inter•noise 96
25th Anniversary Congress – Liverpool

INCE: 61

DIRECT LOUDNESS SCALING OF TRAFFIC NOISE IN NATURAL SETTINGS

J Hellbruck & A Zeitler

Environmental and Health Psychology, Catholic University of Eichstaett, Germany

1. INTRODUCTION

Several psychoacoustical studies have been conducted on the matter of continuous subjective evaluation of traffic noise. The procedures used in such experiments have been essentially based on the method of continuous judgement by category which was introduced by Namba and his colleagues in the seventies [1, 2]. This method has been adopted especially by German researchers. Thus, several procedures of continuous loudness respectively noisiness scaling are known. Namba and his colleagues prefer a 7-point-category scale usually ranging from „not noisy at all" to „intolerably noisy". The subjects are instructed to judge the instantaneous noisiness of sound by pressing one of the 7 buttons on a response box, which contains the seven categories. Weber [3] has linked the categorical judgement to an analogue tracking of the time varying loudness sensations by use of a potentiometer. Fastl [4] uses a continuous cross-modality technique. Subjects are to alter the length of a line on a computer screen according to loudness changes.

Our approach on the continuous subjective evaluation of time varying loudness of traffic noise is in some aspects different from the former procedures.

1. We adopted a scaling procedure which has recently become popular in the field of audiometry, at least in German speaking countries. This procedure is called category subdivision scaling (CS Scaling). It is used in the field of audiometry under the label „Würzburger Hörfeld" („Würzburg Hearing Field", WHF)). The CS Scaling was introduced by Heller [5; cf. also 6]. Contrary to the usual type of category scale, the categories here are finely graduated, allowing differentiation within each category. The scale has five verbally distinguished categories, each divided into 10 graduations. Thus, a 50-point scale results comprising the five categories, "very soft" (1-10), "soft" (11-20), "medium loud" (21-30), "loud" (31-40) and "very loud" (41-50). However, this

scale allows to go beyond 50 in order to express that the noise was perceived as unpleasantly or even painfully loud. The advantage of this procedure over other classical category scales, usually comprising 5, 7 or 9 categories, is that transitions between categories can be directly scaled instead of being calculated from repeated measurements and statistical analysis.

2. We do not evaluate the instantaneous loudness but the loudness of consecutive short-term intervalls. According to the results of auditory memory research, the long auditory store is lasting as long as 20s [7]. Our subjects were advised to give an integrated loudness judgement every 15s.
3. Studies on subjective evaluation of noise are mostly conducted in psychoacoustical laboratories using sounds played back from tapes. An advantage of laboratory studies is their high internal validity, because experiments can be repeated as often as desired. However, the external validity is unknown. If contextual factors in the natural surrounding would significantly influence loudness judgements, then we could not expect a satisfactory correlation between laboratory and field studies. In our study, we have also used our technique in the natural environment. We suppose that the CS Scaling is suited for use in the complex natural environment due to the combination of spontaneous categorization with daily used verbal categories and the within-category fine-grading which provides a high scaling resolution.

The questions proposed in our study are:

1. Are the distributions of consecutive loudness judgements comparable with those found in audiometrical loudness scaling sessions where typically short-lasting discrete sounds are used?
2. What is the correlation between energy equivalent sound pressure level and consecutive loudness judgements measured with CS Scaling?

2. EXPERIMENTS

We carried out a series of 3 experiments. In the first, short-term narrow-band noises and natural sounds were judged with the CS Scale (WHF). In the second, a recording of a consecutive sequence of a traffic noise situation was evaluated each 15s as described above. In the third experiment, the same procedure as used in the second experiment was adopted in the field.

In all experiments the same five volunteers (three females and two males, all with normal hearing) served as subjects. All subjects are students of our University. The ages ranges between 21 and 28 years.

Experiment I: Narrow-band noise and discrete natural sounds
In an audiometry chamber, 64 third-octave band noises with center frequencies at 0.5, 1, 2, and 4 kHz graduated in 5 dB steps were randomly presented through loudspeakers. The duration of the sounds was 2s. The subjects judged the loudness of the sounds using the CS Scaling. The procedure was identical with the WHF procedure. In the second part of this experiment, natural sounds, such as bird whistle, telephone ringing,

starting car engines etc. were presented. The duration of the sounds was about 3 - 4s. All sounds, i.e. the narrow-band noises as well as the natural sounds, in the respective dB steps are recorded on the Westra Audiometry CD #7. In addition, the CD contains CCITT noise for calibration purpose [8].

Both trials were repeated four times on different days by each subject.

In Figure 1 the data is shown. The main result is that the variability of the loudness judgements is nearly the same in both trials. Natural sounds did not cause a higher variability in the judgements. However, the central tendencies indicate a systematic difference. Natural sounds at the lower end of the dB scale are evaluated louder than the synthetic sounds.

