



HAL
open science

Continuous assessment of the unpleasantness of a sound

Emilie Geissner, Etienne Parizet

► **To cite this version:**

Emilie Geissner, Etienne Parizet. Continuous assessment of the unpleasantness of a sound. *Acta Acustica united with Acustica*, 2007, 93, pp.469-476. hal-00849426

HAL Id: hal-00849426

<https://hal.science/hal-00849426>

Submitted on 31 Jul 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Continuous assessment of the unpleasantness of a sound

short title: continuous assessment

Emilie Geissner*, Etienne Parizet

Laboratoire Vibrations Acoustique, INSA de Lyon, 25 bis avenue Jean Capelle, 69621 Villeurbanne Cedex, France

November 21, 2006

Abstract

The goal of the present study was to estimate the accuracy of a continuous assessment method used for sounds of long duration. Such sounds were created by mixing stationary low-frequency random noise and twelve events of shorter duration (1 s to 11 s). These events were obtained from three stationary unpleasant sounds, the amplitude of which being varied in four different ways. Long duration sounds were presented to listeners who had to continuously assess their unpleasantness by moving a cursor sliding along a five-levels scale (from 'not at all unpleasant' to 'extremely unpleasant'). A second experiment, using a paired comparison procedure, gave a reference of the unpleasantness of the three sounds and the background noise used to synthesize the whole stimuli. The comparison of the results of both experiments allowed to confirm the validity of the continuous assessment, though its results were slightly dependent on the duration of the events.

*e-mail:emilie.geissner@lva.insa-lyon

1 Introduction

Over the past forty years, most of listening test experiments used for the evaluation of sound quality had dealt with sounds of short duration (typically less than 15 s.). However, real sounds in everyday life are usually long and unstationary (railway and road traffic, inside car noise, speech, music. . .). Some specific rating methods were therefore proposed to evaluate the instantaneous perception of such long and unstationary sounds. One of them is the continuous assessment method, which consists in continuously evaluating a perceptual parameter, such as loudness or brightness, while listening to the sound sequence. Most studies about continuous assessment were related to loudness evaluation and used rating scales of various types (categorical, analog and analog-categorical). First of all, Kuwano and Namba used a categorical scale to evaluate the loudness of traffic noise [11, 13], inside car noise [15], noise from aircrafts [17] and other real sources [18]. The subject had to push one of several computer keys that were attributed labels from 'very soft' to 'very loud'; this task had to be repeated as soon as the subject noticed any variation of his/her sensation while listening to the sound. Kuwano and Namba emphasized the fact that this task was rather easy for the listeners and could be realized without any specific training. However, the main drawback of that procedure appeared to be that the categories were not sensitive enough to accommodate small variations in hearing sensation. To overcome this drawback, Kuwano and Namba [14] also used an analog scale, as well as Fastl et al. [5] in studies comparing railway and road traffic noise loudness. In that case, the subject had to adjust the length of a line on a screen so that it represented his/her perception of the instantaneous loudness. Using this method, Kato et al. [12] proved that the limitation of subjects' motor ability could prevent them from relating rapid variation of loudness (as those from a hammering sound for example). Another device, developed by Weber [29], combined the two previous scales: the listener assessed the loudness of traffic noise by sliding a cursor along an analog scale graduated with five categories. This device was also used in studies from Hellbrück et al. [10] (loudness of road and railway noises, a procedural difference being that the scale was divided in seven categories, each of them being sub-divided

in ten levels) and Susini et al. [23, 25] (loudness of pure tones of varying amplitudes). The latter authors also used a cross-modal matching device with force feedback [24]: the listener had to adjust a muscular force sensation to the perceived loudness of urban environment noises and interior car sounds. Continuous evaluation of more complex sound attributes have been conducted less frequently: examples can be given about speech quality [7, 8], brightness of various sounds [9] or pleasantness of sound in a driving bus [20]. Research on music perception has also used that technique: Namba et al. [19] asked listeners to evaluate several descriptors while listening to a symphonic music excerpt. Other studies continuously evaluated two dimensions (arousal and valence) that can explain the emotional situation of a listener [16, 22].

These last examples suggested that continuous evaluation of pleasantness could be realized by subjects. Yet, some questions about the accuracy of this method could be raised. In the case of loudness evaluation, by using 1kHz pure tones with time-varying levels and different rates of level variations (for example, the level could be linearly increased between 60 and 80 dB SPL in 2, 5, 10 or 20 seconds), Susini et al. [25] made clear that the continuous evaluation varied with the increase rate: the longer the ramp, the greater the loudness estimation. For sounds made of an increasing and decreasing ramp, it also appeared that the evaluation at the end of the signal was different from the one at its beginning, though the physical levels were identical. Part of these effects could be related to loudness perception, as other authors showed that loudness estimation of plateau at the end of increasing or decreasing sweeps could vary according to the sweep rate [2, 3]. But they could also be due (at least partly) to some motor limitation of the subjects: while detecting very slow loudness variations, listeners might not be able to precisely describe these variations. In the study from Parizet et al. [20], another effect was noticed about impulsive sounds (as door closure sounds): the evolution of the subject's answer was quicker when the sound appeared than when it stopped. In the case of two successive impulsive sounds, this effect could lead to a higher evaluation of the second one. All these points questioned the validity of the analog categorical

continuous assessment: if the rate of the evolution of a sound were a too important bias, the use of this method for the detection of most unpleasant events would be of few accuracy. So the aim of the present study was to evaluate the accuracy of that method when used for unpleasantness assessment; for that purpose, the unpleasantness of three sounds of different timbres were assessed using the analog categorical continuous procedure and also a paired-comparison experiment, used as a reference.

