



**HAL**  
open science

## Noise assessment in a high-speed train

Etienne Parizet, Nacer Hamzaoui, Johan Jacquemoud

► **To cite this version:**

Etienne Parizet, Nacer Hamzaoui, Johan Jacquemoud. Noise assessment in a high-speed train. *Applied Acoustics*, 2002, 63, pp.1109-1124. hal-00849433

**HAL Id: hal-00849433**

**<https://hal.science/hal-00849433>**

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.

## Noise assessment in a high-speed train

Etienne PARIZET, Nacer HAMZAOUI, Johan JACQUEMOUD  
Laboratoire de Vibrations et d'Acoustique  
Insa Lyon  
25 bis, avenue Jean Capelle  
69621 Villeurbanne Cédex – France  
Email : [parizet@lva.insa-lyon.fr](mailto:parizet@lva.insa-lyon.fr)

### Abstract

Binaural sound recordings in various positions in a high-speed train have been used as stimuli in listening tests. In a first experiment, sounds were submitted to subjects at their real levels. Loudness revealed to be the most important parameter; it could be very well described by A-weighted levels. In a second experiment, loudness of stimuli were equalised; strong inter-individual differences appeared, which are due to different appreciations of the low-frequency content of sounds. Usual physical or psycho-acoustical parameters failed to correctly described that perceptual parameter; a simple parameter based on the specific loudness calculations was developed and could correctly describe the results.

### Introduction

As compared to noise outside train or inside cars, the perception of noise in trains has not been studied intensively. Hardy [1] focused on the fact that, as noise in train is not stationary, its character is very time-variable (but this is also the case for car noise). For stationary samples, he reported that the use of "Room Criterion" (which has been defined by Blazier [2] for the evaluation of HVAC noise) can represent sound perception, while A-weighted SPL can not. Letourneaux and Guerrand [3] have modified a sound sequence recorded in a high-speed train in order to reduce its level either at some particular frequency peaks, either in

broader frequency bands; their listening tests showed that the latter modifications gave the best perceptual results, especially when the bands were in the low frequency range. They could not model their results by the help of A-weighted level, but the variations of this level within their stimuli were rather small (less than 3 dB(A)). Patsouras et al. [4], focusing on speech intelligibility in trains, checked that an overall level increase of three types of simulated train noise (pink noise filtered in order to increase the level of different frequency bands) gave the same reduction of sound quality, as evaluated by listeners.

Because acoustical environment can vary a lot in a train, recent studies use more open procedures, like questionnaires to passengers (Mzali [5], Khan [6]). Such studies reveal the importance of noise in overall comfort evaluation (Khan reported that 67% of passengers relate that they are annoyed by noise during their journey) and also the fact that claims about noise are not always related to running noise of the train (mobile telephone ringing and children crying are also annoying !).

The goal of this study is to evaluate perceptual prominent factors on sound perception for noises recorded in a high-speed train. For sake of simplicity, it was decided to focus on stationary noises; but the influence of factors like speed train, the nature of the track (i.e. a track specially designed for high-speed trains or an older track) and positions in the train was also taken into account.

## **Stimuli**

Stimuli have been recorded by a dummy head (Bruel&Kjaer) located at different seats in a two-decks train.

Other parameters were also varied :

- Train speed (160, 220 or 300 km/h);

- type of track (a classical one or a new one, specially built for high speed trains).

Different samples (duration : 10 seconds) were selected in the whole set of recordings, in order to vary those parameters. The set of sound samples is presented in table A.

*[ Table A to be inserted around here ]*

A short fading (10 ms linear fade) was applied at the beginning and at the end of each sample in order to avoid disturbing noise.

