

The Calculation of Loudness of Loudspeakers during Listening Tests

R.M. Aarts

Philips Research Laboratories
Eindhoven, The Netherlands

**Presented at
the 88th Convention
1990 March 13–16
Montreux**



AES

This preprint has been reproduced from the author's advance manuscript, without editing, corrections or consideration by the Review Board. The AES takes no responsibility for the contents.

Additional preprints may be obtained by sending request and remittance to the Audio Engineering Society, 60 East 42nd Street, New York, New York 10165 USA.

All rights reserved. Reproduction of this preprint, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

AN AUDIO ENGINEERING SOCIETY PREPRINT

The calculation of loudness of loudspeakers during listening tests

R.M. Aarts

Philips Research Laboratories
5600 JA, Eindhoven, The Netherlands

Abstract

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 use of Zwicker's loudness model appears to be the best technique for adjustment of the inter-loudness 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 scaling of loudness, a brief description is given of a computer program for 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.

Secondly, 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 are the same in comparison 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 lead to misleading loudspeaker quality assessments.

0.1 Definitions

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

0.1.1 Loudness

That attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from soft to loud.

0.1.2 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.

0.1.3 Critical bandwidth

The widest frequency band within which the loudness of a band of continuously distributed random noise of constant band sound pressure level 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.

0.3 Loudness of complex sounds

Over the past five decades various methods of evaluating loudness of complex sounds from objective spectrum analysis have been proposed. The earliest attempt to use al-

gebraic 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, by Zwicker [5, 6]. A-weighting is a widely used method, based on an early 40 phon contour it is a rough approximation of an ear weighting. It is traditionally applied in sound level meters to measure the loudness of signals or to determine the annoyance of noise. 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 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 an striking example that for wideband signals the A-weighted method is generally too simple.

1 The calculation of loudness

A computer program has been written according to both Stevens' and Zwicker's method as specified in the ISO532 standard: 'Acoustics - method for calculating loudness level' [1]. The input for both programs is a file containing SPL values measured in one third octave bands. The essential parts of both methods will be briefly discussed in the following sections.

1.1 The 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 the 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 in sones:

$$S = .01 (p - p_0)^6, \quad (1)$$

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

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

or

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

where P is the loudness in phons, leading to the ISO532 definition: 'One Sone is the loudness of a sound whose loudness level is 40 phons'. It is important to discriminate between loudness levels which 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 octave G for group (Frequenzgruppen or critical bands), D for diffuse and F for free field. Method 532A (Stevens) assumes that the measurements are taken in a diffuse sound field, whilst Method 532B (Zwicker) allows diffuse and free field conditions to be set as a parameter 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 the 532A method is briefly discussed. The SPL of each third is converted to a loudness index using a table based on subjective measurements. The total loudness in sones(OD) S_i is calculated by means of the equation

$$S_i = S_m + F (\Sigma S_i - S_m), \quad (3)$$

where S_m is the greatest of the loudness indices and ΣS_i is the sum of the loudness indices of all the bands. For one-third octave bands the value of F is 0.15, however if the data is measured in different bands one has to use 0.3 for octave bands and 0.2 for

one-half octave bands. The total loudness may be converted into loudness level in phons(OD) using equation (2b).

1.3 Zwicker's method: 532B

An early version is described in [5] and later it is refined see eg. [11,12,13]. In the following the essential steps in the procedure will be discussed. The spectrum measured in third-octaves is converted into frequency bands having a bandwidth approximating the critical bands of the human ear. This warping of the frequency axis into the critical band rate, with unit Bark, is depicted in Fig. 2. The thirds up to 90 Hz are assembled as the first critical band, the next three thirds (90 Hz to 180 Hz) as the second band and the bands ranging from 180 Hz 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:

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

where L_{125} etc., is the measured one-third octave band SPL for the band with centre frequency 125 Hz, and L_{c2} is the level in the second critical band. The relation between the frequency f_i and the third number i is given by

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

The SPL of each critical band is transformed into a loudness index. 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 90dB SPL. When a band has a lower loudness index level compared to its neighbors, the excitation within that band makes no contribution to the overall result. The total loudness is finally calculated by integrating the loudness indices over all the critical bands, resulting in the loudness in sone(GF) or sone(DF). To integrate the 'tail' as in Fig.3 with sufficient accuracy, each critical band is divided into sub-bands 0.1 Bark wide.