2 Experiment

2.1 Stimuli

2.1.1 Sounds

According to Terhardt and Stoll [26], Aures [1] or Zwicker [31], roughness and sharpness are two elementary auditory attributes of sound that reduce pleasantness and sound quality. With this in mind, three typical sounds were synthesized, each at the sampling rate of 44.1 kHz. Two of these were amplitude modulated pure tones, with a rough character: the first one a 1 kHz tone modulated at 70 Hz and the second one a 4 kHz tone modulated at 50 Hz. In both cases, the modulation factor was 1. The third sound was a 8 kHz pure tone, i.e. a very sharp sound. Each of these principal sounds was mixed with a low frequency random noise (white noise band-pass filtered between 20 and 200 Hz), of the same loudness level as the signal itself. In the following, the three mixed signals will be labeled as S1, S2 and S3. In order to reduce the influence of loudness to unpleasantness, these three sounds were adjusted for similar loudness values. The loudness was evaluated according to ISO 532B standard using the MTS Sound Quality software (version 3.7) from the recordings realized with a dummy head (Bruel & Kjaer type 4100), the headphones used for sound presentation (Sennheiser HD600) being placed on it. The values of loudness, roughness and sharpness of the three signals are presented in Table 1. It should be noted that the roughness value of S1, 1.88 asper, and S2, 0.945 asper, seem high, but roughness

increases with loudness [32].

2.1.2 Temporal envelopes

The high-frequency content of a selected signal, i.e. its principal sound, was multiplied by a particular temporal envelope, which provided an 'event' of 17 second duration. Three different 'single event' envelopes were employed each being a sequence of 5 distinct intervals: 1) zero value, 2) linear increase to unity, 3) unity, 4) linear decrease to zero and 5) zero. The following values were used for different intervals (see table 2). E1 and E3 were inspired by real events (for example, E1 by the closing of a door and E3 by the arrival of a vehicle); E2 represented an intermediate between these two extreme cases. Yet one envelope, E4, named as 'double event' was used: this envelope corresponded to two single events envelopes separated only by a 2 second interval of zero value.

2.1.3 Long sound sequence generation

By using three sounds and four envelopes, twelve events could be obtained. For each subject, these events were added together in a random order, which provided a 204 s long sequence. Due to the differences in the envelopes intervals, the shortest time between two consecutive events was ≈ 6 s (in the case of two successive long single events) so that the listener had enough time to complete his/her evaluation of an event before the next one occurred.

2.2 Listening tests protocol

Each subject had to realize two listening tests. Half of the jury started with the paired comparison test and the other half with the continuous assessment of the 204 s sequence. The two procedures are presented in the following sections.

2.2.1 Paired comparison test

Three seconds long excerpts of the stationary sounds S1, S2 and S3 and the background noise alone (with the same duration) were used as stimuli. Sounds were played over a set of headphones (Sennheiser HD600) in a listening room isolated from outside. First of all, the four sounds were presented to the subject; then the 6 pairs corresponding to that number of sounds (which represents half of the matrix without the diagonal) were built using a Ross series [4]. Before building the series, the order of sounds was randomly rearranged, in order to eliminate any order effect in the pair series. Also, the last pair of the series was presented again at its beginning as a training one. After listening to each pair, the subject had to answer the question 'Which is the most unpleasant sound?' (translation of 'Quel est le son le plus désagréable?') using an analog categorical scale with five labels ('A is much more unpleasant than B', 'A is more unpleasant than B', 'A and B are equally unpleasant' and so on).

2.2.2 Continuous assessment method

The analog continuous scale method presented by Weber [29] was retained with the same device as the one used by Parizet et al. [20]. The listener held a box with a cursor sliding along a scale divided into five categories, selected according to criteria from Fields et al. [6]: 'Not at all unpleasant', 'Slightly unpleasant', 'Moderately unpleasant', 'Very unpleasant', 'Extremely unpleasant'. The cursor controlled a potentiometer which made it possible to modify the amplitude of a 1 kHz sinusoidal signal produced by an external generator. The listener's task was to adjust the position of the cursor so that his/her sensation could be represented by the semantic scale. The signal thus controlled was sampled and recorded by the audio sound card of a PC (at the sampling rate of 8 kHz). At the same time, the audio sequence was played over the headphones in the listening room. At the end of the assessment, the envelope of the signal amplitude was calculated using Hilbert transform and down-sampled at 50 Hz. This envelope represented the instantaneous position of the cursor.

The continuous evaluation session was split up into five steps. First, the task to be performed was explained to the listener. Then the whole sound sequence was presented to the subject so that he/she could become familiar with it before the actual assessment. Thirdly, he/she made a first evaluation while listening again to the sequence. Then he/she was asked to evaluate the difficulty and the duration of that task, by answering to the two questions: 'how difficult was that task?' (the possible answers being 'very difficult', 'difficult', 'nor easy nor difficult', 'easy', 'very easy') and 'how long did you feel it was?' (the possible answers being 'very short', 'short', 'nor short no long', 'long', 'very long'). The goal of questioning the subject about the perceived duration of the task was to have an idea about its acceptability. Finally, the sound sequence had to be continuously evaluated once more. This last continuous evaluation had two purposes: first, having two continuous assessment answers, correlation coefficient could be computed and could be used to verify the test repeatability. Secondly, the first assessment was considered as a training period for the listener. It was then assumed that during the second assessment, the listener was familiar with the evaluation method and so that his answer was more reliable.

