NOISE ANNOYANCE STUDY IN THE CAB OF MOBILE WORK MACHINES

Hanna Järveläinen, Panu Maijala and Matti Karjalainen
Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Box 3000, 02015 HUT, Finland

Abstract

Noise annoyance in the cab of mobile work machines was investigated by subjective listening tests. A number of psychoacoustic indices were fitted into the test results by a linear regression model. A binaural system for evaluating sound quality was developed. It consists of tools and methods on how to record, edit using digital signal processing, and present sound samples binaurally to many subjects simultaneously. Also, a technique was developed for making binaural recordings using the real human head.

1. Introduction

The A−weighted noise level in the cab of a modern work machine is usually less than 85 dB, which is the lowest level considered harmful to hearing. Yet the A−weighted sound level is the most widely used noise measure today. This is paradoxic, since it is a poor descriptor of the perceived loudness and annoyance. For example, the A−weighted sound level underestimates the loudness of low frequency sounds as well as broadband noise and sounds that include strong tonal components [1].

We measured noise annoyance in the cab of work machines. Three types of work machines were investigated, each individually. At first, sound samples were recorded binaurally. Annoyance was then evaluated by listening tests. The test results were used as a basis for a mathematical model, by which annoyance can be objectively estimated without further listening tests. The models are based on sound level measures and psychoacoustic indices such as loudness, sharpness and roughness [5].

2. The system for evaluating sound quality

The binaural technology [2] originates in the philosophy that our perception of sound is controlled by the sound pressure signals at the eardrums. Thus authentic auditory experience is reproduced if the real life sound pressures are exactly reproduced at the listener’s eardrums. Today, most binaural recordings are done with an artificial head replicating the human head in average dimensions and details as well as approximate hardness and softness of skin and bone.

Part of the sound samples were recorded using the dummy head and part by the real head technique, i.e. inserting small microphones into the driver’s ears. This way recordings could be made during operation. The sound samples were edited using digital signal processing. In the listening experiment they were presented binaurally to many subjects simultaneously. The subjects answered by pressing the mouse buttons (see Fig. 1).

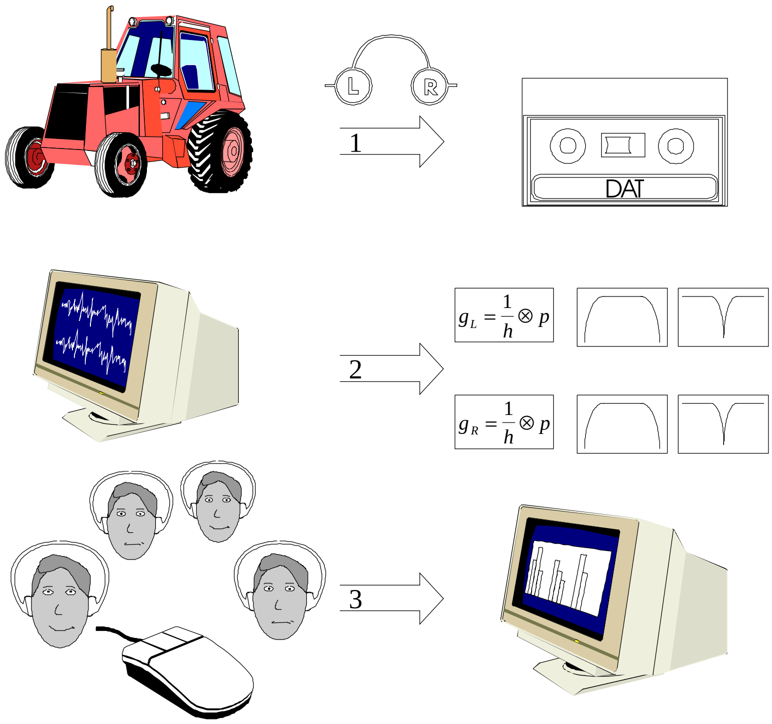

Figure 1. The binaural sound quality evaluation system. The sound samples are recorded binaurally to DAT tape (1). The samples are edited using digital signal processing (2) and presented to many subjects simultaneously in the listening test (3). The subjects give their answers by pressing the mouse buttons. The results are analyzed by statistical means.

3. Listening tests

The chosen sound samples of 5−second duration were analyzed in order to find signal components which might affect annoyance. Those components were then removed one by one from the original signal by digital filtering. For instance, tonal components were attenuated as well as broadband noise in the frequency band of 500...3000 Hz. The original samples were compared with the filtered ones in the listening experiment.

