Relation between the overall unpleasantness of a long duration sound and the one of its events : application to a delivery truck

Emilie Geissner[a] and Etienne Parizet[b]

[a]Laboratoire Vibrations Acoustique - INSA Lyon, 25 bis avenue Jean Capelle, Bâtiment Saint-Exupéry, F-69621 Villeurbanne cedex, France
[b]Laboratoire Vibrations Acoustique, Insa Lyon, 25 bis, av. J. Capelle, 69621 Villeurbanne Cedex, France
emilie.geissner@insa-lyon.fr

The goal of this study was to investigate the link between the continuous assessment of an unstationary long duration sound composed of several sound events and the overall judgments of the corresponding excerpts. For that purpose, a sound sequence of a delivery truck was evaluated by 16 listeners during a laboratory test: first, subjects had to continuously quantify the perceived annoyance of the sequence by moving a sliding cursor along a five levels graduated scale and then give a global rating by using the same scale. In a second step, listeners had to express their overall judgment of unpleasantness for eight excerpts of the delivery sequence. As previously shown for loudness by Kuwano and Namba (1985), the global rating of the unpleasantness of long sound could not be estimated by the arithmetic mean of the continuous assessment. It also appeared that the overall judgment corresponds to the arithmetic mean on the local values of unpleasantness of each main sound event. This last result was similar to the conclusions of Hellbrück et al. (2001) for the loudness scaling of traffic noise.

1 Introduction

As a part of a project conducted by Volvo3P (Lyon) and some other partners, a study was conducted in order to identify the most unpleasant sounds emitted by a medium-size lorry during a delivery. To assess such a long sound sequence, it was decided to use the continuous assessment method, previously used to evaluate the loudness of traffic noise [1] [2], car interior noise [3] or aircraft noise [4]. The analogical continuous scale method presented by Weber [5] was chosen: while listening to the sound sequence, the subject had to slide a cursor along a five level graduated scale. The applicability of the method to the subjective criterion of unpleasantness was first checked on synthetic time-varying events [6] and on a real sound sequence [7]. In both cases, it was shown that this method can be used to evaluate unpleasantness of long duration sounds. In addition to this continuous assessment, an overall judgment can be asked to subjects, who have to give a global rating of the sound sequence they just listened to. The same scale and experimental device can be used for the both assessments. In many studies about loudness evaluation [1][2] [8][9], overall scores appear to be higher than the corresponding time averages of the continuous assessment answers. A questionnaire after an experiment about the loudness of time-varying pink noise [10] enabled listeners to express their way of global rating: for some of them, their overall judgment was mainly based on the average of their continuous evaluation whereas others mentioned their global rating was mainly influenced by sound events with high levels of loudness. Kuwano and Namba [1] suggested taking into account the loudest sound events of the sequence. But, in the case of traffic noise, Hellbrück et al [2] showed that the overall judgment of loudness corresponded to the arithmetic mean of the ratings of separated sound event. Conversely, when the global rating was given five minutes after the end of the listening, its value was close to the time average of the continuous assessment answer. Authors concluded that the continuous assessment reflects the immediate loudness sensation (stored in the short term memory), whereas overall judgment corresponds to an mean of the loudness of the main events (stored and recalled from the long-term memory).

The goal of this study was to investigate the link between the continuous assessment and overall judgment in the case of annoyance using a sound sequence of a lorry: firstly, subjects continuously assessed the annoyance of the delivery sound sequence by sliding a cursor along a scale graduated from “not at all unpleasant to ”extremely unpleasant”. Then they attributed a global rating of annoyance on the same scale. In a second step, subjects had to give an overall judgment of annoyance for eight chosen excerpts of the delivery sequence. The overall judgments of each excerpt could then be compared to the continuous assessment and global rating of the whole sequence.