2.2.3 Listeners

A total of 37 people, 30 men and 7 women, aged from 18 to 57 (average: 26) participated in the experiment. They did not report any hearing problem.

3 Results

3.1 Feasibility of the continuous assessment

The feasibility of the task and the reliability of the listener were evaluated from answers to the short questionnaire and the correlation coefficients computed between the two continuous assessments. As shown on figure 1, the continuous evaluation was described as a rather easy task (1a),

the duration of which being acceptable (1b). By looking at the individual unpleasantness evaluations, it appeared that five listeners did not achieve the task. Either they did not understand the instructions, or they could not use the sliding cursor as required. It was decided to remove those listeners from the jury. The correlation coefficient between the first and the second answer was then calculated for each remaining subject. Figure 2 presents the distribution of correlation coefficient values; all listeners had a correlation coefficient of more than 0.6, which indicated a strong relation between their two evaluations and was similar to what had been observed in a previous study [20]. In the following results, only answers to the second assessment were used, because, as previously mentioned, they were assumed to be more reliable.

3.2 Various ways of continuously assessing

As Weber showed about loudness evaluations [29], the way in which each listener interpreted the task he/she had to carry out and the way he/she used the analog categorical scale varied from one subject to another. The majority of subjects translated the loudness level experienced as immediately as possible into a movement of the response cursor when three of them altered the position of the cursor on the scale less frequently and responded by integrating. In the present study, this second way of answering was not identified, but four types of answers were identified from the global trend of the answers. Figure 3 presents four examples of typical answers. For five listeners, the background unpleasantness level was zero. Thus emerging peaks only corresponded to the occurrence of emerging events (see figure 3a). The second type (nine listeners, figure 3b) consisted in fixing the reference level of the background noise at a constant level over the entire sequence. The first two groups of listeners probably understood that the background noise was the same at all time and thus rated its unpleasantness as a constant value. They might have focused their attention to the events assessment only.

The third type was defined from six listeners who gave an answer which pointed to a global up-

ward trend: the background noise seemed to be perceived as more and more unpleasant, though the unpleasantness level of the background noise remained constant between two events (figure 3c). For such subjects, the increase of unpleasantness at the time of an event may be more relevant than the maximum value obtained at the plateau of the event. It could be noted that the inverse trend was not identified, i.e. no listeners evaluated the background noise as less and less unpleasant during the sequence. Regarding the last type of answers, no particular trend appeared: those twelve listeners might have taken the background noise unpleasantness into account but they did not seem to pay attention in retrieving a precise level after each isolated event (figure 3d).

Various attempts to classify such answers were conducted: numerous hierarchical cluster analysis were performed, based on listeners answers or corresponding temporal spectra but none of them lead to a satisfactory classification.

3.3 Unpleasantness scores

After this overall inspection of the continuous assessment answers, the post-processing was focused more precisely on the twelve events. Thus, for each listener's answer, five specific data were collected for every event: the maximum of the listener's answer to the event, the assessment of the background noise (i.e. the evaluations of the zero values intervals 1 and 5 in table 2), the precise time at which the listener began to react to the increase ramp of the envelope (i.e. interval 2) and the precise time at which he/she finished to react to the decrease ramp of the envelope (i.e. interval 4).

Overall unpleasantness of each sound type (S1, S2, S3 and BG) was derived from the continuous evaluations by averaging the maxima of individual answers for the corresponding events (i.e., for each listener, four data were taken into account). Results are plotted in bright gray bars on figure 4 (mean scores within their 95% confidence interval). The sharp sound S3 was rated

as the most unpleasant one, followed by the two rough sounds S1 and S2, the background noise being the least unpleasant. Such scores were also computed from the results of the paired comparison experiment. A mean preference matrix was calculated by averaging individual matrices. Then scores were computed using a Thurstone model (case V) [28, 27]; these values and their 95% confidence intervals (obtained from a bootstrap method) are shown on figure 4 in dark gray bars. The Thurstonian model was chosen because it provided a way to build an interval scale from paired comparisons, which was similar to results obtained from the continuous assessment. As shown on figure 4, the same classification arose from the two experiments, which proved that the continuous evaluation provided a reliable evaluation of the unpleasantness of sounds.

Some predictive models were used, as sensory pleasantness [1, 32], unbiased annoyance (UBA) [32] or psychoacoustic annoyance (PA) [30] but only sensorial consonance model proposed by Terhardt and Stoll [26] could predict the subjective evaluation. This is not surprising as loudness (which was similar between the three sounds of that experiment) is not an input parameter of that model, while it is of major importance in the other models. Sensorial consonance takes into account roughness, sharpness and tonality. It is defined by its opposite, the disagreement (W^-):

$$W^- = \sqrt{S^2 + 0.25.R^2 + 0.1.(T^-)^2} \quad (1)$$

where R is the roughness, S the sharpness, T- the non-tonality, i.e. the complementary value of the tonality T. Values of the Terhardt and Stoll disagreement are presented on table 3. In the experiment related in [26], tonality of sounds was fixed at 1 or 0 based upon an arbitrary estimation made by the authors. Here it was estimated at 1 for S1, S2 and S3 and 0 for the background noise. The agreement between unpleasantness scores and Terhardt and Stoll model values was good ($R=0.9$) since roughness and sharpness were prominent in the three unpleasant sounds of the experiment.

