*Journal of Sound and Vibration* (1997) **205**(4), 523–531

# CALIBRATION OF NOISE ANNOYANCE RATINGS AT WORK

M. Tesarz and A. Kjellberg

*National Institute for Working Life, S-171 84 SOLNA, Sweden*

and

K. Holmberg and U. Landström

*National Institute for Working Life, Box 7654, S-907 13 Umeå, Sweden*

(*Received* 3 *March* 1997)

Three methods for calibrating noise annoyance ratings were tested in a group of 206 persons from different work places. One was Master scaling, which involves calibration for differences in mean ratings as well as the slope of the psychophysical function for seven levels of a pink noise reference sound. The other two methods calibrated only for differences in mean ratings of reference sounds. The reference sounds were either the pink noises or verbal descriptions of three everyday sounds. Calibration for differences in mean ratings leads to a sixfold increase of the variance explained by sound level. Calibrating for differences in slope did not lead to any further improvement. The described sounds worked only slightly worse than the recorded sounds as reference sounds.



## 1. INTRODUCTION

Sound level has generally been found to explain less than 20% of the individual differences in noise annoyance [1]. One reason for the low correlation is that frequency weightings were developed to predict loudness, not annoyance. Although loudness is an important determinant of annoyance, many other factors have a great impact on the annoyance response [1, 2]. Another reason for the low correlation is that the respondents in studies of noise annoyance typically make one annoyance rating using a scale with verbally defined steps. Obviously, a scale category such as *rather annoyed* may be interpreted in different ways as a result of individual and situational characteristics. The aim of the present study was to test different ways to reduce this problem by calibrating the annoyance ratings.

Berglund and Berglund [3] proposed Master scaling as a way of calibrating ratings. This method requires subjects to make magnitude estimations of the loudness of a reference noise at several levels. The individual functions relating sound level to loudness are determined, and their deviations from a Master function, regarding both intercept and slope, are used as a basis for calibrating the individual ratings. In laboratory studies of e.g., power line noise, Berglund *et al*. [4] demonstrated that the relation between sound level and loudness was dramatically increased by this calibration.

In this study [4], it was possible to present the target stimulus as one stimulus in a random series of reference stimuli. Such rating conditions could not be created in the present field study of noise at the work place. The work place noise was instead judged separately from the recorded reference stimuli. This is likely to be the only practical way to make these measurements in most field studies of noise annoyance at work. One

0022–460X/97/340523 + 09 $25.00/0/sv971021 

consequence of this procedural difference was that it was not possible to use magnitude estimation. A graphic scale with some verbally defined points was instead used. Another reason for using ratings on an absolute scale of annoyance was that they may provide a basis for discussions about what should be regarded as an acceptable sound level. This change of rating method to one with an upper numerical limit is likely to change the relation between annoyance and sound level. Using magnitude estimation, *logarithmic* ratings would be expected to be a linear function of sound level (dB) [5]. When using the graphic scale, this is likely to be true for the non transformed linear ratings, as is generally found with category rating methods [6].

Another important difference between the present study and those reported by Berglund and co-workers is that subjects rated annoyance and not loudness. The change from loudness to annoyance ratings may also affect the form of the function relating ratings to sound level; when annoyance is rated, it is quite possible that several levels above the hearing threshold may be given a rating of 0.

The Master scale calibration procedure must be adapted to these expected changes in the psychophysical function. The Master scaling involves a calibration for individual differences in both mean ratings and the slope of the psychophysical function. A simpler procedure is to calibrate only for differences in mean reference ratings of the reference stimuli. Berglund *et al*. [4] found that ratings became more closely related to sound level when the slope was also taken into consideration. It remains to be shown that this conclusion is valid for studies of the present type.

It is a rather cumbersome task to present recorded reference sounds for each participant, and such a procedure is impossible to use in noise surveys in which individual exposure levels are not measured. Thus, an alternative means of calibration is needed which also would be practicable in questionnaire studies. The present study tested one possible method; some well-known types of noise were described, and the subject was asked to rate how annoyed he would be if he were exposed to these noises at work. The work place ratings were calibrated for differences in means of these reference ratings.