## **First experiment**

In the first test, sounds have been submitted to listeners, through headphones (Sennheiser HD600), at their real levels, in a quiet and isolated listening room. **The calibration was realised by placing the headphones on the same type of dummy head and checking that the loudness values of sounds were the same as those obtained on the recording sounds.**

### ***Procedure***

We used a classical pair comparison method on a 5 levels preference scale. After hearing each pair, the listener had to indicate the most pleasant sound on a scale consisting of 5 levels ("B is much more pleasant than A", "B is more pleasant than A", "A and B are equivalent", etc...). He could listen to the pair as often as he wants to.

The overall test was run by a Matlab program on a PC computer which established pairs, presented them through the sound card, presented the answering scale on the screen, and stored the listener's answer as an integer between  $-2$  et  $+2$ .

Pairs have been made up by the technique of the Ross series [7], which allows to further off as much as possible two successive presentations of each sound. Beforehand, the 8 samples were randomly arranged, which assured that pairs were presented in a different order for each listener.

The whole set lead to a total number of 28 pairs, plus two used as a training at the beginning of the test.

First of all, the purpose of the study and the task to achieve were presented to the listener. Then the set of samples was presented so as to make the listener used to the context of the test.

The duration of the test was between 15 to 30 minutes, according to the number of repetitions of the pairs needed by the listener.

### ***Listeners***

48 people participated to the test; most of them were students. The average age of the panel was 25 (range from 18 to 55) and the proportions of men and women were equivalent. **Their hearing ability was not measured, but none of them reported any hearing impairment.** Each listener received a gift for his participation.

### ***Results***

#### Analysis of answers

The averaged merit scores are computed by

$$H_n = \sum_{i \neq n} P_{in} \quad (1)$$

where  $P_{in}$  is the preference within pair (i,n) :

$$P_{in} = -\frac{1}{N} \sum_{k=1}^N P_{in}^k \quad (2)$$

$P_{in}^k$  being the answer of the k listener for the {i,n} pair and N the number of listeners.

( $H_n$ ) scores are presented in figure 1. The amplitude of results is large (from -10 to +10, for a possible maximal variation from -14 à +14), which means that inter-individual differences are small.

*[ Figure 1 to be inserted around here ]*

### Preference model

The sound level undoubtedly controls these results, because the variations of the levels between samples are very important (up to 14 dB(A)). Figure 2 presents the relation between averaged preference scores and overall A-weighted level of each sound; the correlation is very good (R= - 0,98).

*[ Figure 2 to be inserted around here ]*

In this case, the use of more accurate indicators (as Zwicker's loudness) is not necessary.

This result is clearly different from those obtained by Hardy [1] and Letourneaux et al. [3], who found that A-weighted sound level did not lead to a good description of the subjective assessment. In the case of the study from Letourneaux et al., level differences between stimuli was quite small (3 dB(A)) as compared to differences in our stimuli (12 dB(A)), which can

explain why A-weighted level was not a good descriptor. The reason is not so clear in Hardy's study (amplitude of 10 dB(A) between stimuli). Indeed, the weak link between A-weighted level and subjective assessment in this study is mainly due to one stimulus (sound 1) whose level is low but which is evaluated as annoying by listeners. The timbre of that sound may be particular, which could explain its bad evaluation.

Here, A-weighted sound level can even correctly describe the preference probabilities ( $P_{ij}$ ) of the possible pairs, as it can be seen in figure 3.

*[ Figure 3 to be inserted around here ]*

This figure emphasises the saturation appearing with strong differences of level : when the difference is greater than a dozen of decibels, the loudest sound is considered as “much less pleasant” by all the listeners.

The relation between level differences and the average preference within each pair can be represented by the following expression :

$$P_{ij} = A \cdot \text{Arc tan} \left( \frac{\Delta_{ij}}{B} \right) \quad (3)$$

where  $\Delta_{ij}$  is the difference of the sound level in dB(A), A and B being constants (the Excel solver finds  $A=-1,76$  and  $B=6,06$ , which gives the curve represented in figure 3).

