	0.70

Table 3. Analysis of variance.

Source	SS	df	F	F_c
Loudspeakers	223.72	5	43.39	6.6
Program	0.03	1	0.03	1.8
Interaction	2.07	5	0.40	6.6
Residual	49.50	48		
Total	275.31	59		

4) Jazz, excerpt from "Down home rag" (0:13–0:31), by the Dutch Swing College Band (track 8) from a demo disk of Philips no. 810 027-2.

5) Classic, excerpt from part II of Tchaikovsky's 5th symphony (7:15–7:44), Chicago Symphony Orchestra (conducted by G. Solti), Decca no. 425 516-2.

3.3.2.2 *Subjects.* In this experiment seven male and three female subjects participated, ranging from 21 to 46 years of age, all of whom were tested for normal hearing (<20 dB hearing loss, ISO 389).

3.3.3 Results

The responses of all subjects were measured in decibels relative to the absolute level of LS_1 . The individual responses were averaged and recorded in Table 4, while the ANOVA results are given in Table 5.

It appears from Table 5 that the loudspeakers are the only statistically significant factor ($p < 0.0001$). The influence of the program alone, as well as the interaction between program and loudspeakers, are not statistically significant. There remains a substantial residual variance however, due, in addition to normal errors such as inadvertence of the subjects, to the lack of a full consensus among subjects. Another reason for the residual variance is that not every subject receives the same signal due to the poor directivity of some loudspeakers, their positioning, and the different ear positions of the several subjects.

4 COMPARISON OF COMPUTED AND MEASURED LOUDNESS

The computed loudnesses for the loudspeakers mentioned in Sec. 3 are recorded in Table 6. Loudnesses LS_2 – LS_7 are relative to the loudness level of LS_1 . The results of the second experiment (the average over subjects and pink noise as program) are also gathered in Table 6 in the last column.

The error of an objective measure is considered to be the difference between the loudness levels of the subjective method and an objective method. These values are given in Table 7. The entry HT^2 is Hotelling's (squared) generalized student ratio or, briefly, Hotelling's T^2 [24].

Clearly, Table 7 shows that the dBA values can differ considerably from the subjective measurements. The statistical significance of the data presented in Table 7 can be tested against the following zero hypothesis: The differences between the various methods (dB, dBA, ISO-A and ISO-B) and the subjective ratings are due to change variations only. Using the T^2 values from Table 7 one can not reject the zero hypothesis for ISO 532B. The hypothesis is rejected for the other three methods. The entry α in Table 7 denotes the level of significance of the T^2 test: it is the probability of making the decision to reject the zero hypothesis when, in fact, it is true. One may conclude that the ISO 532B method

Table 4. Table of means; values in decibels.

	PN	Ferry	Rabbit	Jazz	Classical	Average
LS_2	-0.90	-1.40	-0.85	-1.62	-1.17	-1.19
LS_3	-0.57	-0.63	-0.35	0.13	0.32	-0.22
LS_4	-1.40	-1.23	-1.35	-1.10	-1.70	-1.36
LS_5	1.17	1.17	0.85	1.00	0.70	0.98
LS_6	5.75	5.02	5.09	5.32	5.60	5.36
LS_7	-0.13	-0.15	-0.10	0.42	-0.35	-0.06
Avg.	0.65	0.46	0.55	0.69	0.57	0.58

Table 5. Analysis of variance.

Source	SS	df	F	F_c
LSS	1545.54	5	256.58	5.1
Program	1.97	4	0.41	5.8
Interaction	20.58	20	0.85	2.6
Residual	325.27	270		
Total	1893.36	299		

Table 6. Computed loudness values.

	SPL (dB)	SPL (dBA)	ISO 532A [dB(OD)]	ISO 532B [dB(GD)]	Subjects (dB)
LS_1	65.0	59.6	75.3	80.7	—
LS_2	-0.4	-2.0	-1.0	-1.4	-0.9
LS_3	0.7	-1.8	-1.1	-0.9	-0.6
LS_4	-1.3	-2.8	-2.5	-2.1	-1.4
LS_5	2.3	-0.7	-0.1	0.2	1.2
LS_6	5.3	4.2	5.7	5.0	5.8
LS_7	—	—	—	—	-0.1

Table 7. Difference between subjective and objective measurements.