2 The Loudness program

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

Table I

noise	400Hz Level dB	SPL dB	Level dBA	Judged level dB	Stevens OD Phons	Zwicker GD Phons	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	-
Sine1k	-	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

The table has the following vertical entries:

- the type of noise spectra
- the noise level in the third of 400 Hz.
- the total SPL of the spectrum, i.e. the value obtained by simply adding the sound pressure in all the thirds.
- the total A-weighted power of the noise spectrum, i.e. the value obtained by using the A-weighting formula [14].
- the loudness in phons(OD) (method 532A) calculated by assuming diffuse field measurement conditions.
- the loudness in phons(GD) (method 532B) calculated by assuming diffuse field measurement conditions.

- the figure numbers referring to the corresponding spectrum.

The table 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 Hz to 11 kHz and scaled to a noise power of 74 dB.
- Sine1k (Fig. 3) is 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).

2.1 Comparison of 532A en 532B

At first glance there is quite a difference when the figures 4b-6b are compared with their counterparts 4c-6c, but one has to bear in mind that the frequency axes are different (as is discussed in section 2.3).

Another apparent difference is the low frequency gain for 532B, which may be on account of the addition of the first three groups of thirds into critical bands, as discussed in section 2.3. 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 models 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 section 4, we were interested in relative loudness levels only.

3 Calculation of loudness

As is necessary for loudness calculations, the responses of the loudspeakers first have to be determined. Six loudspeakers were used throughout this paper for this purpose. The loudspeakers, hereafter labelled LS1-LS6, were of different brands and covered a wide price and quality range. 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 analyser and a Brüel & Kjær 4134 microphone pointed towards the ceiling, whilst the loudspeaker being tested was reproducing pink noise. The microphone was placed in the middle of the listener 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. The calculated loudness versus the critical band rate is shown in Fig. 9.

4 Subjective loudness measurements

The perceived sound quality of a loudspeaker and its relation to the various physical properties of the loudspeaker has been a subject for discussion and research for a long time, see e.g. [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 as it is well known that a higher reproduction level, or loudness level, of a loudspeaker can lead to a higher appreciation score than that of another one 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 Korpassy [22]. They found that the rank order of the loudspeakers according to the subjective quality judgements was in good agreement with the rank order obtained by the corresponding calculated loudness.

4.1 First Experiment

4.1.1 Aim

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

4.1.2 Method

Usually, loudness is determined directly using a pure tone at 1kHz, and is obtained at other frequencies indirectly by means of loudness matching. However, with this experiment only the relative loudness of several loudspeakers is considered, which eases the subjects' task considerably.

4.1.2.1 Listening Conditions

The subjects were seated in front of the loudspeakers, at a distance of 3.5m, see Fig. 10. The listening room was a soundproof room of dimensions: length 8.35 m, breadth 4.50 m and height 2.62 m, and fulfills the requirements of the IEC 268-13 [23]. The room is arranged and equipped as a normal living room with chairs and furniture. Diffusion of the sound field is created by a glass window, bookcases and framed pictures. The room has a reverberation time of about 500ms at 125Hz, which gradually decreases to about 300ms at 4kHz.

4.1.2.2 Technical Equipment

We used six different loudspeakers, including the standard (LS1). An ostensible 7th loudspeaker (LS7) was used, but this was physically 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 CD player to the power amplifier. Each loudspeaker could be attenuated by the experimenter, by adjusting the knob corresponding to the loudspeaker which was playing. To eliminate possible cues the knobs could not be seen by the subjects.

4.1.2.3 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 were recently tested for normal hearing and they could all be considered as otologically normal persons.

4.1.2.4 Procedure

The stimuli were presented by reproducing pink noise via six different loudspeakers, LS1-LS6. The subjects could compare the loudspeakers LS2-LS7; with the reference one (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 the loudspeakers LS2-LS7 to the reference one, which took approximately 10 minutes. After a short break of about two minutes, 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 respectively, were considered as a factor to study the ability of the subjects to reproduce the same loudness level as during the first run.

4.1.3 Results

Analysis of variance (ANOVA) [23, 24, 25, Appendix A] was performed to analyse 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 a factor, however their variation was taken in account also. Formally this is called a 6X2 design with repeated measures on the same elements. All the responses of the subjects are in dB, relative to the absolute level of LS1. The individual responses are averaged and listed in the Table of Means in Table IIa. The results of the ANOVA computations are recorded in Table IIb. The ratio of the resulting variance quantities will possess an F distribution, which can be considered as a generalization of Student's t distribution [24]. When the F values exceed a certain critical F 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.