3.4 Influence of the temporal envelope on the unpleasantness level

The next part of the analysis was devoted to the influence of temporal envelopes on listeners' evaluations. First of all, for single events only, the importance of the increase ramp duration of the envelope (corresponding to interval 2 in table 2) was checked. The set of unpleasantness values extracted from individual answers was examined through an analysis of variance (repeated measures) in order to evaluate the relative influence of the sound type and the temporal envelope on unpleasantness. It appeared that for single events, the influence of the temporal evolution was significant ($F(2, 29) = 24.635, p < 0.001$) but was smaller than the one of the type of sound ($F(2, 29) = 72.041, p < 0.001$). That effect can be seen on figure 5: the longer the increase of signal level, the greater the increase of unpleasantness at the maximum level of the signal. Listeners probably tried to follow the increasing curve of the event instead of really representing their perceived unpleasantness. The same phenomenon was noticed by Susini et al. [25] in their study of the continuous loudness estimation of 1kHz pure tones: the maximum values of continuous judgments increased with the duration of the linear level ramp for an identical dynamic range. The influence of the increase time on the unpleasantness level could be due to some loudness changes; as demonstrated by Canévet et al. [3] or Schlauch [21], the longer the ramp, the greater the loudness increase. Nevertheless, from results from Canévet et al., it appeared that, in the case of 1 kHz or 4 kHz pure tones with level increasing from 45 to 75 dB during 1.8, 10 or 50 s, this effect was not significant (while it was for sounds decreasing in level). Therefore the results of the present study cannot be explained by such loudness effects. As can be seen in figure 5, there was no significant interaction between these two influent factors (sound type and temporal envelope).

The second point concerns the two-peaks events (E4). Unpleasantness values obtained for the second peak were slightly higher than for the first peak; the difference, though it was slight (approximately one tenth of a category of the scale), was significant ($F(1, 25) = 6.2; p < 0.05$). This difference is more important than the one observed in the case of inside bus noise [20]

(approximately one sixth of a category) though, in that previous study, the two impulsive events (opening and closing of doors) were separated by a longer time interval (10 seconds).

3.5 Complementary results

The data collected from individual answers were used to compute two values that characterized the listener's answer to an event. The first one was the *reaction time*, defined as the time elapsed between the beginning of an event and a modification of the listener's evaluation (figure 6a). Similarly, the *relaxation time* was defined as the time elapsed between the end of the event and the moment when the listener's answer reached a stable value, corresponding to the evaluation of the background noise only after the event (figure 6b).

Reaction times were calculated only for events with E1 and E4 temporal envelopes since the precise time a listener began to react to the increasing ramp of E2 and E3 events was hardly detectable. Attention was then paid on the difference between the two peaks of the double event E4. The jury was made up of 25 listeners who detected the beginning of both peaks. Results of the ANOVA proved a statistically significant difference ($F(2, 22) = 41.101, p < 0.001$) between the reaction times of the two peaks: the averaged value is 1.07 s. for the first event and 0.89 s. for the second one, the influence of the type of sound being neglectable. Listeners were less surprised by the occurrence of the second event than they were for the first one.

Relaxation times were very difficult to determine for four listeners; the analysis was then based on the data obtained from 28 listeners. In the case of single events, the temporal envelope appeared to be the only influential parameter ($F(2, 25) = 5.098, p < 0.01$), though its influence could not be characterized: for the S3 sound type, a clear upward trend was observed (the longer the increasing ramp duration, the greater the relaxation time), but this did not hold for S1 and S2. In the case of E4 (double short event), the relaxation time was smaller for the second event

than for the first one ($F(1, 24) = 12.217, p < 0.01$), as it had been noted for reaction times. A less significant interaction between the two factors, i.e. sound type and temporal envelope, also appeared ($F(2, 24) = 3.802, p < 0.05$).

4 Conclusion

The main goal of the study was to know if the most unpleasant events of a long sound sequence could be identified using the continuous evaluation method. Previous studies have shown that this method can be used to loudness estimation, as a strong correlation was found between averaged answers and classical sound measures as A-weighted level or Zwicker's loudness. As such standard estimations do not exist for unpleasantness, the present study used a paired comparison test as a reference in order to estimate the accuracy of unpleasantness evaluations by a continuous assessment. For the three sounds used in the experiment, unpleasantness could be correctly measured from a continuous evaluation using an analog-categorical scale, in spite of the bias due to the temporal evolution of the signals (between the shortest and the longest ramp duration, the difference of unpleasantness had the order of magnitude of one category of the scale). The reason for that bias is not yet clearly understood: it seems to be too important to be related to a loudness increase (which, as shown in the literature, is very small for such ramp durations). It might be possible that listeners, beside achieving the task, have unconsciously tried to follow the increasing ramp of the envelope. For real long sound sequences, this result gives an idea of the accuracy that can be expected from that evaluation method.

5 Acknowledgements

This study was financially supported by a grant of the Région Rhône-Alpes.