### Link with physical parameters

Lastly, the predominant physical parameter is the train speed. Figure 4 represents the averaged scores of preference, ordered in an ascending way. Different symbols are used for the different train speeds, (300 km/h : crosses; 220 km/h : black diamonds; 160 km/h : open diamonds) so that the influence of this parameter clearly appear.

*[ Figure 4 to be inserted around here ]*

## Conclusion

The results of this first experiment are what they could be expected : inter-individual variability is very low, loudness is the first order parameter for preference and is related to train speed.

## **Second experiment**

The purpose of the second experiment was to go further than the first order sound parameter identified in the previous experiment.

### ***Procedure***

Previous stimuli have been equalized in loudness so as to suppress the influence of this parameter. This equalization consisted in modifying the overall level of sounds (by multiplying numerical data) in order to adjust **time-averaged** loudness to a value close to 10 sones (computed according to ISO 532 B, implemented in the *Mts Sound Quality* software). Actually, this target value was the average of original sequences loudness. After this equalization, the computed **time-averaged** loudness of stimuli varied between 9,8 and 10,2 sones.

Test procedure was the same as the one which had been used for the previous experiment. The only difference is that each listener carried on two listening tests :

- for the first one, sounds were presented to him through earphones as in the first experiment ;
- for the second one, a complementary sub-woofer (Altec Lansing 1632 A) was used. The goal of this test was to evaluate the importance of very-low frequencies perception through the body, because ordinary sound reproduction of dummy head recordings



(through earphones) get rid of such a body perception. A special device was developed, with a two-ways cross-over filtering at 63 Hz, so that frequencies below this limit were emitted by the subwoofer and frequencies above it by earphones. The relative gains of the different amplifiers were carefully adjusted so as the frequency response of the whole device, when measured at the ear of the listener, was the same as the frequency response obtained in the first test, when the whole frequency range was emitted by earphones. In another study, dealing with truck interior noise, it has been shown that the use of the subwoofer can increase the realism of sound reproduction [8].

Half of the panel began with the first test (only earphones), and the other half with the second one (earphones and subwoofer).

### ***Listeners***

Listeners are those of the previous test with very few exceptions. In this case, the jury was made up of 47 people (from 18 to 55 years old, the average age being 25 and the proportions of men and women being the same).

### ***Results***

#### Analysis of answers

First of all, it was found that averaged results are very close for each of the procedures (with or without the use of the subwoofer), so that they were mixed together (figure 5). This conclusion is different from the one obtained in the case of truck noise, which may be due to the fact that loudness was more important for truck noise (about 25 sones instead of 10 sones for the train noises).

*[ Figure 5 to be inserted around here ]*

We have thus decided to gather the overall results (which makes a set of 94 answers).

Then it can be noticed that the amplitude of averaged merit scores is much less important than in the first experiment (from -2,5 to 4,5, whereas the scores varied between -10 and +10 in the first experiment). Generally speaking, this reduction can be due to three causes :

- more cautious listeners' answers (who wouldn't use the whole scale for their answers), indicating reduced differences between noises ;
- unsteady estimations of each listener, which can lead to circular errors (A is preferred to B, and B to C but C is preferred to A). Such circular errors indicate a difficult test (which can be the case if differences between noises are small) ;
- a large inter-individual variability (preferences are different !).

It seems that these three reasons have to be taken into account in that experiment. First of all, the **root mean squared** value of answers is lower than the one which had been obtained in the first experiment (1,13 against 1,45), indicating a more limited use of the scale. Secondly, the rates of circular errors are higher. These rates have been computed in the following way : if  $P_{ij}$ ,  $P_{jk}$  and  $P_{ik}$  indicate the answers of a listener for the {i,j}, {j,k} and {i,k} pairs, a circular error is detected if

$$\left\{ \begin{array}{l} P_{12} > A \text{ and } P_{23} > A \text{ and } P_{13} < -A \\ \text{or} \\ P_{12} < -A \text{ and } P_{23} < -A \text{ and } P_{13} > A \end{array} \right. \quad (3)$$

A being a fixed level (here  $A = 0$ , which means that an error is not counted if the listener evaluates two of the three noises as equivalent). Then the circular error rate of the listener is computed by :

$$\mathcal{X}_n = \frac{1}{A_t^3} \sum_{1 \leq i, j, k \leq t} \delta_{ijk} \quad (4)$$

where  $\delta_{ijk}$  is equal to 0 or 1 according to equation (3),  $n$  indicating the listener and  $t$  the number of presented sounds [9]. The rate of circular errors, averaged over all listeners is 2.7% for the first experiment and 12.7% in the second one, indicating that this second one is more difficult.