Table IIa. Table of Means (values in dB).

	PN1	PN2	Avg.
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
Avg.	0.72	0.68	0.70

Table IIb. Analysis of Variance Table.

Source	SS	df	F	F_c
LSS	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		

The entries SS are the sum of squares and the entries df the number of degrees of freedom for the particular factors, respectively (the definitions are given in Appendix A). It appears from the table 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 are 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 minutes, with a good accuracy. One has to bear in mind that the timbre of all the loudspeakers is very different.

4.2 Second experiment

4.2.1 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.

4.2.2 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.

4.2.2.1 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.

- a) Pink noise, from the audio frequency test sample No. 3 Philips No. 410 055-2, which was also the noise source used in the first experiment.
- b) 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.
- c) 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.
- d) 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.
- e) Classic, excerpt from: part II of Tchaikovsky's 5th symphony (7:15-7:44), Chicago Symphony Orchestra (cond. by G. Solti), Decca No. 425 516-2.

4.2.2.2 Subjects

In this experiment 7 male and 3 female subjects participated, ranging from 21 to 46 years of age, all of whom were tested for normal hearing.

4.2.3 Results

All subjects' responses were measured in dB relative to the absolute level of LS1. The individual responses were averaged and recorded in Table IIIa, while the ANOVA results are in Table IIIb.

Table IIIa. Table of Means (values in dB).

		PN	Ferry	Rabbit	Jazz	Clas.	Avg.
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
<hr/>							
Avg.		0.65	0.46	0.55	0.69	0.57	0.58

Table IIIb. Analysis of Variance Table.

Source	SS	df	F	Fc
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		

It appears from the table 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 amongst the 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 position of the several subjects (ranging from midgets to tall ones).

5 Comparison of computed and measured loudness

The computed loudnesses for the loudspeakers mentioned in section 3 are recorded in Table IV; the loudnesses (LS2-LS7) are relative to the loudness level of LS1. The results of the second experiment (the average over subjects and pink noise as program) are gathered in the same table. The error of an objective measure is considered to be the difference between the loudness levels of the subjective and an objective method respectively. The entries Emax and SS are the maximal error and the variation in the errors (sum of squares of the differences between the error and average error) respectively, while Etot are the sums of squared errors.

Clearly, the table shows that the dBA values can differ considerably from the subjective measurements. The Etot and SS values for the A weighting exhibit high and low values respectively, which indicates a bias value. However, this is not an ordinary bias, because the loudness levels are already relative (to LS1), besides such a bias will be dependent on the choice of the reference loudspeaker. The values obtained by the loudness method according to Zwicker's model show the best resemblance with the subjective measurements.

Table IV

	SPL dB	SPL dBA	ISO532A dB (OD)	ISO532B dB (GD)	subj. dB
LS 1	65.0	60.0	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	5.8	5.0	5.8
LS 7	0.0	0.0	0.0	0.0	-0.1
Emax	1.3	1.9	1.3	1.0	-
Etot	3.4	10.8	3.2	2.5	-
SS	2.2	0.4	1.4	0.3	-

6 Conclusions

The best technique for adjustment of the inter-loudness levels of loudspeakers during listening tests is Zwicker's loudness model. The A-weighting method is risky and it is inadvisable to use it 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 some variance between the subjects, however analysis of variance shows that the only significant factor is the difference between the loudspeakers and not the program or an interaction between them.

7 Acknowledgments

The author expresses his thanks to Gillian Booles who conducted all the listening tests. He would also like to thank all the subjects who were willing and patient listeners.

8 References

- [1] International standard ISO 532-1975(E), Acoustics - Method for calculating loudness level, (1st ed.1975) 1977.
- [2] International standard ISO 226-1987(E), Acoustics - Normal equal-loudness level contours, 1987.
- [3] L.F.Richardson, J.S. Ross, Loudness and telephone current, J. Gen. Psychol., pp. 121-164, 1916.
- [4] H. Fletcher, 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, C. Dallmayr, Letter to the Editors: BASIC-Program for calculating the loudness of sounds from their 1/3-oct band spectra according to ISO 532B, Acustica, Vol. 55, pp. 63-67, 1984.
- [7] F.H. Britain, The loudness of continuous spectrum noise and its application to loudness measurements, J. Acoust. Soc. Am., Vol. 11, pp. 113-117, 1938.
- [8] B. Scharf in: Handbook of Perception Vol.IV, Ed. E.C. Carterette, Academic Press, New York, 1978.
- [9] S.S. Stevens, Procedure for Calculating Loudness: Mark VI, J. Acoust. Soc. Am., Vol. 33, N. 11, pp. 1577-1585, 1961.
- [10] S.S. Stevens, Perceived Level of Noise by Mark VII and Decibels (E), J. Acoust. Soc. Am., Vol. 51, N. 2 (Part 2), pp. 575-601, 1972.
- [11] E. Zwicker, R. Feldtkeller, Das Ohr als Nachrichtenempfänger, S. Hirzel Verlag, Stuttgart, 1967.