## 2. METHOD

### 2.1. PARTICIPANTS AND WORK ENVIRONMENTS

The group consisted of 338 persons, 139 men and 199 women, with a mean age of 39·7 years (range: 17–64) from 23 different work places. The distribution of the participants at each work place was as follows: offices ($n = 75$), dental laboratories ($n = 71$), control rooms ($n = 56$), a library ($n = 10$), a laundry ($n = 25$), cashier's desk ($n = 12$), classrooms ($n = 13$), assembly work ($n = 17$), student's writing room ($n = 31$), a large scale kitchen ($n = 9$) and a kindergarten ($n = 9$). The remaining ten persons were evenly distributed at work places such as a lunch room at a public school, a public swimming pool and a repair shop. The sound level was below 85 dB(A) in all cases.

### 2.2. PROCEDURE AND EQUIPMENT

A noise level meter (Brüel & Kjær model 2231) with a Brüel & Kjær 4155 microphone was placed beside the participant's workplace, and a calibration preceded the period of measuring. The noise was recorded during a period of 15 minutes, during which the person worked as usual. Noise annoyance during the measurement period was rated on a 100 mm scale ranging from "not at all annoying" to "almost unbearable" and with five further verbally labelled points on the scale. The distances between the labelled points were empirically determined to correspond to differences in the strength of annoyance. The scale is therefore treated as a ratio scale. It has previously been used in a series of work place and laboratory studies [7, 8].

The same scale was also used for ratings of how annoyed the participants thought they would be if they were exposed to three sounds at work: The sound of a vacuum cleaner, a hand-held hairdryer and a sewing machine. The questionnaire furthermore contained questions about working conditions and different noise characteristics.

Finally, pink noises (reference stimuli) were presented to the participant from a digital audio tape recorder (TEAC DA-P20) through earphones (Sennheiser HD 250 linear II). The pink noise was presented twice at seven sound pressure levels (range 20–80 dB(A)), each with a duration of 15 s. The 14 stimuli were given in one random order with a pause of 5 s between the sounds. Participants made their rating of annoyance during the pauses.

### 2.3. SOUND ANALYSES

For all the recordings, the sound level, dB(A), was measured as the equivalent level over the final ten minutes of the recording. The first five minutes were treated as a habituation period. Zwicker loudness was also calculated using a real-time analyzer (Brüel & Kjær 2144) with an option for the purpose (Zwicker Loudness Option—Brüel & Kjær type 7638).

### 2.4. SCALING AND CALIBRATION OF ANNOYANCE RATINGS

The annoyance ratings of the recorded pink noise and imagined sounds were used as reference ratings, and were used to calibrate the ratings of noise annoyance at work.

Calibration for differences in mean ratings of the reference sounds was obtained by forming ratios of ratings of annoyance at work and of the reference sounds. Ratios were calculated using as denominators either single or the mean ratings of the imagined sounds or the mean ratings of four of the recorded pink noises (40–70 dB). The 40–70 dB range was chosen since the noise in the work places lay in this interval in almost all cases. To make these calibrated ratings directly comparable with the uncalibrated ones, the ratio was multiplied by 34, which was the grand mean of the annoyance ratings.

The linear regression of the mean pink noise annoyance ratings on sound level was calculated and used as a master function. The function was based only on the four highest sound levels, which yielded ratings above zero from all subjects. Inclusion of the two lowest levels resulted in a deviation from linearity, which simply reflected the fact that the annoyance threshold varied between individuals and that there were more zero ratings at 20 than at 30 dB. The same function, based on the levels given a non-zero rating, was also calculated for each individual. To obtain Master scale calibrated annoyance scale values, the individual annoyance functions were transformed to the Master function using the following formula given by Berglund *et al.* [4]: $y' = By/b - Ba/b + A$ where $y'$ is the perceived annoyance in Master scale units, $B$ is the slope of the Master function and $A$ its intercept. $y$ is the individually obtained rating of the work place noise and $b$ is the slope of the individual function relating annoyance to sound level of the pink noise. $a$ is the intercept for this individual function.

## 3. RESULTS

In connection with the measurement, notes were made of what source was the major determinant of the sound level. The participants also reported what noise source they had in mind when they made their ratings. Only the 206 cases for whom these two sources coincided were selected for the analyses of the relation between annoyance and sound level. This group is labelled *Noise source group* in the following.

In comparison with the previous laboratory studies by Berglund *et al*., the determination of the individual psychophysical functions rested upon very few ratings. A large error in the slope and intercepts would of course disfavour the Master scaling calibration. A further selection was thus made of persons for whom $r^2$ for the sound level-annoyance function was at least 0·90; this left 115 persons in a *Good fit group*.