Finally, a clustering of the panel (using the K-means technique) reveals that this panel can be separated in more homogeneous groups. The K-means technique [10] leads to different solutions, according to the number of groups chosen by the user; here, the three groups solution was selected, because it gave two advantages with regard to more important numbers :

- numbers of listeners are similar in the three groups (respectively 35, 24 and 35);
- for 36 out of 47 listeners, both sets of answers (with or without the subwoofer) have been classified in the same group, which indicates a good stability of the perceptive criteria used for the assessment.

The results of these 3 groups are represented in figure 6. In each group, the amplitude of preference score is greater than the amplitude obtained when averaging preference scores over the whole jury, indicating more homogeneous results.

*[ Figure 6 to be inserted around here ]*

This figure shows that :

- tendencies of groups 1 and 3 are similar, with a lower amplitude in group 1 (9.4 in group 1 and 13.6 in group 3);
- group 2 reveals various inversions with regard to group 3. For instance, sounds 2 and 7 are the preferred ones by listeners from group 2, but they are the least appreciated ones by listeners from group 3.

Moreover, it is to be noted that the averaged rate of circular errors, when computed on the third group alone, is significantly lower than the rates computed for the two other groups :

*[ Table B to be inserted around here ]*

As a great number of circular errors leads to a fewer amplitude of preference scores, it can be estimated that listeners from group 1 express similar preferences than those from group 3, but in a less accurate way.

In summary, the classification phase allows the separation of the jury in three groups :

- Groups 1 and 3 (for a total of 70 answers) show similar tendencies, which have been reduced in group 1 because of more important rates of circular errors ;
- Group 2 (24 answers) is composed of listeners with opposed preferences.

It was not possible to explain the composition of the group from any individual criterion; the only known variable criterion of the panel was the sex of the listener, which is not related to any group.

### Preference model

Several classical sound parameters have been computed by Mts Sound Quality software :

- Linear, or weighted (A-B-C) level;
- Sharpness, fluctuation strength and roughness ;
- Tonality ;
- Spectral center of gravity;
- Intelligibility.

It was tried to use these parameters in linear or logarithmic models of experimental merit scores. None of these parameters gave a reliable model for preference on its own. The best

models are obtained with sharpness : in that case, correlation coefficients between estimated and real preference scores are  $-0.64$ ,  $0.74$  and  $-0.39$  for the three groups. This gives an indication about the reason why listeners from group 2 have expressed opposite preferences to other listeners : they seem to appreciate rather noises **with a high frequency content**, while other listeners prefer low-frequency ones.

Also, a multivariate linear regression modelling did not give useful results, with the exception of group 2, whose preference scores can be represented by a linear combination of sharpness and tonality ( $R^2 = 0.86$ , figure 7). **The relation is :**

$$\text{merit score} = -31.1 + 68.9 * \text{Tonality} + 36.2 * \text{Sharpness} \quad (5)$$

**where Sharpness is computed using Zwicker's procedure [11, page 242] and expressed in acum, while Tonality is computed following Terhardt's model [12] and expressed in tu.**

*[ Figure 7 to be inserted around here ]*

So the use of classical sound parameters did not lead to good representations of the results from the three listeners groups.