References

- [1] W. AURES. Berechnungsverfahren für den sensorischen wohlklang beliebiger schallsignale. *Acustica*, 59:130–141, 1985.
- [2] G. CANEVET and B. SCHARF. The loudness of sounds that increase and decrease continuously in level. *Journal of the Acoustical Society of America*, 88(5):2136–2142, 1990.
- [3] G. CANEVET, R. TEGHTSOONIAN, and M. TEGHSONIAN. A comparison of loudness change in signals that continuously rise or fall in amplitude. *Acta Acustica united with Acustica*, 89:339–345, 2003.
- [4] H.A. DAVID. *The method of paired comparison*. Oxford University Press, New York, 1988.
- [5] H. FASTL, S. KUWANO, and S. NAMBA. Assessing the railway bonus in laboratory studies. *Acta-Acustica Acustica*, 17(3):139–148, 1996.
- [6] 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.
- [7] L. GROS and N. CHATEAU. Instantaneous and overall judgments for time-varying speech quality: assessments and relationships. *Acta Acustica united with Acustica*, 87:367–377, 2001.
- [8] M. HANSEN and B. KOLLMEIER. Continuous assessment of time-varying speech quality. *Journal of the Acoustical Society of America*, 105(5):2888–2899, 1999.
- [9] D. HEDBERG and JANSSON C. Continuous rating of sound quality. Master thesis, Karolinska Institutet, 1998.
- [10] J. HELLBRÜCK, A. ZEITLER, and M. GOLD. Continuous loudness scaling of traffic noise in outdoor settings and under laboratory conditions. In *InterNoise 1997*, Budapest, HUNGARY, August 1997.

- [11] I. HIRANO, A. KINOSHITA, H. SHIBUYA, S. NAMBA, and S. KUWANO. Subjective evaluation of car interior noise in a consecutive series of driving situations. In *InterNoise 1994*, pages 897–900, Yokohama, JAPAN, August 1994.
- [12] T. KATO, S. NAMBA, and S. KUWANO. Continuous judgment of level-fluctuating noise. In *InterNoise 1994*, pages 1081–1084, Yokohama, JAPAN, August 1994.
- [13] S. KUWANO and S. NAMBA. Continuous judgment of level-fluctuating sounds and the relationship between overall loudness and instantaneous loudness. *Psychological Research*, 47:27–37, 1985.
- [14] S. KUWANO and S. NAMBA. Continuous judgment of loudness and annoyance. In *Fetchner Day, 6th Annual Meeting of the International Society for Psychophysics*, pages 129.–172, Würzburg, GERMANY, 1990.
- [15] 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.
- [16] C.K. MADSEN. Emotional response to music as measured by the two-dimensional crdi. *Journal of Music Therapy*, 34:187–199, 1997.
- [17] S. NAMBA and S. KUWANO. Continuous judgments of noise events. In *5th Oldenburg Symposium on Psychological Acoustics*, pages 217–226, Oldenburg, GERMANY, 1991.
- [18] S. NAMBA, S. KUWANO, H. FASTL, T. KATO, and J. KAKU. Estimation of reaction time in continuous judgment. In *18th International Congress on Acoustics*, Tokyo, JAPAN, April 2004.
- [19] S. NAMBA, S. KUWANO, T. HATOH, and M. KATO. Assessment of musical performance by using the method of continuous judgement by selected description. *Music Perception*, 8(3):251–276, 1991.

- [20] E. PARIZET, N. HAMZOU, L. SEGAUD, and J.R. KOCH. Continuous evaluation of noise uncomfort in a bus. *Acta Acustica united with Acustica*, 89:900–907, 2003.
- [21] R.S. SCHLAUCH. A cognitive influence on the loudness of tones that change continuously in level. *Journal of the Acoustical Society of America*, 92(2):758–765, 1992.
- [22] E. SCHUBERT. *Measurement and Time Series Analysis of Emotion in Music*. PhD thesis, University of New South Wales, Sydney, Australia, 1999.
- [23] P. SUSINI. *Perception évolutive et globale de sons non stationnaires*. PhD thesis, Université du Maine, 1999.
- [24] P. SUSINI and V. MAFFIOLO. Loudness evaluation of urban soundscapes by a cross-modal matching method. In *ASA & Forum Acusticum joint meeting*, Berlin, GERMANY, March 1999.
- [25] P. SUSINI, S. McADAMS, and K.S. BENNETT. Global and continuous loudness estimation of time-varying levels. *Acta Acustica united with Acustica*, 88(3):536–548, 2002.
- [26] E. TERHARDT and G. STOLL. Skalierung des wohlklangs von 17 umweltschallen und untersuchung der beteiligten hörparameter. *Acustica*, 48:247–253, 1981.
- [27] L. THURSTONE. A law of comparative judgment. *Psychological Review*, 34:273–286, 1927.
- [28] L. THURSTONE. The method of paired comparison for social values. *Journal of Abnormal and Social Psychology*, 21:384–400, 1927.
- [29] R. WEBER. The continuous loudness judgment of temporally variable sounds with an ”analogue” category procedure. In *5th Oldenburg Symposium on Psychological Acoustics*, pages 267–289, Oldenburg, GERMANY, 1991.
- [30] U. WIDMANN. A psychoacoustic annoyance concept for application in sound quality. *The Journal of the Acoustical Society of America*, 101(5):3078, 1997.