[12] E. Paulus, E. Zwicker, Programme zur automatischen Bestimmung der Lautheit aus Tertzpegeln oder Frequenzgruppenpegeln, *Acustica*, Vol.27, Heft 5, pp. 253-266, 1972.

[13] E. Zwicker, *Psychoakustik*, Springer-Verlag, New York, 1982.

[14] R.M. Aarts, Algebraic Expression for A-Weighting Response, *J. of Sound and Vibration*, Vol. 115, N. 2, p. 372, 1987.

[15] E. Lübecke, G. Mittag, 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.*, Vol. 30, No.6, pp. 431-445, June 1982.

[18] F.E. Toole, Subjective Measurements of Loudspeaker Sound Quality and Listener Performance, *J. Audio Eng. Soc.*, Vol. 33, No.1/2, pp. 3-32, Jan./Feb. 1985.

[19] F.E. Toole, Loudspeaker Measurements and Their Relationship to Listener Preferences, Part 1: *J. Audio Eng. Soc.*, Vol. 34, No.4, pp. 227-235, April 1986. Part 2: *J. Audio Eng. Soc.*, Vol. 34, No.5, pp. 323-348, May 1986.

[20] A. Gabrielsson, B. Lindström, Perceived Sound Quality of High-Fidelity Loudspeakers, *J. Audio Eng. Soc.*, Vol. 33, No.1/2, pp.33-53, Jan/Feb 1985.

[21] Y. Tannaka, T. Koshikawa, Correlations between soundfield characteristics and subjective ratings on reproduced music quality, *J. Acoust. Soc. Am.*, Vol.86, No.2, Aug. 1989.

[22] A. Illényi, P. Korpássy, Correlation Between Loudness and Quality of Stereophonic Loudspeakers, *Acustica*, Vol. 49, No. 4, pp. 334-336, Dec. 1981.

- [23] International Electrotechnical Commission (IEC), Publication 268-13, Sound system equipment, Part 13: Listening tests on loudspeakers, Geneva, Switzerland, 1985.
- [24] B.J. Winer, Statistical principles in experimental design, McGraw-Hill Book Company, New York, 1962.
- [25] A. Gabrielsson, Statistical treatment of data from listening tests on sound-reproducing systems, Report TA No. 92, Dept. of Technical Audiology, Karolinska Institutet, Sweden, 1979.

Appendix A

Computed quantities

In a two-way cross-classification there are r 'rows' (the loudspeakers) and c 'columns' (the program), with m observations (the 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.

The cell means

$$y_{.ij} = \sum_{k=1}^m y_{kij} / m. \quad (A1)$$

The row (loudspeakers) means

$$y_{.i.} = \sum_{j=1}^c y_{.ij} / c. \quad (A2)$$

The column (program) means

$$y_{..j} = \sum_{i=1}^r y_{.ij} / r. \quad (A3)$$

The grand mean

$$y_{...} = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^m y_{kij} / r c m. \quad (A4)$$

Sums of squares.

$$\text{Rows} = SS1 = c m \sum_{i=1}^r (y_{.i.} - y_{...})^2. \quad (A5 - a)$$

$$\text{Columns} = SS2 = r m \sum_{j=1}^c (y_{..j} - y_{...})^2. \quad (A5 - b)$$

$$\text{Interaction} = SS3 = m \sum_{i=1}^r \sum_{j=1}^c (y_{.ij} - y_{.i.} - y_{..j} + y_{...})^2. \quad (A5 - c)$$

$$\text{Residual} = SS4 = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^m (y_{kij} - y_{.ij})^2. \quad (A5 - d)$$

$$\text{Corrected total} = SS5 = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^m (y_{kij} - y_{...})^2. \quad (A5 - e)$$