2 Protocol

2.1 Stimuli

The vehicle used in this study was a truck from the Renault Trucks Midlum product range, equipped with a side door and a hydraulic tailgate. The delivery took place on an ISO 10844 [11] acoustic track, sound being recorded using a Bruel & Kjaer dummy head placed next to the track. Recording conditions are based on the PEAK report specifications [12] and adapted to represent real conditions of a normal delivery. The whole sound sequence lasted 2 min 30 (wave file, 16 bit- signed, stereo, recorded at 44100 Hz and down sampled at 22050 Hz). Eight excerpts corresponding to identified unpleasant sound events [7] were extracted from this sequence:

- E1: vehicle arrival on the delivery stage, with braking and switching off the engine in front of the dummy head (15 s),
- E2: opening and closing of the driver’s cab door (7.3 s),
- E3: opening of the tailgate (14.5 s),
- E4: unloading the hand pallet truck from the tailgate platform + pallet truck rolling outside the trailer + loading the pallet truck onto the tailgate platform (16.5 s),
- E5: closing of the tailgate (14.5 s),
- E6: opening and closing of the passenger’s cab door (9 s),
- E7: reverse motion + idle (31.7 s),
- E8: acceleration with gear changes and departure from the delivery area (18.5 s).

2.2 Experimental device and protocol

Although the experiment was carried out in a laboratory, listeners were told to figure out a real life situation (“in the early morning, a truck arrives for a goods delivery next to your home”). The evaluated subjective criterion was thus "annoyance" [13]. Subjects could hear sounds through headphones (Stax Lambda Pro).

For the continuous assessment, the listener had in front of him a box with a cursor sliding along a continuous scale divided in five categories, build with the Fields et al. criteria [14]: "not at all annoying", "slightly annoying", "moderately annoying", "very annoying" and "extremely annoying". The listener's task consisted in adjusting the cursor position so that his sensation could be represented on the semantic scale. He had to perform that task two times, in order to get familiar with it and the sound sequence. The same device was used for the overall judgement: after having listened to a sequence, the subject was asked to adjust the cursor position according to his evaluation of the annoyance of the sequence.

Then, the eight excerpts were randomly presented to the listener, who had to evaluate the overall annoyance of each of them using the same device.

The duration of the experiment was 15 min. The jury was made of 16 people between 21 to 51 years old (average: 30.5), with 2 women and 14 men. Most of them leaved in urban or sub-urban areas. Only 3 of them were naive subjects (no previous participation in any subjective study); the 13 others had already participated to continuous evaluation experiments and could then be considered as experienced subjects.

3 Results

3.1 Perceived annoyance of the whole delivery sound sequence

Only answers to the second continuous assessment were analyzed since the listener answer was considered as more reliable than the one of the first assessment. A mean answer was calculated by averaging the individual answers to the continuous assessment. On Fig. 1, the most annoying sound events are easily identified: the first one was the departure of the truck with the acceleration (at 130 s), followed by the reverse motion and the idle (between [100-125 s]), then the truck arrival with the braking and the switching off the engine [0-10 s], the shock of the pallet truck when loading onto the tailgate platform (at 59 s), the closing of the tailgate [62-76 s] and the handling of the pallet truck outside the trailer [44-58 s]. In a previous study [7], the unpleasantness of a longer version of the delivery sound sequence was evaluated by continuous assessment and free verbalizations. By comparing the continuous assessment results of the two studies, it appears that the same annoying sound events were identified.

Nevertheless, some sound events that could not be so easily identified in [7] are recognized on Fig. 1: the handling of the cab doors at the beginning and at the end of the sequence ([16-27 s] and [90-97 s]), the opening of the tailgate [30-45 s], pallet truck handling outside [43-60 s]. This "better accuracy" could be explained by the way of continuously answering. In a study about the evaluation of loudness of passing-by cars, Weber [5] showed that a first group of subjects tried to convert as quickly as possible their instantaneous impression of loudness by a rapid motion of the sliding cursor along the scale to express their judgment, whereas a second group moved the cursor less frequently along the scale (only when large loudness alterations happened). It seemed these last listeners answered "by integrating". In the present study, by looking at the individual continuous assessment answers, it appears that there was no integrating-type answer: every subject tried to convert his/her perception of annoyance as quickly as possible without "integrating". This could be explained by the composition of the jury: most of the subjects were experienced ones, i.e. familiar with the use of the sliding cursor to express their "instantaneous" sensation. This was confirmed by comparing the composition of the jury (rate of naive listeners) and the rate of integrating answers of the two studies (see Table 1). The way of answering (by integrating or not) might be due to the training: the more subjects familiar with the assessment method, the less integrating answers.

The comparison of the two experiments showed that main results were the same, thus confirming that this method can be used to identify the most unpleasant sound events of a long duration sound sequence

Experiment	Integrating answers	Naive subjects
Geissner et al. (2006) [7]	22%	66%
Present study	0%	18%

Table 1- Rates of integrating answers and naive subjects for two continuous assessment studies on the delivery sound sequence.