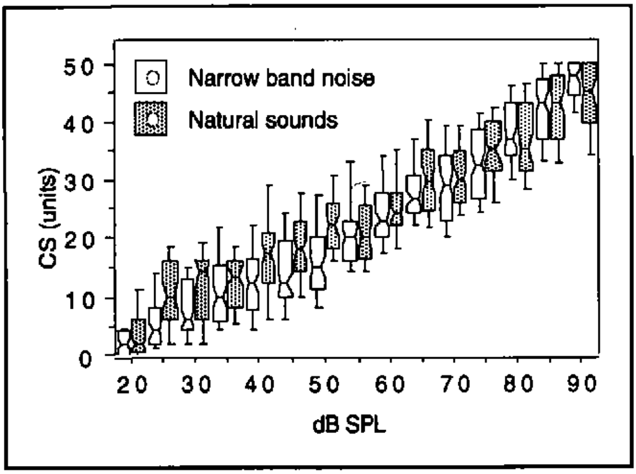

Fig. 1: Box plots of CS-Loudness Scalings of narrow-band noises and natural sounds, averaged over all frequencies. The 5 horizontal lines display the 10th, 25th, 50th, 75th and 90th percentiles of the data. The notches represent 95% confidence intervals around the median

Experiment II: Consecutive loudness scaling of recorded traffic noise

In the laboratory, the group of the five subjects was presented with a traffic noise scene which was stereophonically recorded and reproduced by DAT through two loudspeakers. The scene involved traffic at a gated railway crossing. Vehicle noise and a passing train were the main acoustic impressions. The 5 participants made loudness judgements every 15 seconds representing the 15-second period which had just passed. The numerical scaling values were immediately written down on a provided form. The equivalent sound pressure level (Leq,A) was calculated for each of the 15s periods. In order to determine retest-reliability, the same subjects repeated the scalings after two months.

In Figure 2 the correspondence between Leq,A and the CS-Loudness Scalings can be seen. It is noteworthy that the standard deviations of the loudness scalings are in correspondance to the semi-interquartils shown in Fig. 1. The correlation between these scalings and the retest was r = .95. Since there was a rather long period of time between test and retest, one can assume that the high correletion of r = .95 is not a pseudo correlation due to memory effects.

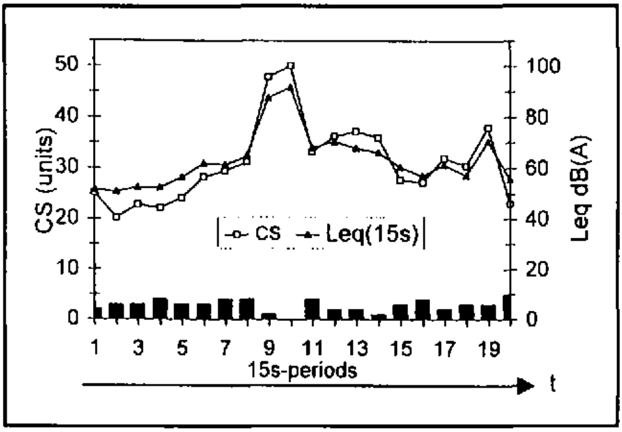

Fig. 2: Judgements of a 5min-section of the 15min sequence of the traffic noise scene played back in the labaratory. Triangles stand for the Leq calculated for each 15s-period (right Y-axis), squares for the mean CS (left Y-axis) over the respective 15s. Bars indicate the standard deviation of the CS judgements.

Experiment III: Consecutive loudness scaling in natural settings

In Experiment III the subjects basically evaluated traffic noise in the same way as they had done in Experiment II, but this time they did so in the real environment. The tests were carried out near a village on a main road at a gated level crossing and were located ca. 7 meters away from the edge of the road and also from the railway track. In fact, the setting was the same as the one used for the recording in Experiment II. The noise-level readings were obtained by a sound-level meter, type B&K 2231 with short-term Leq module, which registered and stored the Leq second by second. The 5 evaluators sat to the side of the sound-level meter to avoid disturbing the readings and gave their judgement every 15s as in Experiment II. A flashlight, which could be easily seen in daylight, signalled them when to make the judgement. Thus, they could keep their attention focussed on the sound environment.

Two assessment sittings were carried out at the test sight on each day appointed for the test, each lasting approximately 20 minutes. The assessments were carried out on five different days, always at the same time in order to keep the noise pattern similar which is primarily induced by the train timetable. Alltogether, we gathered judgements on ca. 800 15s intervals. In Figure 3 the data collected in all sittings are displayed with regard to the dB scale, in each case averaged over the scalings of the 5 observers. A curve was fitted to the data using a modified Fechner function. Two other curves are displayed which were fitted to the data shown in Figure 1.

