GENIFER FEASABILITY STUDY - ASSESSING RAILWAY NOISE INSTANTANEOUS ANNOYANCE WITH IN-SITU LISTENING TEST

Manuel Hellot[1*] **Robin Mafféïs**[1] **Matthieu Sineau**[1]
Anne-Sophie Evrard[2] **Jean-Philippe Regairaz**[3]
[1] Bruitparif, observatoire du bruit en Ile-de-France
32 boulevard Ornano, 93200 Saint-Denis, France
[2] Univ Lyon, Univ Gustave Eiffel, Univ Claude Bernard Lyon 1, UMRESTTE, UMR_T9405, F-69500 BRON, France
[3] SNCF-Réseau, Direction Générale Industrielle et Ingénierie, Direction Technique du Réseau
15/17 rue Jean-Philippe Rameau, 93210 La Plaine Saint-Denis, France

ABSTRACT

Railway noise is one of the potential obstacles to the development of this low carbon means of mobility, and it is necessary to study its acoustic parameters in the involved annoyance for better characterization. In this frame, the GENIFER feasibility study, aimed at improving knowledge of the acoustic factors involved in instantaneous annoyance due to railway noise. This study implemented a first phase of instantaneous annoyance ratings using an electronic device with 62 people living near railway lines, followed by a second phase involving in-home listening panels with 33 of these people. During this second phase, and from binaural digital audio recordings of a representative sample of trains, they were asked to rate the instantaneous annoyance caused by passing trains. The results showed that sound indicators such as SEL or LAmax were the best predictors of instantaneous annoyance for non-equalized audio sounds, while psychoacoustic indicators such as tonality and roughness were the best predictors for equalized sounds. A hierarchical classification highlighted two groups of 'raters' with age as a potential distinguishing factor, with younger individuals assigning lower instantaneous annoyance ratings than other individuals. Finally, the two methods of assessing instantaneous annoyance were compared and showed different results.

*__Corresponding author__: manuel.hellot@bruitparif.fr.

Keywords: *Noise, railway, annoyance, listening test, descriptors, indicators, instantaneous.*

1. INTRODUCTION

Although railway transportation is a low carbon means of mobility, its development is held back in part by the noise pollution it can cause. Authorities need to understand the mechanisms that can affect annoyance due to railway noise to adapt current regulations and train operations as effectively as possible. In this perspective, the GENIFER feasibility study [1] investigated the instantaneous annoyance caused by passing trains to better understand it, with the aim in the long term, of developing new noise indicators more correlated with the annoyance perceived by residents than energetic indicators such as Leq. Following an initial survey in 2023 aimed to collect the instantaneous annoyance associated with trains passing by using a remote control [2], a second phase of survey was carried out in 2024, based on listening panels at participant's homes

One objective of this second phase was to evaluate event-based indicators correlated with instantaneous annoyance under controlled listening conditions (via headphones), and to compare these results with those obtained in the first phase [2], where annoyance was assessed in situ using a remote control during real noise exposure at participants' homes.

2. MATERIAL AND METHODS

2.1 Sound samples

Sound recordings were conducted using a Head Acoustic SQadrigra II system in the immediate vicinity of the railway tracks and analyzed with Arthemis© software. A total of six hours of recordings, divided into three separate sessions, were collected to obtain approximately 10 to 15 train samples, with two samples per train type. The recordings were made at a distance of 18 meters from the railway tracks in front of the nearest houses. The type of train was defined as follows: urban passenger trains (RER), old generation regional trains (CORAIL), new generation regional trains (TER_NG2N), new generation regional short trains (TER_AUTORAIL), and freight (FRET). The study site contains 5 train tracks (ordered north to south, V2B, V2, V1, EV1 and V1).

The six-hour recording dataset was filtered to retain only 11 representative sound samples, selected based on the preferential tracks used by each train type and their $L_{Aeq,Tevt}$ noise levels [2]. Specifically, the selected samples had $L_{Aeq,Tevt}$ values close to the median noise level measured at the study site by the medusa sensors [2]. These samples are listed in Table 1, with two samples each for TER_NG2N, CORAIL, and FRET trains, and five samples for RER trains. In the first part of the test, sound sample durations were standardized to focus on sound characteristics. Freight train noise events tended to be longer than 30 seconds, so their sound samples have been shortened to be approximately 15 seconds. Unfortunately, no samples of new-generation regional short trains (TER_AUTORAIL) could be retained after filtering, as this type of train ran infrequently on the site, and no sample was obtained without interference from other noises.