Figure 1 shows mean ratings of the reference sounds as a function of sound level for the *Good fit group*. The fitted linear function was used as a Master function.

The strength of the relations between different scalings of annoyance and two exposure measures (dB(A) and Sone) was determined by calculating product moment correlations, with the annoyance measures expressed in both linear and logarithmic units. Seven subjects with zero ratings were excluded in the analyses of *log* values, five subjects in the *Good fit group* and two in the *Noise source group*. Correlations for *log* values are not given for the non-transformed and master scaled ratings, since these analyses yielded much lower correlations than those of linear values. For the other calibrated values, correlations for linear and logged values were of similar magnitude. Table 1 shows these correlations for the *Noise source group* and the *Good fit group*.

The correlation with non-transformed ratings was somewhat higher in the *Good fit group* than in the *Noise source group*. All calibration methods lead to substantially higher correlations with only minor differences between the methods. Calibration based on ratings of imagined sounds tended to yield somewhat lower correlations. The standard errors for the difference between correlations varied between 0·08 and 0·11. Thus, only the differences between non-transformed and calibrated scales were statistically significant ($t = 2{\cdot}00$–$3{\cdot}1$, $p < 0{\cdot}05$).

In Figures 2–4, the uncalibrated and calibrated annoyance ratings are plotted as a function of sound level for the *Noise source group*. It is evident from the figures that the scatter was considerably reduced by the calibration. Exclusion of outliers with a strong influence on the slope of the regression line did not change the relative differences between the correlations. Figure 5 plots the Master scale calibrated values against the non-transformed ratings. As shown by the figure, the lowest calibrated value was given to a subject who had rated his annoyance as 30, whereas many subjects who had rated themselves to be not at all annoyed received a fairly high calibrated value.

Both the ratings of the recorded and the imagined reference sounds were found to be negatively correlated with the sound level in the work place ($r = -0{\cdot}33$ and $-0{\cdot}21$, respectively). The mean ratings of the recorded and the imagined reference sounds ratings correlated moderately ($r = 0{\cdot}46$). The Master scale calibrated ratings and the ratings

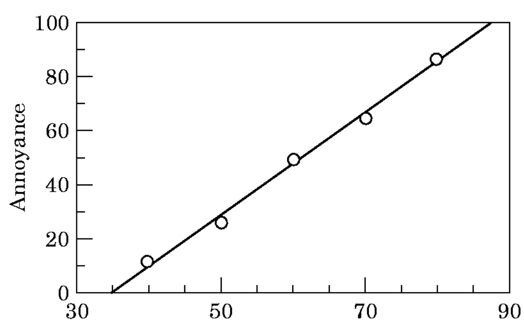


Figure 1. Mean ratings of the pink noise reference sounds for the *Good fit group* ($n = 115$). This function was used as the Master function in the Master function calibration of the annoyance ratings. $y = -65{\cdot}590 + 1{\cdot}8892x$, $R_2 = 0{\cdot}994$.

TABLE 1

*Product-moment correlations between different scalings of work place noise annoyance, sound level* (dB(A)) *and loudness* (Sone) *in the* "Noise source" *and* "Good fit groups"; *Master scaling was based on the* dB-*annoyance Master function shown in Figure* 1

| Annoyance scales | Noise source group ($n = 206$) dB(A) | Noise source group ($n = 206$) Sone | Good fit group ($n = 115$) dB(A) | Good fit group ($n = 115$) Sone |
|---|---|---|---|---|
| Non transformed ratings | 0·17 | 0·23 | 0·28 | 0·35 |
| Master scale calibrated values | 0·42 | 0·45 | 0·45 | 0·50 |
| Calibrations for differences by forming means with mean ratings of different reference sounds: The recorded pink noise, 40–70 dB(A) | | | | |
| linear values | 0·40 | 0·41 | 0·45 | 0·50 |
| *log* values | 0·44 | 0·47 | 0·41 | 0·43 |
| Imagined sound, hairdryer | | | | |
| linear values | 0·24 | 0·23 | 0·21 | 0·20 |
| *log* values | 0·34 | 0·33 | 0·34 | 0·33 |
| Imagined sound, sewing machine | | | | |
| linear values | 0·38 | 0·42 | 0·35 | 0·36 |
| *log* values | 0·42 | 0·42 | 0·42 | 0·42 |
| Imagined sound, vacuum cleaner | | | | |
| linear values | 0·34 | 0·38 | 0·31 | 0·37 |
| *log* values | 0·32 | 0·35 | 0·28 | 0·31 |
| Imagined sounds, mean of hairdryer, sewing machine and vacuum cleaner | | | | |
| linear values | 0·37 | 0·41 | 0·33 | 0·36 |
| *log* values | 0·38 | 0·37 | 0·35 | 0·36 |
| Imagined sounds, mean of sewing machine and vacuum cleaner | | | | |
| linear values | 0·38 | 0·43 | 0·35 | 0·40 |
| *log* values | 0·37 | 0·39 | 0·34 | 0·36 |