- [31] E. ZWICKER. A proposal for defining and calculating the unbiased annoyance. In *5th Oldenburg Symposium on Psychological Acoustics*, pages 187–202, Oldenburg, GERMANY, 1991.
- [32] E. ZWICKER and H. FASTL. *Psycho-acoustics - facts and models, 2nd edition*. Springer-Verlag, Berlin, 1999.

List of Tables

- | | | |
|---|--|----|
| 1 | Sound quality metrics computed with the MTS Sound Quality software for the 1 kHz tone modulated at 70 Hz (S1), 4 kHz tone modulated at 50 Hz (S2), 8 kHz pure tone (S3) and the background noise (BG). Roughness and sharpness are based on the Aures [1] works. | 19 |
| 2 | Composition of each 'single event' based on a sequence of 5 intervals: interval 1 corresponds to a period of zero value of the sound, interval 2 to a linear increase to the unity, interval 3 to unity, interval 4 to a linear decrease to zero and interval 5 to zero. | 19 |
| 3 | Unpleasantness of the four sounds of the present experiment, predicted according to the model proposed by Terhardt and Stoll [26]. S1 is the 1 kHz tone modulated at 70 Hz, S2 the 4 kHz tone modulated at 50 Hz, S3 the 8 kHz pure tone and BG the background noise. | 19 |

Sound label	Loudness (ISO 532B) (soneGF)	Sharpness (acum)	Roughness (asper)
S1	52	0.631	1.880
S2	49.6	0.837	0.945
S3	50.1	1.191	0.241
BG	40.7	0.538	0.271

Table 1: Sound quality metrics computed with the MTS Sound Quality software for the 1 kHz tone modulated at 70 Hz (S1), 4 kHz tone modulated at 50 Hz (S2), 8 kHz pure tone (S3) and the background noise (BG). Roughness and sharpness are based on the Aures [1] works.

envelope	name of the event	interval 1	interval 2	interval 3	interval 4	interval 5
<i>E1</i>	short single event	7.9 s	0.1 s	1 s	0.1 s	7.9 s
<i>E2</i>	medium single event	5 s	5 s	1 s	0.1 s	5.9 s
<i>E3</i>	long single event	3 s	10 s	1 s	0.1 s	2.9 s

Table 2: Composition of each 'single event' based on a sequence of 5 intervals: interval 1 corresponds to a period of zero value of the sound, interval 2 to a linear increase to the unity, interval 3 to unity, interval 4 to a linear decrease to zero and interval 5 to zero.

Sound type	Disagreement (W^-)
S1	1.13
S2	0.96
S3	1.19
BG	0.63

Table 3: Unpleasantness of the four sounds of the present experiment, predicted according to the model proposed by Terhardt and Stoll [26]. S1 is the 1 kHz tone modulated at 70 Hz, S2 the 4 kHz tone modulated at 50 Hz, S3 the 8 kHz pure tone and BG the background noise.

List of Figures

1	Ratings results: (a) difficulty of the continuous evaluation and (b) test duration.	21
2	Distribution of the correlation coefficients between the two continuous evaluations of each listener.	22
3	Four types of listener continuous assessment answers. (a) <i>constant and null basic level</i> , (b) <i>constant basic level</i> , (c) <i>increasing basic level</i> and (d) <i>random basic level</i> . (The scale is divided into five categories. 0: not at all unpleasant, 0.25: slightly unpleasant, 0.5: moderately unpleasant, 0.75: very unpleasant, 1: extremely unpleasant.)	23
4	Unpleasantness mean scores evaluated by paired comparison test (dark gray bars) and by continuous evaluation (bright gray bars) with 95% confidence intervals. (The scale is divided into five categories. 0: not at all unpleasant, 0.25: slightly unpleasant, 0.5: moderately unpleasant, 0.75: very unpleasant, 1: extremely unpleasant.)	24
5	Perceived unpleasantness for the three sound types (<i>S1</i> , <i>S2</i> and <i>S3</i>) and temporal envelopes (<i>t1</i> , <i>t2</i> and <i>t3</i>). (The scale is divided into five categories. 0: not at all unpleasant, 0.25: slightly unpleasant, 0.5: moderately unpleasant, 0.75: very unpleasant, 1: extremely unpleasant.)	24
6	Temporal data definitions: (a) reaction time and (b) relaxation time.	25