	SPL (dB)	SPL (dBA)	ISO 532A [dB(OD)]	ISO 532B [dB(GD)]
LS ₂	0.53	-1.14	-0.05	-0.46
LS ₃	1.26	-1.15	-0.53	-0.33
LS ₄	0.13	-1.36	-1.13	-0.69
LS ₅	1.11	-1.90	-1.32	-1.02
LS ₆	-0.50	-1.58	-0.08	-0.80
HT ²	14.94	54.32	16.58	5.94
α	$<10^{-3}$	$<10^{-4}$	$<10^{-3}$	~ 0.06

provides results similar to those of the subjects. The three other methods are not consistent with the subjective ratings.

5 CONCLUSIONS

The best technique for adjusting the interloudness levels of loudspeakers during listening tests is using the ISO 532B method. It provides similar results as the population of 10 subjects. The A-weighting method is not recommended for accurate loudness balancing. The results of the computed loudness level of pink noise reproduced by several loudspeakers agree very well with the average loudness level adjusted by several subjects. The average loudness for a varied repertoire is very similar to the computed loudness for pink noise as the program. There is a little variation among the subjects. However, analysis of variance shows that the only significant factor is the difference between the loudspeakers. The loudness levels are hardly influenced by the program choice.

6 ACKNOWLEDGMENT

The author wishes to thank Gillian Booles, who conducted all the listening tests: He would also like to thank all the subjects who were willing and patient listeners. He is indebted to the anonymous reviewers for their comments and suggestions.

7 REFERENCES

- [1] ISO 532-1975(E), "Acoustics—Method for Calculating Loudness Level," 1st ed. 1975, Int. Standards Organization (1977).
- [2] ISO 226-1987(E), "Acoustics—Normal Equal-Loudness Level Contours," Int. Standards Organization (1987).
- [3] L. F. Richardson and J. S. Ross, "Loudness and Telephone Current," *J. Gen. Psychol.*, pp. 121–164 (1916).
- [4] H. Fletcher and W. A. Munson, "Loudness, Its Definition, Measurement and Calculation," *J. Acoust. Soc. Am.*, vol. 5, pp. 82–108 (1933).
- [5] E. Zwicker, "Ein Verfahren zur Berechnung der Lautstärke," *Acustica*, vol. 10 (1960).
- [6] E. Zwicker, H. Fastl, and C. Dallmayr, "BASIC-Program for Calculating the Loudness of Sounds from Their 1/3-oct Band Spectra According to ISO 532B,"

Acustica (Letter to the Editors), vol. 55, pp. 63–67 (1984).

- [7] F. H. Brittain, "The Loudness of Continuous Spectrum Noise and Its Application to Loudness Measurements," *J. Acoust. Soc. Am.*, vol. 11, pp. 113–117 (1939).
- [8] B. Scharf, in *Handbook of Perception*, vol. 4, E. C. Carterette (Ed.). (Academic Press, New York, 1978).
- [9] S. S. Stevens, "Procedure for Calculating Loudness: Mark VI," *J. Acoust. Soc. Am.*, vol. 33, pp. 1577–1585 (1961).
- [10] S. S. Stevens, "Perceived Level of Noise by Mark VII and Decibels (E)," *J. Acoust. Soc. Am.*, vol. 51, pp. 575–601 (1972).
- [11] E. Zwicker and R. Feldtkeller, *Das Ohr als Nachrichtenempfänger* (S. Hirzel, Stuttgart, FRG, 1967).
- [12] E. Paulus and E. Zwicker, "Programme zur automatischen Bestimmung der Lautheit aus Tertzpegeln oder Frequenzgruppenpegeln," *Acustica*, vol. 27, pp. 253–266 (1972).
- [13] E. Zwicker, *Psychoakustik* (Springer, New York, 1982).
- [14] R. M. Aarts, "Algebraic Expression for A-Weighting Response," *J. Sound Vibration*, vol. 115, p. 372 (1987).
- [15] E. Lübcke, G. Mittag, and E. Port, "Subjektive und Objektive Bewertung von Machinengeräuschen," *Acustica*, vol. 14, pp. 105–114 (1964).
- [16] M. Jahn, "Subjektive und Objektive Bewertung von Machinengeräuschen," *Acustica*, vol. 16, pp. 175–186 (1965/1966).
- [17] F. E. Toole, "Listening Tests—Turning Opinion into Fact," *J. Audio Eng. Soc. (Engineering Reports)*, vol. 30, pp. 431–445 (1982 June).
- [18] F. E. Toole, "Subjective Measurements of Loudspeaker Sound Quality and Listener Performance," *J. Audio Eng. Soc.*, vol. 33, pp. 2–32 (1985 Jan./Feb.).
- [19] F. E. Toole, "Loudspeaker Measurements and Their Relationship to Listener Preferences, Parts 1 and 2," *J. Audio Eng. Soc.*, vol. 34, pp. 227–235 (1986 Apr.); pp. 323–348 (1986 May).
- [20] A. Gabrielsson and B. Lindström, "Perceived Sound Quality of High-Fidelity Loudspeakers," *J. Audio Eng. Soc.*, vol. 33, pp. 33–53 (1985 Jan./Feb.).
- [21] Y. Tannaka and T. Koshikawa, "Correlations between Soundfield Characteristics and Subjective