calibrated by the mean ratings of recorded sounds were fairly highly correlated ($r = 0{\cdot}87$). The correlations between Master scale calibrated ratings and the ratings calibrated by the ratings of imagined sounds were almost as high as the correlations between the latter and the ratings calibrated by the ratings of recorded sounds ($r = 0{\cdot}54$ and $r = 0{\cdot}58$).

### 3.1. INFLUENCE OF THE CALIBRATION ON THE EFFECTS OF SENSITIVITY-RELATED VARIABLES

Differences in mean ratings of reference sounds may partly be a result of differences in response styles and scale interpretation. However, they may also reflect real differences in sensitivity caused by both individual and situational characteristics. To test whether the calibration reduced individual differences of the latter type, the effects of sensitivity and two sensitivity-related variables (work load and the time spent in conversation) were tested both for uncalibrated and calibrated ratings. Table 2 shows that the effects of sensitivity

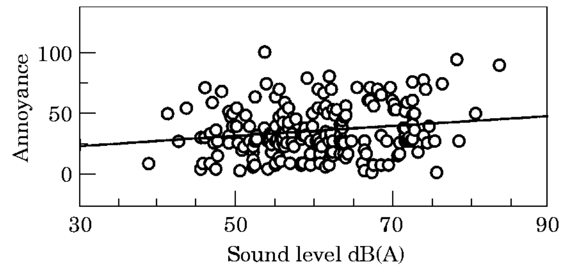


Figure 2. Uncalibrated annoyance ratings as a function of sound level in dB(A) in the *Noise source group* ($n = 206$). $y = 8{\cdot}0586 + 0{\cdot}42783x$, $R_2 = 0{\cdot}032$.

and work load, with one exception, were markedly reduced by all calibration methods, whereas the effect of conversation was affected in a less consistent way. The largest standard errors were in most cases found for the values calibrated using ratings of imagined sounds and the lowest for the non-transformed ratings.

## 4. DISCUSSION

The results clearly showed that the relation between sound level and annoyance was considerably strengthened when the ratings were calibrated. The proportion of the variance that was explained by the sound level thereby increased up to six times. The other main result was that there were no striking differences between the calibration methods in this respect. Master scaling did not yield a higher correlation than calibration for differences in mean level. Thus, no further improvement seems to be achieved by also calibrating ratings for individual slope differences.

Even though calibration based on ratings of the recorded and the imagined reference sounds yielded about the same correlations with sound level, the choice between them is not without consequence, as indicated by the moderate correlation between the two types of calibrated values. This is also shown by the fact that differences between groups differing in sensitivity and sensitivity-related variables are not affected similarly by the three calibration methods. Mean differences between sensitivity and work load groups are reduced but not at all eliminated by all calibration procedures. It should also be noted that calibrations increased the standard errors, and that this increase in most cases was most pronounced for values calibrated for mean ratings of imagined sounds. The effect of the time spent in conversation during the work day was not reduced by the calibrations based on recorded sounds but was totally eliminated by that based on imagined sounds. These results suggest that the response to the reference sounds does not reflect the individual

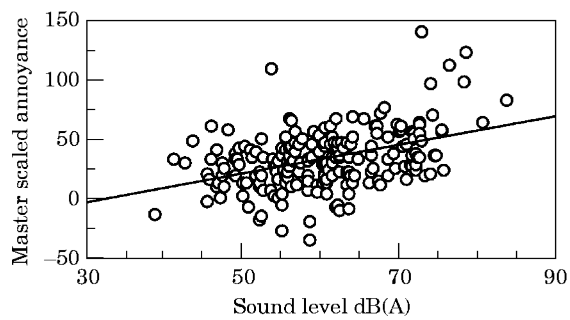


Figure 3. Master scale calibrated ratings as a function of sound level in dB(A) in the *Noise source group* ($n = 206$). $y = -40{\cdot}001 + 1{\cdot}2260x$, $R_2 = 0{\cdot}176$.