As spectral balance seems to be an important sound feature, we tried to express it from specific loudness functions of the stimuli. For that purpose, the loudness model from Moore and Glasberg [13] was used. This model differs from Zwicker's one on the following points :

- the assumed transfer function of outer and middle ear;
- the calculation of excitation patterns;
- the critical bandwidth, especially for the frequencies below 500 Hz;
- the computation of specific loudness from the excitation patterns.

The software made available by Moore and Glasberg was used. Its input is the average spectrum of noise expressed in third-octave bands (between 40 Hz and 16 kHz); in this

experiment, the spectrum averaged over the two ears for each frequency band was first computed, because the stimuli were dichotic ones. The output is the overall loudness and the specific loudness curve associated to the spectrum.

These specific loudness curves are shown in figure 8; clear differences between noises can be seen. For some of them (2, 5, 6, 7), specific loudness is maximum in the first critical bands, while others present important specific loudness in higher bands (1,4, 8 for example).

*[ Figure 8 to be inserted around here ]*

The proposed criterion is therefore the ratio of loudness in the bands 0..2 Bark and 13..18 Bark (when expressed in Hz, these bands correspond to 0 to 200 Hz and 2000 to 4500 Hz). In order to express this criterion in Phones, the criterion is defined as "loudness balance" :

$$LB = 10 \cdot \log_2 \left( \frac{S_{0..2}}{S_{13..18}} \right) \quad (5)$$

As overall loudness of sounds are nearly equal,  $LB$  represents the relative contributions of very low and sharp bands of noise to loudness.

Linear regression between  $LB$  and preference scores computed in each group gave correlation coefficients presented in table C; these coefficients express significant relations between  $LB$  and preference scores. Also, the opposition of preference between group 2 on one hand and groups 1 and 3 in the other hand can be seen in table C. This inter-individual variability, as regard to the low-frequency content of sounds, had already been shown in the case of road noise in cars [14].

*[ Table C to be inserted around here ]*

Figure 9 shows the relation between preference scores and estimated scores from each model using  $LB$ .

*[ Figure 9 to be inserted around here ]*

And figure 10 presents the relation between loudness balance differences of two sounds and the preferences within the pairs for the three groups of listeners. Correlation coefficient are respectively 0.65, -0.85 and 0.75, indicating that loudness balance can correctly represent the preference.

*[ Figure 10 to be inserted around here ]*

Of course, this work on its own does not allow to use such models in technical requirements. More work would be needed to check the validity of this model when used in different situations (new type of trains, different speeds, and other listeners). This only shows that sound features that were important for perception were correctly identified.

### Link with physical parameters

Figure 11 presents LB values obtained for the eight stimuli, ordered in an ascending way. Stimuli recorded in the low room are indicated by an open diamond, while those recorded in the upper room are indicated by a black diamond.

*[ Figure 10 to be inserted around here ]*

Sounds 7, 2 and 6, representing the maximal values of the criterion, are also those having the strongest energy in low frequency, which can be linked to an aerodynamic noise being more intense in the high room than in the low room (in the high room, aerodynamic noise transmitted through the roof is an important source in the very low frequency region).

## **Conclusion**

The conclusions of this study are the following ones :

- the first influencing factor of interior noise perception in a high-speed train is loudness, which can be correctly described by the overall A-weighted level of the signal. The

influence of loudness is quite the same for every listener, and loudness is mainly due to the speed of the train;

- when the influence of loudness is eliminated, perception is different among listeners. Most of them (70%) prefer the noise to be in the low frequency range while some other listeners prove opposite preference;
- this second-order sound feature can not be described by existing sound parameters; even sharpness fails to be a good descriptor;
- the examination of specific loudness curves can represent this perception and this can be formalised by an indicator expressing the balance between the contribution to loudness of the lowest critical bands and those between 13 and 18 Bark;
- this indicator shows clear differences between the upper and lower rooms of the train, which may be due to different contributions of noise source in the two rooms.

## **Acknowledgements**

The authors are grateful to the Research Center of the French National Railway Society (Sncf), who submitted the study and allowed the publication of the results.