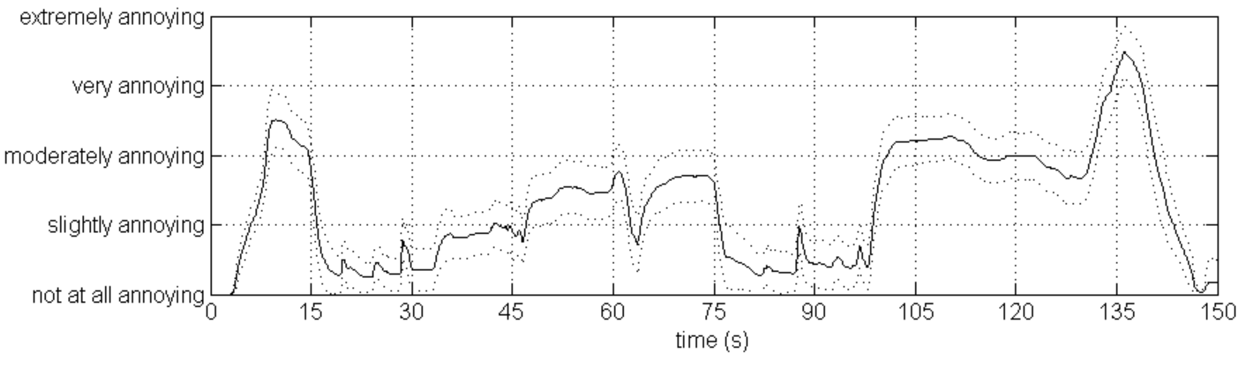

Fig. 1 – Mean continuous assessment answer of the whole delivery sound sequence (standard deviation in dotted lines)

3.2 Perceived annoyance for the eight excerpts

Results from the continuous assessment

For each excerpt, measuring the maximum local value of the corresponding interval on the individual continuous assessment answers enables to calculate a continuous assessment score by excerpt. A mean continuous assessment scores is then calculated by averaging the individual scores. As illustrated on Fig. 2, the most annoying sound source was the engine, followed by the pallet truck and the tailgate.

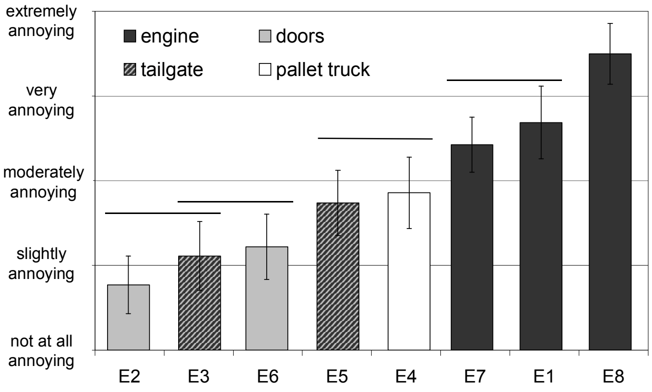

Fig. 2 – Classification of the 8 excerpts by maximum values to the continuous assessment with corresponding 95% confidence intervals. Non-statistically significant differences ($p \geq 0,05$) are represented by horizontal lines.

Table 2 enables to compare the ranking of the 8 excerpts with the ranking of the corresponding sound events of [7]. The same classification was obtained: the engine was the most annoying source, then the handlings of the pallet truck and the tailgate. The stability of the results confirms the repeatability and thus the reliability of the continuous assessment for the evaluation of the subjective criterion of annoyance.

Results from the overall judgment of the 8 excerpts

Fig. 3 shows the mean overall judgment scores of the 8 excerpts computed from the individual global ratings of annoyance. The main source of annoyance was still the engine (E8 with the final acceleration, E1 with the truck arrival and E7 with the reverse motion and the idle), then the pallet truck handling outside the trailer (E4), the closing of the tailgate (E5), and finally the opening of the tailgate (E3) and the cab doors handling (E2 and E6).