The test method was pairwise comparison. The listeners judged which sound sample was more annoying. The answer “equally annoying” was not allowed. Each sound was tested against every other in order AB and BA. The “annoyance score” for each sound sample was computed as the total number of times that a listener judged the sample more annoying than its pair. Although this method led to a numeric value, no numeric information could be obtained from the test results. The score was used merely to put the sound samples in order according to annoyance.

To determine a reliable estimate of the real annoyance order, the test results were analyzed by statistical means. To evaluate the performance of individual listeners the test was repeated, and the correlation coefficient was calculated between the two sets of judgments of each listener. The consistency of each listener’s judgments was estimated by the consistency coefficient.

There was a disagreement between listeners on certain test samples. This was verified by a nonparametric test of homogeneity of classified data [4]. The listeners were in all three cases quite easily divided into two groups who clearly disagreed on one or two samples. In one case an additional third group was formed.

The significance of the differences in annoyance scores between samples was investigated by nonparametric hypotheses testing. Each sample was assigned a rank according to its relative annoyance. Statistically equally annoying samples shared the same rank.

4. Mathematical modeling

A number of level measures and psychoacoustic indices were calculated for each sound sample. They were sound pressure level, A−, B−, C−, and D−weighted sound level, loudness level, sharpness, roughness, fluctuation strength, and tonality. The objective measures were fitted into the subjective test results by a linear regression model. A stepwise algorithm made it possible to use only a few appropriate indices. Several combinations of indices were tested.

Because the listeners disagreed, the results of the different groups had to be modeled separately. Since it was difficult to determine which group should be chosen as a basis for the actual model, it was decided to model both groups of each machine and investigate their differences.

5. Results

Only loudness and sharpness of the psychoacoustic indices seemed to affect annoyance remarkably. On the other hand, the sound samples were relatively similar in timbre, and the other indices did not vary enough to affect the annoyance order.

The A−weighted sound level was not a good descriptor of annoyance. In two of the three cases it was the worst single level measure and time−varying loudness (the loudness level that is exceeded 10% of the time) the best one. D−weighted sound level was second best. In the third case A− and D−weighted sound levels as well as loudness level were almost equally good. Typically, time−varying loudness alone explained about 75% of the variation of the test results, D−weighted sound level 68…74% and A−weighted level 52…65%. Better results were achieved by combining two or three indices. In all cases sharpness and time−varying loudness level were included in the model. The third index was either D− or A−weighted sound level or loudness level. In one case C−weighted sound level was added. The combination models explained 95…99% of the variation, but their performance was strongly dependent of the listener group.

The relation between D−weighted sound level, sharpness and annoyance is complicated, as the coefficients in the different models varied quite much. It seems that high frequencies affect annoyance more than has been thought. Both D−weighted sound level and sharpness emphasize high frequencies, but compared to the A−weighted sound level the D−weighted level stresses also low frequencies. Thus it cannot be stated that sharpness always increases annoyance. Instead, the spectral balance and the loudness proportions in different frequency bands may be relevant.

It should be noted that the annoyance models cover only the chosen noise problem and limited situations. Even the psychoacoustic indices fail to describe for instance transient sounds or sounds that become annoying in long term. The test does not reveal very detailed factors of annoyance but rather the annoying characteristics of the basic sound.

6. Conclusions

The in−cab noise annoyance was evaluated in three tractors and loaders by subjective listening tests. A system was developed for binaural sound quality evaluation. By using the real head technique, natural driving situations could be included in the listening test. Linear regression models were fitted into the test results. The models are linear combinations of traditional and psycho−acoustic indices. The modeling revealed that the A−weighted sound level is not the best measure for noise annoyance. Instead, loudness level and D−weighted sound level are better individual measures. By combining two or three indices even better results can be achieved.

References

1. Fastl H., (1994), *Psychoacoustics and Noise Evaluation,* Proceedings of 1994 Nordic Acoustic Meeting, Aarhuus, Denmark, p. 1−12.
2. Møller H., (1992), *Fundamentals of Binaural Technology,* Applied Acoustics, vol.36 p. 171−218.
3. Møller H., Jensen C.B., Hammershøi D., and Sørensen M.F., (1997), *Evaluation of Artificial Heads in Listening Tests,* AES Preprint number: 4404 (A1), Convention: 102, (1997 March 22−25).
4. Siegel S., (1956), *Nonparametric Statistics for the Behavioral Sciences,* McGraw−Hill.
5. Zwicker E. and Fastl H., (1990), *Psychoacoustics, Facts and Models,* Springer Verlag, Heidelberg, Germany.