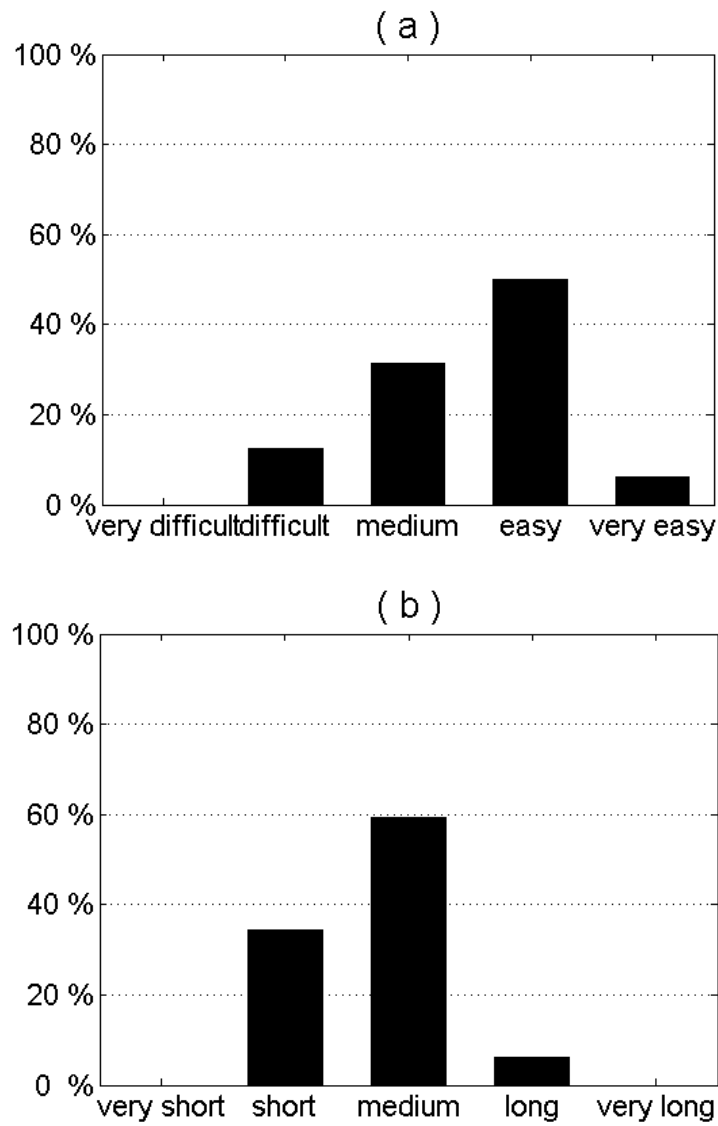


Figure 1: Ratings results: (a) difficulty of the continuous evaluation and (b) test duration.

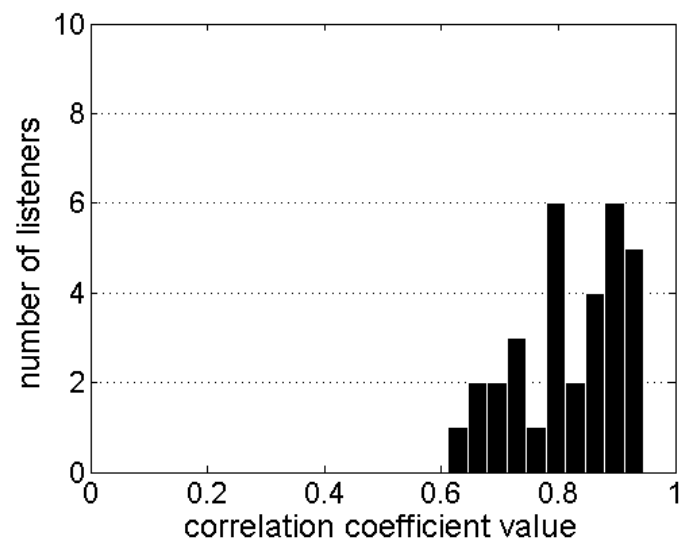


Figure 2: Distribution of the correlation coefficients between the two continuous evaluations of each listener.