Ratings on Reproduced Music Quality," *J. Acoust. Soc. Am.*, vol. 86 (1989 Aug.).

[22] A. Illényi and P. Korpássy, "Correlation between Loudness and Quality of Stereophonic Loudspeakers," *Acustica*, vol. 49, pp. 334-336 (1981 Dec.).

[23] IEC 268-13, "Sound System Equipment, Part 13: Listening Tests on Loudspeakers," International Electrotechnical Commission, Geneva, Switzerland (1985).

[24] B. J. Winer, *Statistical Principles in Experimental Design* (McGraw-Hill, New York, 1962).

[25] A. Gabrielsson, "Statistical Treatment of Data from Listening Tests on Sound-Reproducing Systems," Rep. TA 92, Dept. of Technical Audiology, Karolinska Inst., Sweden (1979).

[26] B. C. Y. Moore, *An Introduction to the Psychology of Hearing* (Academic Press, New York, 1982).

**APPENDIX
COMPUTED QUANTITIES**

In a two-way cross-classification there are r "rows" (loudspeakers) and c "columns" (program), with m observations (subjects) for each row and column combination. Let y_{kij} denote the k^{th} observation at row i and column j . The following quantities are computed:

Cell average

$$y_{\cdot ij} = \sum_{k=1}^m y_{kij}/m \tag{6}$$

Row (loudspeakers) means

$$y_{\cdot i} = \sum_{j=1}^c y_{\cdot ij}/c \tag{7}$$

Column (program) means

$$y_{\cdot j} = \sum_{i=1}^r y_{\cdot ij}/r \tag{8}$$

Grand mean

$$y_{\dots} = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^m y_{kij}/rcm \tag{9}$$

Sums of squares

$$\text{rows } SS_1 = cm \sum_{i=1}^r (y_{\cdot i} - y_{\dots})^2 \tag{10a}$$

$$\text{columns } SS_2 = rm \sum_{j=1}^c (y_{\cdot j} - y_{\dots})^2 \tag{10b}$$

$$\text{interaction } SS_3 = m \sum_{i=1}^r \sum_{j=1}^c (y_{ij} - y_{\cdot i} - y_{\cdot j} + y_{\dots})^2 \tag{10c}$$

$$\text{residual } SS_4 = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^m (y_{kij} - y_{\cdot ij})^2 \tag{10d}$$

$$\text{corrected total } SS_5 = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^m (y_{kij} - y_{\dots})^2 \tag{10e}$$

Degrees of freedom

$$\text{rows } df_1 = r - 1 \tag{11a}$$

$$\text{columns } df_2 = c - 1 \tag{11b}$$

$$\text{interaction } df_3 = (r - 1)(c - 1) \tag{11c}$$

$$\text{residual } df_4 = (m - 1)rc \tag{11d}$$

$$\text{total } df_5 = rcm - 1 \tag{11e}$$

F ratios

$$\text{rows } F_1 = \frac{SS_1/df_1}{SS_4/df_4} \tag{12a}$$

$$\text{columns } F_2 = \frac{SS_2/df_2}{SS_4/df_4} \tag{12b}$$

$$\text{interaction } F_3 = \frac{SS_3/df_3}{SS_4/df_4} \tag{12c}$$

THE AUTHOR



Ronald M. Aarts was born in Amsterdam, The Netherlands, in 1956. He received a B.Sc. degree in electrical engineering in 1977, then joined the optics group of Philips Research Laboratories where he was engaged in research into servos and signal processing for use in both video long-play players and Compact Disc

players. In 1984 he joined the acoustics group of the Philips Research Laboratories and was engaged in the development of CAD tools for loudspeaker systems.

Mr. Aarts has published a number of technical papers and reports and holds several patents in his field. He is a member of the AES and the ASA.