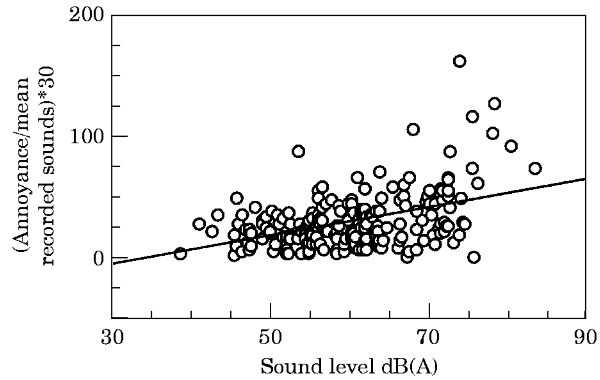


Figure 4. Ratings calibrated for differences in mean ratings of recorded reference sounds (pink noise) as a function of sound level in dB(A) in the *Noise source group* ($n = 206$). $y = -40{\cdot}824 + 1{\cdot}1747x$, $R_2 = 0{\cdot}189$.

sensitivity differences caused by situational factor considered in this question. They might also mean that the calibration based on ratings of imagined sounds adds too much to the error variance to make discriminations between some groups possible. Previous studies [9] also indicate that the method that uses imagined sounds is not always effective. It is therefore probably wise to use recorded reference sounds whenever possible.

Another conclusion from the effect of calibration on the group differences is that the calibration methods reduces differences in rating style, rather than differences in sensitivity.

Two persons reporting zero annoyance could receive widely different ratings after Master scale calibration. The calibrated ratings may also be higher among such persons than for persons who rated themselves to be rather annoyed. This consequence of the Master scaling procedure indicates that persons giving zero ratings cannot be handled by this calibration method. In studies of loudness, it might be argued that these persons should be excluded from the analysis, but this is hardly defensible in studies of annoyance.

Calibration for differences in slope did not improve the correlation with sound level. Three deviations from the procedure used by Berglund in her studies of the Master scaling procedure might have been critical for this result. First, ten minutes of work place noise was compared with 15 s periods of the reference noises. Secondly, the work place and reference sounds could not be presented in the same series of stimuli. Finally, it is possible that individual differences in annoyance functions are less generalizable than differences in loudness functions from one type of sound to another. However, it should be noted that

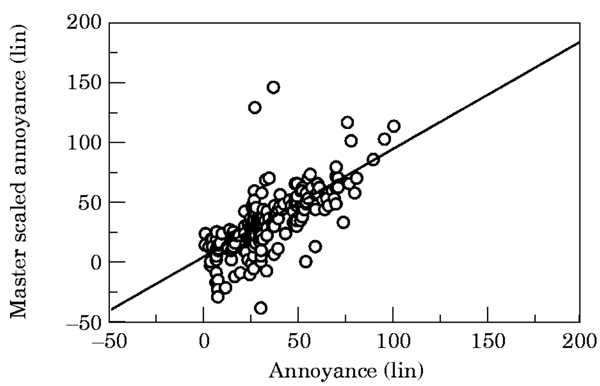


Figure 5. Master scaled ratings plotted as a function of uncalibrated ratings. $y = 2{\cdot}3421 + 0{\cdot}90340x$, $R_2 = 0{\cdot}488$.