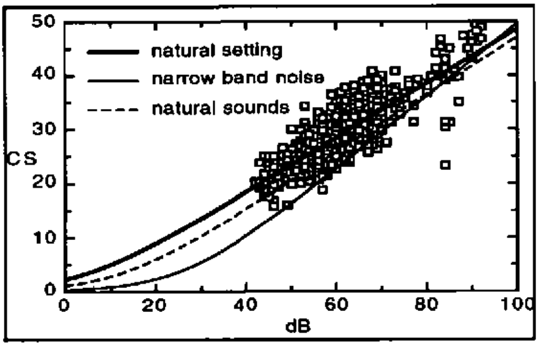

Fig. 3. CS data averaged over the judgements of the 5 observers in all sittings under natural conditions in regard to the dB scale. Compare the fitted curve with those from Experiment I (see also text).

From these results shown in Figure 3 one may conclude that loudness scalings of narrow-band noise do not predict the loudness scalings of real noise in the natural environment. However, the loudness scalings of the WHF-natural sounds evaluated in the laboratory, are better correlated with the scalings in the field. The systematics insinuated in Figure 3, however, need to be tested in additional experiments. What stands out in the data shown in Figure 3, is the fact that some of the mean judgements at the high end of the dB scale largely deviate from the fitted curve. What is the reason for causing such outlyers? Figure 4 illustrates this phenomenon by zooming a specific time section of one of the 10 sessions.

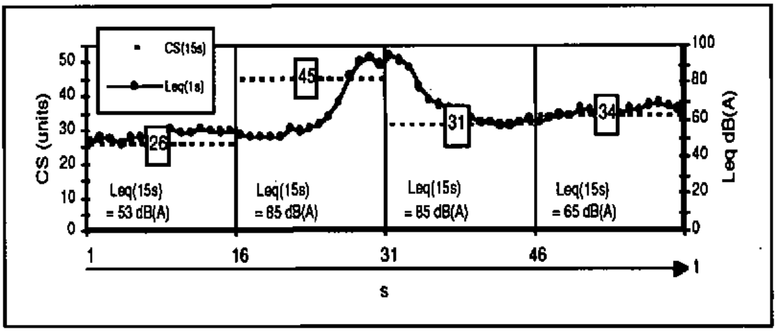

Fig. 4. The temporal course of the sound level of a train as it approaches, passes and recedes. The dots display 1-s-Leqs of 4 consecutive intervalls of 15s each. Though the approaching as well as the receding of the train (2nd and 3rd section) have the same Leq of 85 dB(A), the judgements differ by more than one category (numbers in boxes).

In this Figure four consecutive 15s-intervalls are displayed. Each point marks a 1s Leq . Here, a train is shown as it approaches, passes, and recedes. Though the Leqs of both the approaching and the receding of the train are equal, the loudness scalings differ drastically, by almost 15 points on the scale: The approaching train was judged as „very loud" whereas the receding train was judged as „between middle loud and loud".

3. CONCLUSIONS

From our study we draw the following conclusions:

The CS Scaling is an appropriate method even for loudness scalings of consecutive noise sequences. Intra- and interindividual variability of scalings is not higher than it is found in the well-established WHF. The question, however, is whether this applies to other than the well-trained observers in our study too. Therefore, additional experiments with observers who are inexperienced are needed.

WHF loudness judgements based on natural sounds seem to predict the loudness judgements in the field better than those judgements which are related to synthetic sounds.

The equivalent sound pressure levels are well correlated with the loudness judgements, at least for the relatively short intervals of 15s. However, even in such short periods of time loudness judgements are governed by specific contextual factors, such as the time course of sound pressure level or the sound pressure level immediately preceding the judgement [cf. also 9]. More data analysis than could be done so far, and presumably further experiments are necessary to explore details. By now, we can already state that the temporal structures of the periods to be evaluated have to be taken into account.

4. REFERENCES

[1] S. Namba, S. Kuwano and T. Nakamura, J. Acoust. Soc. Jpn., 34, 29-34 (1978).
[2] S. Namba and S. Kuwano, In: A. Schick, J. Hellbrück and R. Weber (eds.), Contributions to Psychological Acoustics. Results of the fifth Oldenburg Symposium on Psychological Acoustics (pp. 217-226). (BIS, Oldenburg, 1991).
[3] R. Weber, In: A. Schick, J. Hellbrück and R. Weber (eds.), Contributions to Psychological Acoustics. Results of the fifth Oldenburg Symposium on Psychological Acoustics (pp. 267-294). (BIS, Oldenburg, 1991).
[4] H. Fastl, In: A. Schick, J. Hellbrück and R. Weber (eds.), Contributions to Psychological Acoustics. Results of the fifth Oldenburg Symposium on Psychological Acoustics (pp. 205-216). (BIS, Oldenburg, 1991).
[5] O. Heller, Psychol. Beiträge, 27, 478-493 (1985).
[6] J. Hellbrück, D. Thomamüller und A. Zeitler, Audiol. Akustik, 34, 60-81 (1995).
[7] N. Cowan, Psychol. Bull., 96, 341-370 (1984).
[8] Westra Audiometrie Disc Nr. 7 (Welden, Westra electronic, 1989).
[9] G. Rosinger, C.W. Nixon and H.E. von Gierke, J. Acoust. Soc. Am., 48, 843-853 (1970).