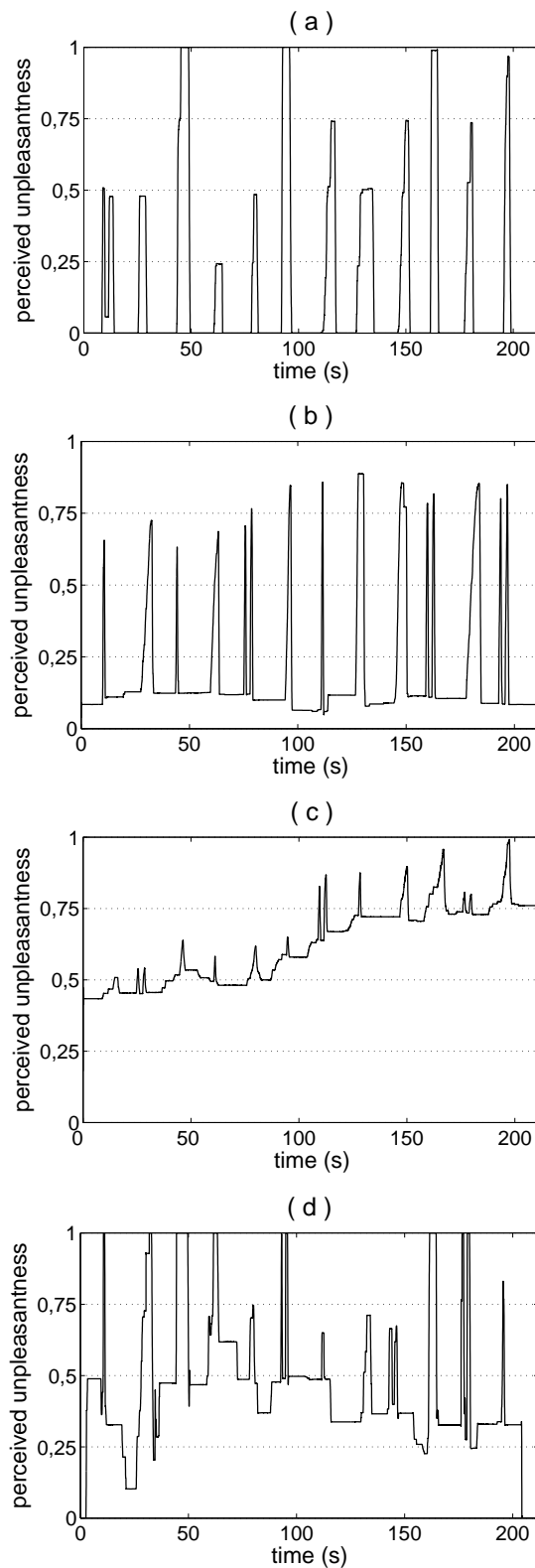


Figure 3: Four types of listener continuous assessment answers. (a) *constant and null basic level*, (b) *constant basic level*, (c) *increasing basic level* and (d) *random basic level*. (The scale is divided into five categories. 0: not at all unpleasant, 0.25: slightly unpleasant, 0.5: moderately unpleasant, 0.75: very unpleasant, 1: extremely unpleasant.)

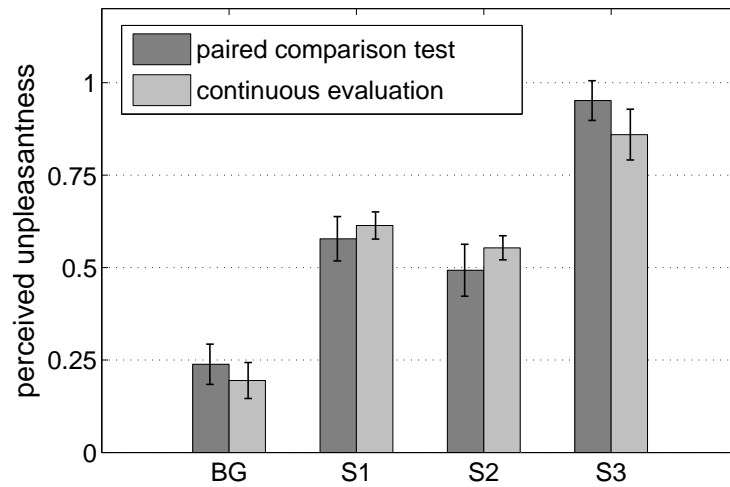


Figure 4: Unpleasantness mean scores evaluated by paired comparison test (dark gray bars) and by continuous evaluation (bright gray bars) with 95% confidence intervals. (The scale is divided into five categories. 0: not at all unpleasant, 0.25: slightly unpleasant, 0.5: moderately unpleasant, 0.75: very unpleasant, 1: extremely unpleasant.)

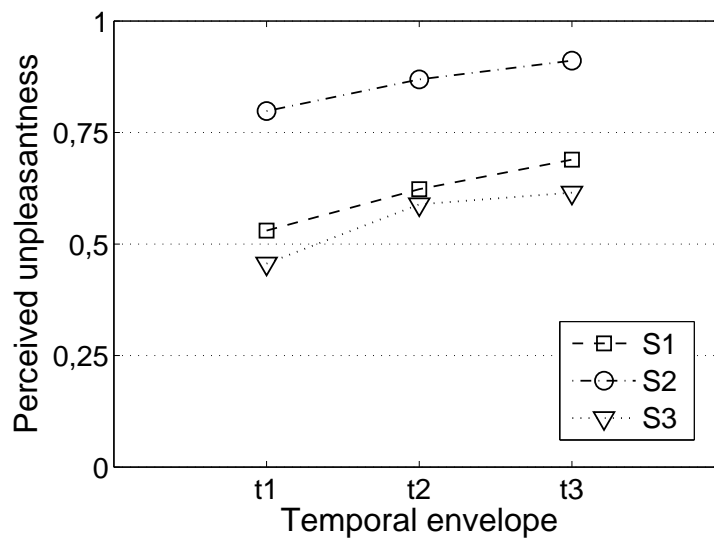


Figure 5: Perceived unpleasantness for the three sound types ($S1$, $S2$ and $S3$) and temporal envelopes ($t1$, $t2$ and $t3$). (The scale is divided into five categories. 0: not at all unpleasant, 0.25: slightly unpleasant, 0.5: moderately unpleasant, 0.75: very unpleasant, 1: extremely unpleasant.)

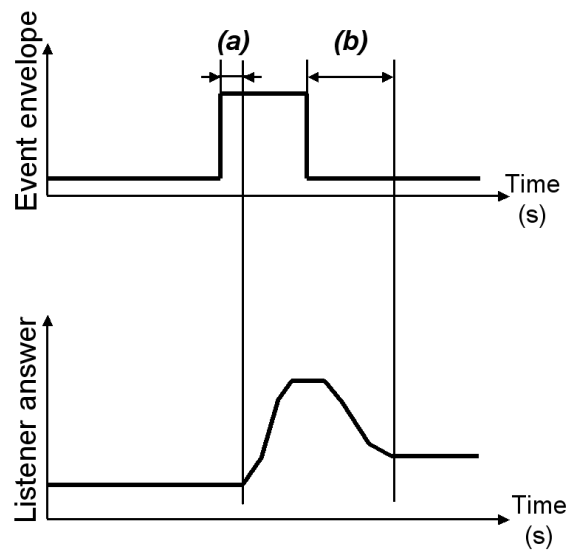


Figure 6: Temporal data definitions: (a) reaction time and (b) relaxation time.