Degrees of freedom.

$$\text{Rows: } df1 = r - 1. \quad (A6 - a)$$

$$\text{Columns: } df2 = c - 1. \quad (A6 - b)$$

$$\text{Interaction: } df3 = (r - 1)(c - 1). \quad (A6 - c)$$

$$\text{Residual: } df4 = (m - 1)rc. \quad (A6 - d)$$

$$\text{Total: } df5 = rc m - 1. \quad (A6 - e)$$

F ratios.

$$\text{Rows: } F1 = \frac{SS1/df1}{SS4/df4}. \quad (A7 - a)$$

$$\text{Columns: } F2 = \frac{SS2/df2}{SS4/df4}. \quad (A7 - b)$$

$$\text{Interaction: } F3 = \frac{SS3/df3}{SS4/df4}. \quad (A7 - c)$$

Figure Captions

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

Fig. 2. The relationship between frequency, third number and critical band rate.

Fig. 3. The transformation from SPL into loudness for a 1 kHz tone at 90 dB.

Fig. 4a. Pink noise spectrum at 96 dB SPL.

Fig. 4b. Pink noise loudness spectrum according 532B [Sone(GD)/Bark].

Fig. 4c. Pink noise loudness spectrum according 532A [Sone(OD)/third].

Fig. 5a. White noise spectrum at 81.2 dBA level.

Fig. 5b. White noise loudness spectrum according 532B [Sone(GD)/Bark].

Fig. 5c. White noise loudness spectrum according 532A [Sone(OD)/third].

Fig. 6a. Red noise spectrum at 67.4 dBA level.

Fig. 6b. Red noise loudness spectrum according 532B [Sone(GD)/Bark].

Fig. 6c. Red noise loudness spectrum according 532A [Sone(OD)/third].

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

Fig. 8. Response of the loudspeakers in the listening room.

Fig. 9. Loudness (soneGD) of the loudspeakers in the listening room.

Fig. 10. Experimental set-up.

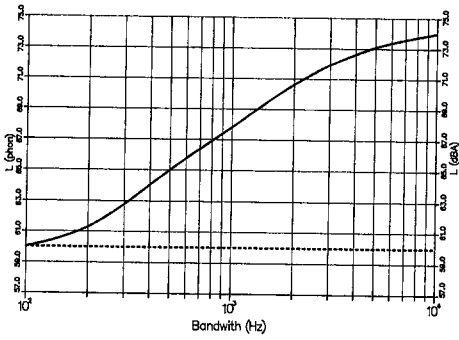


Fig. 1. Loudness (solid curve) and dBA Level (dashed line) vs. bandwidth of white noise.

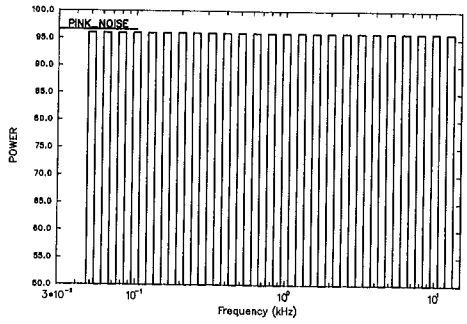


Fig. 4a. Pink noise spectrum at 96 dB SPL.

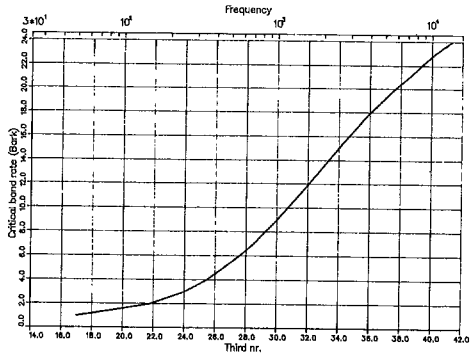


Fig. 2. The relationship between frequency, third number and critical band rate.