Table 1. Characteristics of the non-equalized noise samples

Sample ID	$L_{Aeq,Tevt}$ dB(A)	L_{Amax} dB(A)	$\Delta L_{Ceq,Tevt}$ vs $L_{Aeq,Tevt}$
RER_V2B_10h59	78.2	84.4	2.7
RER_V2B_15h57	72.5	78.8	6.5
RER_V1B_16h04	62.6	67.2	3.0
RER_V1B_12h06	63.8	67.6	4.7
RER_V1B_16h35	71.3	77.6	0.7
TER_NG2N_V1_15h34	71.6	77.4	3.8
TER_NG2N_V2_16h25	78.8	84.8	2.5
FRET_V2B_15h12_coupe_2	76.8	81.4	8.6
FRET_V1B_15h44_coupe	71.4	74.7	11.1
CORAIL_V2_10h17	81.3	85.6	0.9
CORAIL_V2_14h44	80.4	86.2	0.4

In the second part of the test, the same 11 sound samples have been equalized to identical $L_{Aeq,Tevt}$ levels (Leq equalization) to isolate and evaluate how their frequency components influenced annoyance perception, independent of overall noise level differences. Equalizing sounds on certain parameters, such as sound level or duration, allows to limit the influence that their variations could have on annoyance perception, to focus on other acoustic characteristics such as spectral content in this case.

2.2 Participants

Overall, 33 participants from the first phase of the GENIFER study took part in the listening test. Their mean age was 51 ± 14 years with 16 women included. Furthermore, 15 participants had been living in their current residence for less than 10 years at the time of the test. Moreover, participants were classified into three train noise exposure groups based on modeled Lden levels: moderate (<54 dB(A), n=8; intermediate (54-63 dB(A), n=11; and high (>63 dB(A), n=14).).

2.3 Test procedure

The test, conducted under the Head Acoustics SQALA interface, began with a presentation of the objectives of the study and the test procedure. Participants then completed an ear habituation phase, a mandatory step in listening-based tests to familiarize them with the headphone-based listening conditions. This was followed by an introductory scoring phase, designed to help participants understand the interface operation. The first phase of scoring was conducted with non-equalized audio samples which were played and rated using a scale of instantaneous annoyance from 1 to 10, 10 being the maximum annoyance value. The participant could listen to each sound as many times as desired.

At the end of this rating phase, a short interview was conducted to collect participant's perception about railway noise, about the survey and about the two methods used to collect the instantaneous annoyance (remote control vs listening test). Finally, a final scoring phase was carried out with the equalized sounds samples. The duration of the whole test was about 45 minutes. To assess the consistency of the annoyance ratings, intra-participant and inter-participant comparisons were performed. No significant differences were observed between individual judgments or across participants. Consequently, all instantaneous annoyance scores were retained for analysis.

2.4 Statistical methods

Statistical analyses were conducted to examine the relationships between instantaneous annoyance ratings and acoustic descriptors. All computations were performed using Arthemis© software, which provided both classical acoustic indicators (e.g., LAeq, LAmax) and psychoacoustic parameters derived from the Hearing Model developed by Head Acoustics [5]. Furethermore, the following indicators were calculated:

- Impulsiveness: it measured rapid and strong fluctuations in sound level, expressed in impulsiveness units (iu).
- Kurtosis: it represented the peakedness of the sound level distribution, defined as the fourth-order moment of a standardized variable.
- Speech Interference Level (SIL): it corresponded to the unweighted arithmetic mean of octave bands from 500 Hz to 4 kHz.
- Di annoyance index : it corresponded to pyschoacoustic annoyance base on Di et al work [4]

To assess correlations between instantaneous annoyance scores and acoustic descriptors for non-equalized sounds, Pearson's correlation coefficient was used.

Additionally, the distribution of annoyance ratings was analyzed using hierarchical clustering, specifically Ward's method, to identify potential subgroups among participants. A chi-squared test of independence was then performed to determine whether any significant associations existed between clusters and participant characteristics.

For equalized sound samples (i.e., sounds normalized to the same $L_{Aeq,Tevt}$ level), Pearson's correlation was again applied to compare annoyance scores with acoustic indicators from the Arthemis© Hearing Model.

3. RESULTS

3.1 Correlation between annoyance scores and acoustic descriptors for non-equalized sounds

The results showed that instantaneous annoyance scores were highly correlated (R > 0.9), see figure 2, with noise level descriptors such as L_{Amax} or $L_{Aeq,Tevt}$ [3].