TABLE 2

*Mean (standard errors) of uncalibrated and calibrated ratings in groups differing in sensitivity and critical situational characteristics in the* 'Noise source group' ($n = 206$); *p-values for the differences between groups from analyses of variance are also shown*

| Variables and response distributions (n) | Non-transformed ratings | Master scale calibrated values | Calibrated values based on ratings of recorded pink noise | Calibrated values based on ratings of imagined sounds |
|---|---|---|---|---|
| *Sensitivity* | | | | |
| 0 Less than most (32) | 30·4 (3·7) | 36·3 (5·5) | 36·0 (5·5) | 33·8 (6·1) |
| 1 About average (150) | 33·0 (1·7) | 31·6 (2·0) | 32·0 (1·9) | 33·9 (2·5) |
| 2 More than most (24) | 44·0 (4·1) | 38·4 (6·2) | 45·1 (7·7) | 38·5 (5·5) |
| | $p = 0{\cdot}03$ | $p = 0{\cdot}39$ | $p = 0{\cdot}07$ | $p = 0{\cdot}79$ |
| *Work load* | | | | |
| 0 Too Little (6) | 27·7 (11·5) | 24·0 (14·3) | 28·4 (13·1) | 37·8 (14·2) |
| 1 Somewhat (33) | 32·7 (3·8) | 30·8 (4·8) | 34·6 (5·1) | 32·9 (4·4) |
| 2 Neither/nor (152) | 33·3 (1·6) | 33·0 (2·2) | 33·5 (2·2) | 33·1 (2·5) |
| 3 Far too much (15) | 49·7 (4·9) | 42·6 (5·4) | 41·2 (5·0) | 49·2 (10·9) |
| | $p = 0{\cdot}02$ | $p = 0{\cdot}42$ | $p = 0{\cdot}70$ | $p = 0{\cdot}28$ |
| *Conversation (proportion of the work day)* | | | | |
| 0 Not at all (103) | 32·9 (2·2) | 32·4 (2·6) | 33·4 (2·7) | 32·4 (3·0) |
| 1 1/10 of the day (46) | 35·6 (2·6) | 32·8 (3·2) | 33·4 (2·7) | 34·9 (4·0) |
| 2 1/4 of the day (25) | 32·4 (3·5) | 28·4 (5·0) | 36·1 (7·9) | 41·2 (7·1) |
| 3 1/2 of the day (19) | 30·9 (4·9) | 27·3 (5·6) | 26·4 (3·9) | 34·5 (9·7) |
| 4 The whole day (13) | 47·5 (5·1) | 57·8 (10·2) | 49·6 (6·9) | 35·5 (6·3) |
| | $p = 0{\cdot}11$ | $p = 0{\cdot}01$ | $p = 0{\cdot}18$ | $p = 0{\cdot}80$ |

these deviations from Berglund's procedure are often both necessary and desirable in field studies of noise annoyance at work. The results therefore indicate a serious limitation of the use of this calibration method.

## REFERENCES


1. R. F. S. JOB 1988 *Journal of the Acoustical Society of America* **83**, 991–1001. Community response to noise: a review of factors influencing the relationship between noise exposure and reaction.
2. A. KJELLBERG, U. LANDSTRÖM, M. TESARZ, L. SÖDERBERG and E. ÅKERLUND 1996 *Journal of Environmental Psychology* **16**, 123–136. The effects of non-physical noise characteristics, ongoing task and noise sensitivity on annoyance and distraction due to noise at work.
3. B. BERGLUND and U. BERGLUND 1987 in *Environmental annoyance: characterization, measurement, and control* (H. S. Koelega, ed.), 29–44. Amsterdam: Elsevier. Measurement and control of annoyance.
4. B. BERGLUND, U. BERGLUND and S. LINDBERG 1983 *Department of Psychology, University of Stockholm, Report No.* 610. Master scaling of environmental loudness.
5. S. S. STEVENS 1957 *Psychological Review* **64**, 153–181. On the psychophysical law.
6. T. ENGEN 1971 in *Woodworth and Schlosberg's experimental psychology* (J. W. Kling and L. A. Riggs, editors), 47–86. Psychophysics II. Scaling methods. London: Methuen.
7. U. LANDSTRÖM, A. KJELLBERG and M. BYSTRÖM 1993 *Journal of Sound and Vibration* **160**, 533–542. Acceptable levels of sounds with different spectral characteristics during the performance of a simple and a complex non-auditory task.

8. U. Landström, A. Kjellberg, M. Tesarz and E. Åkerlund 1995 *Environment International* **21**, 265–275. Exposure levels, tonal components and noise annoyance in working environments.
9. A. Kjellberg, U. Landström, P. Löfstedt, P. Wide and E. Åkerlund 1988 *National Institute of Occupational Health*, *Solna*, *Sweden*, *Arbete och Hälsa* 1988: 30 Bedömning av upplevd bullerstörning i arbetet.