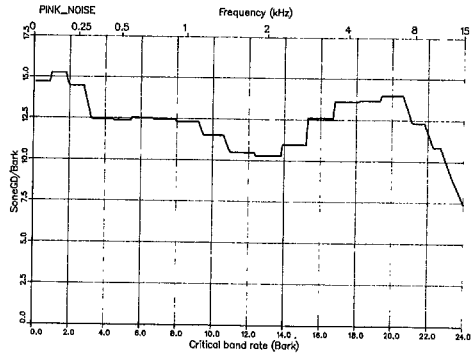


Fig. 4b. Pink noise loudness spectrum according 532B

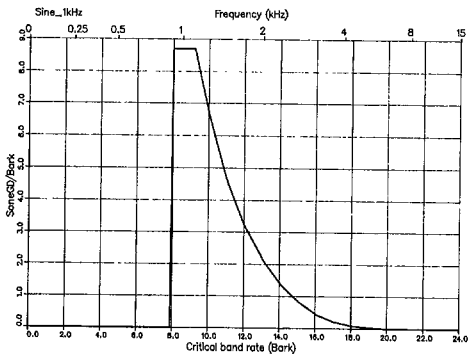


Fig. 3. The transformation from SPL into loudness for a 1 kHz tone at 90 dB.

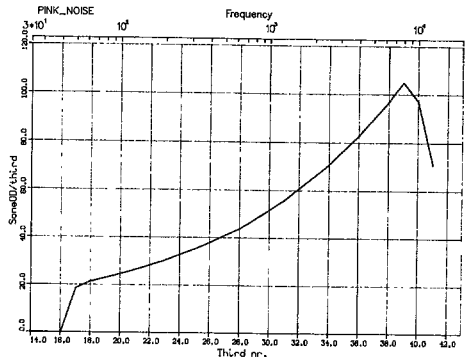


Fig. 4c. Pink noise loudness spectrum according 532A

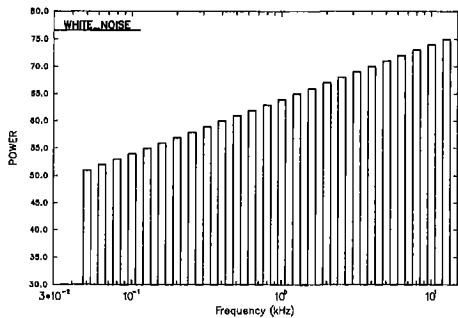


Fig. 5a. White noise spectrum at 81.2 dBA level.

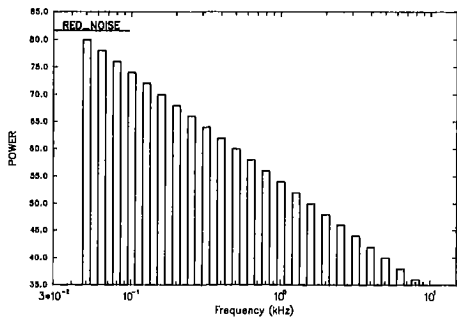


Fig. 6a. Red noise spectrum at 67.4 dBA level.