## Bibliography

- [1] "Measurement and assessment of noise within passenger trains", A.Hardy, JSV 2000 **231**(3), 819 – 829.
- [2] "Revised noise criteria for application in the acoustical design of HVAC systems" Blazier W.E. Noise Control Eng. Journ. 16, 64-73 (1981).
- [3] "Assessment of the acoustical comfort in high-speed trains at the SnCF : integration of subjective parameters", F. Letourneaux, S. Guerrand, JSV 2000 **231**(3), 839 – 846
- [4] "Privacy versus sound quality in high speed trains" C. Patsouras, H. Fastl, U. Widmann, G.Hoelzl, Internoise 1999, 391-396
- [5] "Etude de la qualité du confort acoustique dans les transports ferroviaires : analyse sémantique de questionnaires ouverts" Mzali M., Dubois D., Polack J.D., Letourneaux F., Poisson F., 4<sup>ème</sup> CFA, Lausanne 2000.
- [6] "Determination of sound quality in Swedish passenger trains" M.S. Khan, Internoise 2001
- [7] "Optimum orders for the presentation of pairs in the method of paired comparison" Ross R.T., Journ. Educ. Psychol. **25**, 375-382 (1934).
- [8] "Amélioration du réalisme de la restitution d'enregistrements binauraux par l'utilisation d'une enceinte additionnelle pour les basses fréquences", E. Parizet, M. Mouret, 5<sup>ème</sup> Congrès Français d'Acoustique, 2002.
- [9] "Paired comparison listening tests and circular error rates", E. Parizet, submitted to Acta Acustica / Acustica.
- [10] "Cluster analysis for applications" Anderberg, M. R. , Academic Press, New York, 1973
- [11] "Psycho-acoustics : facts and models", Zwicker E., Fastl H., Springer Verlag, 1990
- [12] "Algorithm for extraction of pitch and pitch salience from complex tonal signals", Terhardt E., Stoll G., Seewann M., J. Acoust. Soc. Am. **71**(3), 679-688 (1973)
- [13] "A model for the prediction of thresholds, loudness and partial loudness" B. Moore, B. Glasberg, T. Baer. (1997), J. Audio Eng. Soc. **45**, 224-240
- [14] "Car road noise annoyance : significant timbre parameters and inter-individual variability", E. Parizet, S. Deumier, E. Milland, Forum Acusticum 1996, Acta Acustica **82**, S-216.

**Tables**

Sound	Speed (km/h)	Track (Conventional / High Speed)	Room (Low / High)	Position in the room (1 – 2 – 3)
1	300	HS	H	1
2	160	C	H	1
3	220	HS	L	2
4	300	HS	L	2
5	160	C	L	2
6	220	C	H	3
7	160	HS	H	3
8	220	C	L	3

Table A : stimuli

Group 1	Group 2	Group 3
14,9	15,1	8,7

Table B : average rates of circular errors for the three groups of listeners (%)

Group 1	Group 2	Group 3
- 0.68	0.85	- 0.78

Table C : correlation coefficients between each group's result and loudness balance criterion

### Figure legends

Figure 1 : Averaged scores of preference (test 1)

Figure 2 : preference scores and overall levels (test 1)

Figure 3 : differences of A-weighted levels and average preference within pairs

Figure 4 : averaged scores of preference by increasing order of this one (test 1)

Figure 5 : averaged preference scores of experiment 2

Figure 6 : averaged score of each group

Figure 7 : preference scores from group 2 and estimated scores from a linear combination of sharpness and tonality

Figure 8 : specific loudness of the stimuli (test 2)

Figure 9 : preference scores and loudness balance for the three groups of listeners from test 2  
(left : Group 1, middle : Group 2, right : Group 3)

Figure 10 : difference of criterion and preference within pairs for the three groups of listeners from test 2 (left : Group 1, middle : Group 2, right : Group 3)

Figure 11 : *LB* values of stimuli