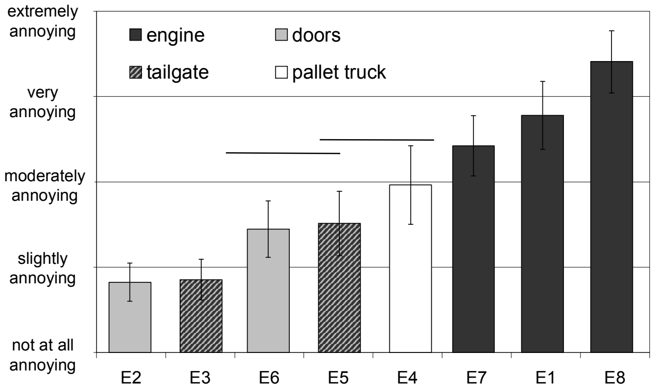

Fig. 3 – Classification of the 8 excerpts by overall judgments scores with corresponding 95% confidence intervals. Non-statistically significant differences ($p \geq 0,05$) are represented by horizontal lines.

By comparing Fig. 2 and Fig. 3, it is obvious that the classification obtained for the continuous assessment method and the overall judgment are the same (a non-statistically significant Student T-test confirms this result).

Excerpts number	Rank in [7] (among 28)	Rank in the present study
E8	1	1
E1	3	2
E7	4	3
E4	11	4
E5	13/14	5
E6	18	6
E3	19/20/21	7
E2	25	8

Table 2 - Comparison of the classification obtained with the continuous assessment evaluation for two studies assessing the same sound sequence.

3.3 Comparison of the perceived annoyance evaluated by the two assessment methods

On Fig. 4 (open circles), the continuous assessment score of each excerpt is plotted versus its corresponding overall judgment. For a given excerpt, the maximum value of the corresponding continuous assessment provides a correct estimation of the corresponding overall judgment.

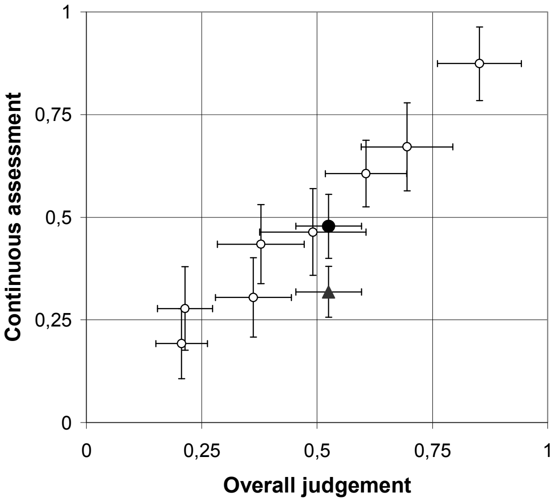

Fig. 4 – Scores of annoyance of the continuous assessment versus the overall judgment, with the corresponding 95% confidence intervals: open circles for the 8 excerpts, filled circle for E_{OJ} (or E_{CA}) against W_{OJ} and filled triangle for W_{CA} against W_{OJ}. The x- and y-axis correspond to the five point analogical continuous scale labeled from “not at all annoying” (0) to “extremely annoying” (1).

To investigate the relation between continuous assessment and overall judgment for the annoyance criterion, four values were computed:

- E_{OJ} : the arithmetic mean of the overall judgment of the 8 excerpts (0,47 on the numeric scale),
- E_{CA} : the arithmetic mean of the maximum local values of the continuous assessment of the 8 excerpts (0,47 on the numeric scale),
- W_{OJ}: the mean overall judgment of the whole sequence, i.e. the arithmetic mean of the individual overall judgments, noted. Its value is close to the label “moderately annoying” (0,52 on the numeric scale),
- W_{CA}: the time average of the continuous assessment answer of the whole sound sequence (0,32 on the numeric scale).

On Fig. 4, the filled triangle represents W_{CA} plotted against W_{OJ}: the time average of the continuous assessment answer is statistically lower than the overall judgment value of the whole sound sequence. This result is similar to the ones obtained in previous studies about loudness evaluation [1], [2], [4], [9]: the time average of the continuous assessment does not enable to estimate the global rating of a long sound sequence.

Since the overall judgment of a given excerpt is correctly estimated by the maximum value of the corresponding continuous assessment, it is obvious that the corresponding arithmetic means, E_{OJ} and E_{CA} are equivalent.

On Fig. 4, the filled circle represents E_{OJ} plotted against W_{OJ}. The difference between W_{OJ} and E_{OJ} is not statistically significant: the overall judgment of a long duration sound sequence could then be obtained from the global ratings of the main sounds events parts of the sequence. This result is similar to the conclusions of Hellbrück et al. [2]: when the overall judgment is expressed just after listening to the whole sequence, it is equivalent to the arithmetic mean of the global ratings of the main sound events separately evaluated.