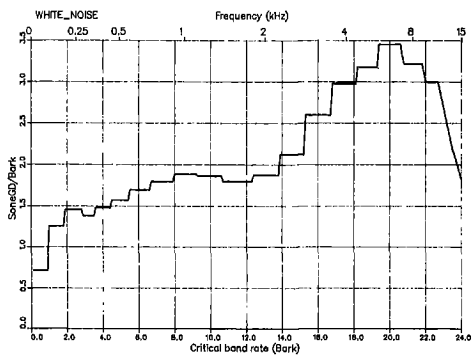


Fig. 5b. White noise loudness spectrum according 532B

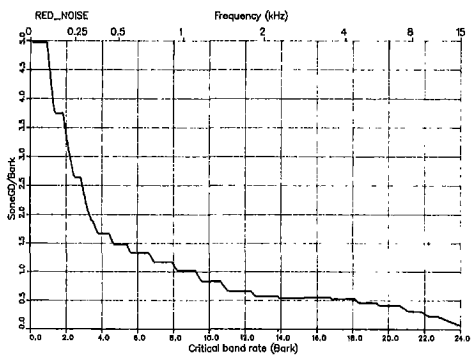


Fig. 6b. Red noise loudness spectrum according 532B

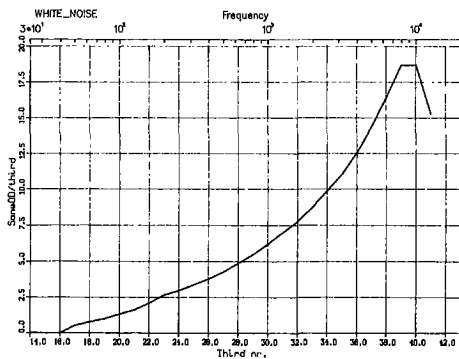


Fig. 5c. White noise loudness spectrum according 532A

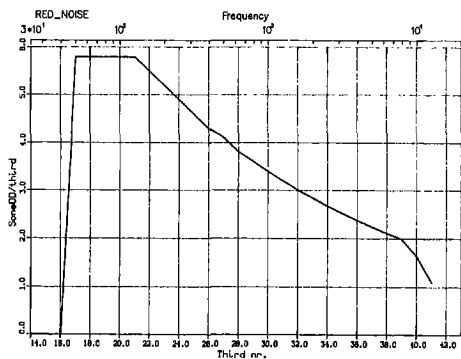


Fig. 6c. Red noise loudness spectrum according 532A

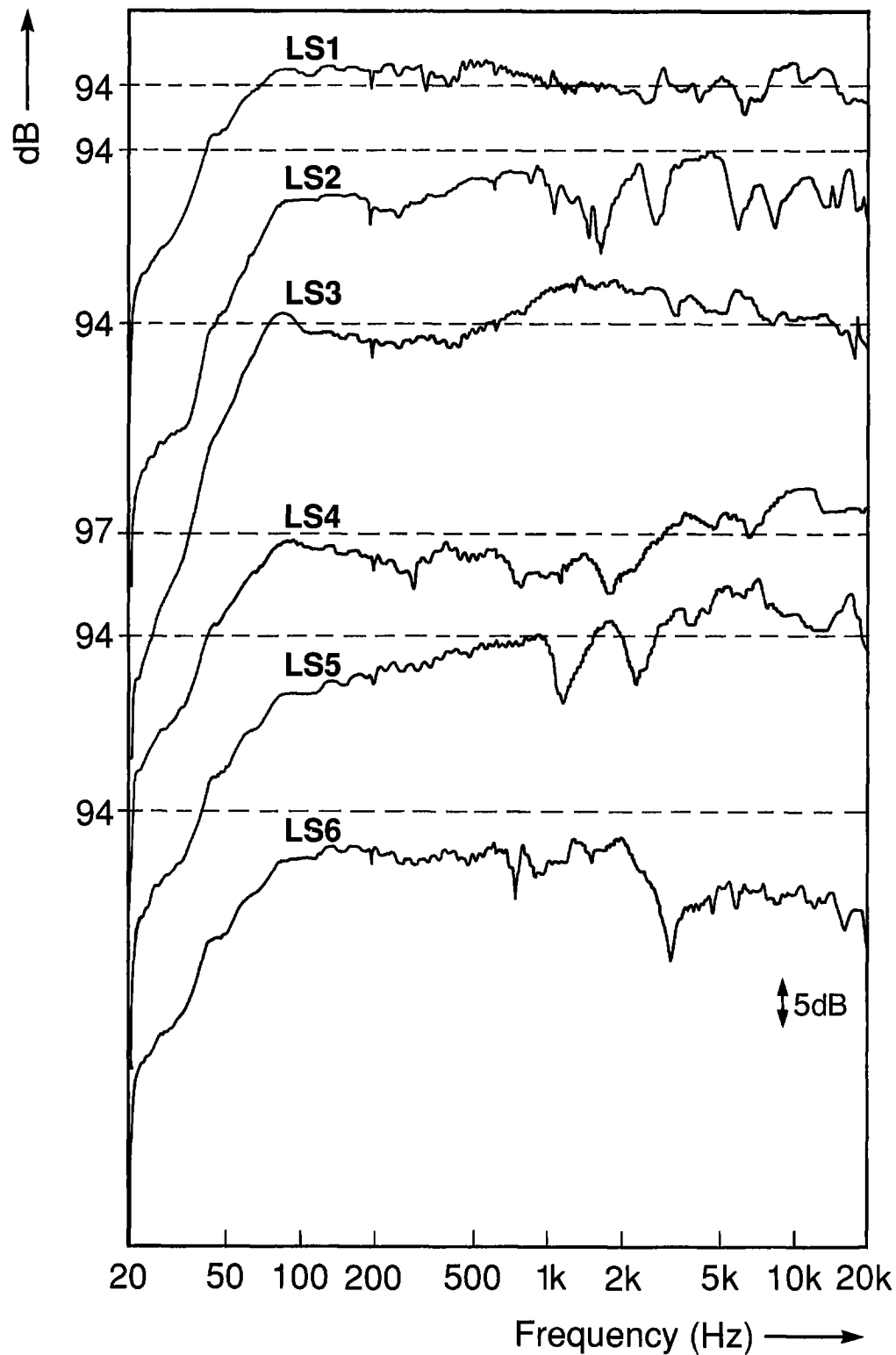


Fig. 7. Free field on-axis response of the six loudspeakers (3V across the terminals).

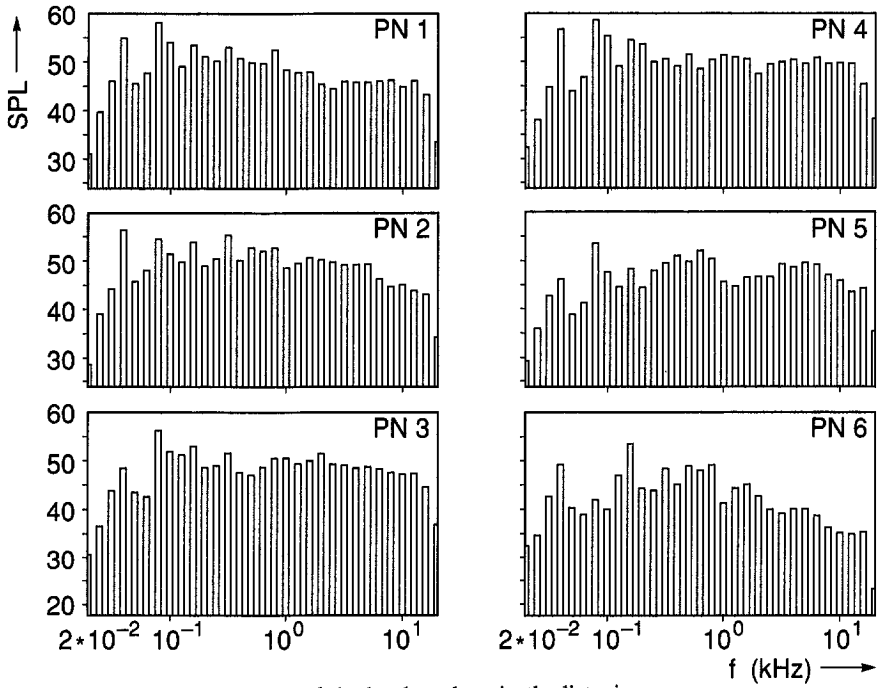


Fig. 8. Response of the loudspeakers in the listening room.

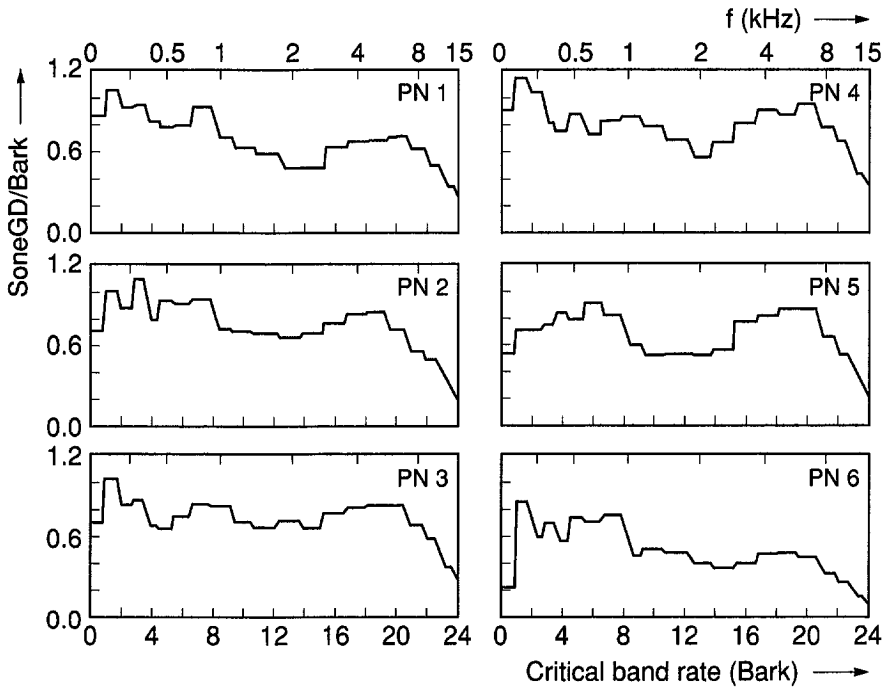


Fig. 9. Loudness (soneGD) of the loudspeakers in the listening room.



Fig. 10. Experimental set-up.