Meanwhile, the difference between the W_{OJ} and E_{CA} is not statistically significant: on Fig. 4, the dark circle could also represents E_{CA} against W_{OJ}. The overall judgment of annoyance of the whole sequence could thus be calculated based on the arithmetic mean of the maximum local values of the continuous assessment.

4 Conclusion

The comparison of the present study with a previous one showed that not only were the same most annoying sound events identified, but that the annoyance classification and the perceived annoyance scores were equivalent as well. This first result confirmed the reliability and the validity of the continuous assessment method for the evaluation of the specific criterion of annoyance. Moreover, the comparison between results from a jury exclusively made of naive subjects and from an other one mainly composed of experienced listeners showed the influence of the training on the continuous assessment. The “by-integrating” way of answering also seemed linked to the composition of the jury.

The other main result was that, the overall judgment of annoyance for a long duration sound sequence could not be estimated by the simple time average of the continuous assessment. The comparison of the overall and continuous scores of annoyance for the main sound events showed the link between annoyance global rating and continuous assessment local scores: the overall judgment of annoyance corresponds to the arithmetic mean of the maximum values of annoyance measured on the continuous assessment answer.

Acknowledgments

This research was supported by a grant from the Région Rhône Alpes. The authors would like to thank Volvo3P for the technical and human support provided to record the delivery sequence.

References

[1] S. Kuwano and S. Namba, “Continuous judgement of level-fluctuating sounds and the relationship between overall loudness and instantaneous loudness”, *Psychological Research*, 47, 27–37 (1985).

[2] J. Hellbrück, T. Kato, A. Zeitler, A. Chick, S. Kuwano and S. Namba, “Loudness scaling of traffic noise : perceptual and cognitive factors”, *Proceedings of the 17th International Congress on Acoustics*, Rome, Italy (2001).

[3] S. Kuwano, S. Namba and Y. Hayakawa, “Comparison of the loudness of inside car noises from various sound sources in the same context”, *Journal of Acoustics Japanese Society*, 18(4), 191–195 (1997).

[4] S. Namba and S. Kuwano, “Continuous judgments of noise events”, *Proceedings of the 5th Oldenburg*

Symposium on Psychological Acoustics, Oldenburg, Germany, 217–226 (1991).

[5] R. Weber, “The continuous loudness judgement of temporally variable sounds with an ”analog” category procedure”, *Proceedings of the 5th Oldenburg Symposium on Psychological Acoustics*, Oldenburg, Germany, 267–289 (1991).

[6] E. Geissner, E. Parizet, “Continuous assessment of the unpleasantness of a sound”, *Acta Acustica united with Acustica*, vol. 93(3), 469-476 (2007).

[7] E. Geissner, E. Parizet, N. Nosulenko, “Perception of delivery truck noise”, *Proceedings of Euronoise 2006*, Tampere, Finland (2006).

[8] S. Kuwano, J. Kaku, T. Kato and S. Namba, “The experiment on loudness in field and laboratory: an examination of the applicability of laeq to mixed sound sources”, *Proceedings of InterNoise 97*, Budapest, Hungary (1997).

[9] H. Fastl, “Evaluation and measurement of perceived average loudness”, *Proceedings of the 5th Oldenburg Symposium on Psychological Acoustics*, Oldenburg, Germany (1991).

[10] S. Springer, R. Weber, A. Schick, “Instantaneaous and overall loudness of temporally variable pink noise”, *Proceedings of the 7th Oldenburg Symposium on Psychological Acoustics*, Oldenburg, Germany, 91-98 (1997).

[11] ISO10844: “Acoustics -- Specification of test tracks for the purpose of measuring noise emitted by road vehicles”, *International Organisation for Standardization* (1994).

[12] A.M. van Noort, P.P. Kooyman and M.G. Dittrich, “Measurement methods for peak noise during loading and unloading”, *TNO Report* (2003).

[13] B. Berglund, U. Berglund and T. Lindvall, “Scaling loudness, noisiness, and annoyance of aircraft noise”, *Journal of the Acoustical Society of America*, 57(4), 930–934 (1975).

[14] J.M. Fields, R.G. De Jong, T. Gjestland, I.H. Flindell, and R.F.S Job, “Standardized general-purpose noise reaction questions for community noise surveys: research and a recommendation”, *Journal of Sound and Vibration*, 242(4), 641–679 (2001).