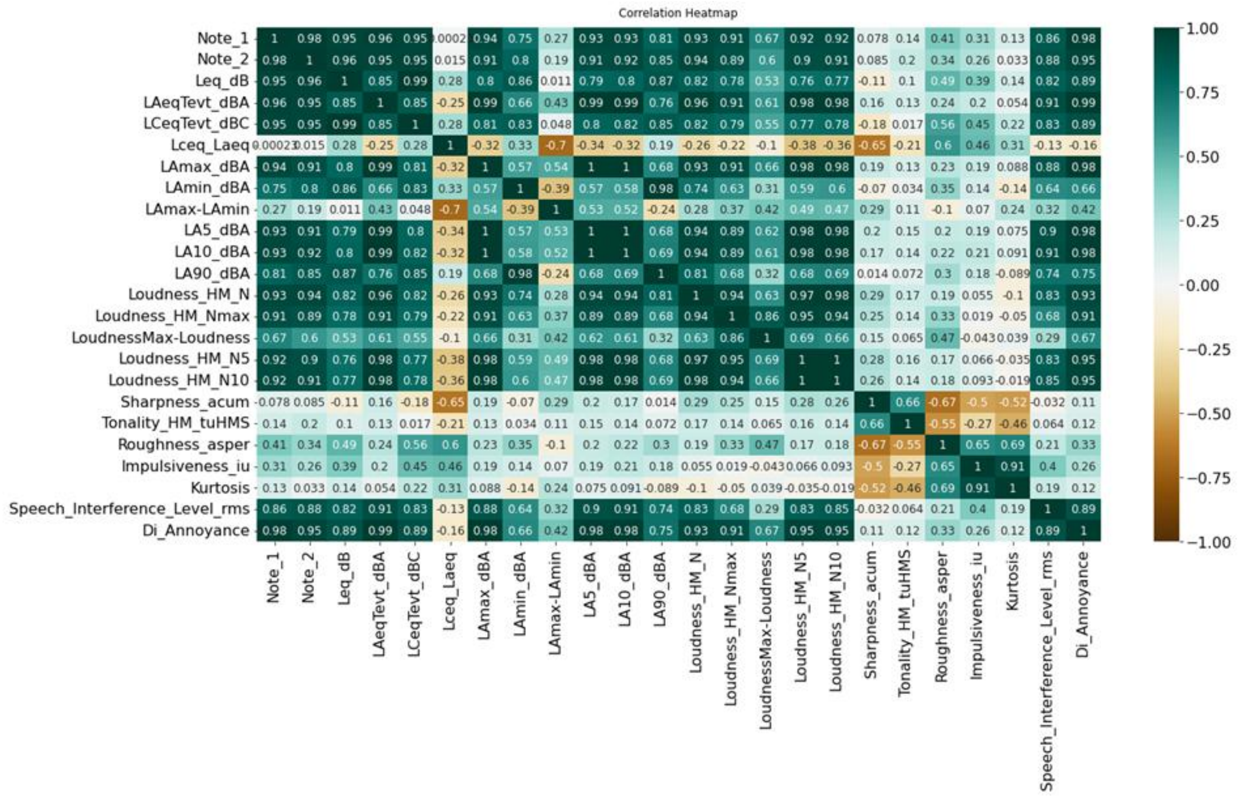

Figure 2. Correlation heatmap between instantaneous annoyance scores and acoustic variables, for non-equalized sounds.

3.2 Instantaneous annoyance ratings distribution and clustering for non-equalized sounds

The distribution of the instantaneous annoyance scores for the non-equalized sounds (figure 3) showed a wide range of scores. Scores that can go up to the maximum scores of 9 to 10. These effects, known as the 'excessive demand effect' and the 'ceiling effect', indicated the existence of groups across individuals.

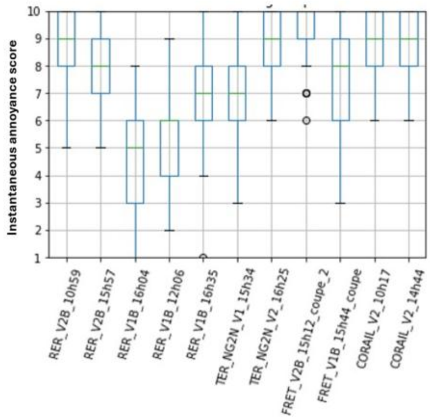

Figure 3. Boxplot for instantaneous annoyance ratings by non-equalized sound samples for all 33 participants.

A hierarchical classification using Ward's method identified two distinct groups in terms of instantaneous annoyance rating (figure 4). A chi-square test of independence showed that only "age" emerged as a potential dependent factor in the formation of the groups (Table 2), in addition to the method of rating.

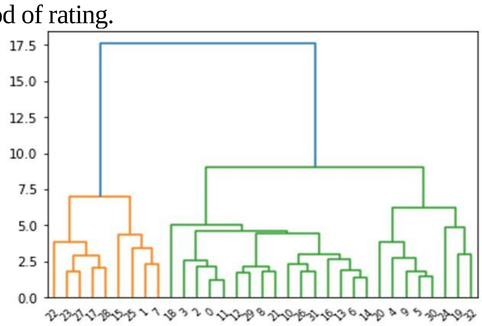

Figure 4. Dendrogram after hierarchical classificiation (with the 33 participants)

Table 2. Results of independence test between clusters and variables

Variable	Chi-Square (χ^2)	Degrees of freedom (df)	p-value
Age	**9.070**	**2**	**0.01**
Assessment of neighborhood	3.621	2	0.16
Area of noise exposure	3.053	2	0.22
Year lived in (*more than 10 years*)	1.223	1	0.27
Weinstein Noise Sensitivity Scale WNSS sensitivity	0.901	4	0.92
Global noise long term annoyance	1.708	2	0.43
Railway noise long term annoyance	3.905	4	0.42

Cluster 1 brings together the youngest participants on the panel (figure 5). This group tended to give lower instantaneous annoyance ratings than the other participants. Similarly, the distribution of the instantaneous annoyance ratings given for the various sound samples by the participants in cluster 1 appeared to be less widespread and less dispersed than that of the second cluster (Figure 6). Thus, it appeared that for this second cluster, certain sounds seemed to be 'the subject of debate' in terms of instantaneous annoyance, in particular the sounds from the RER trains running on track V1B as well as one of the FRET trains.

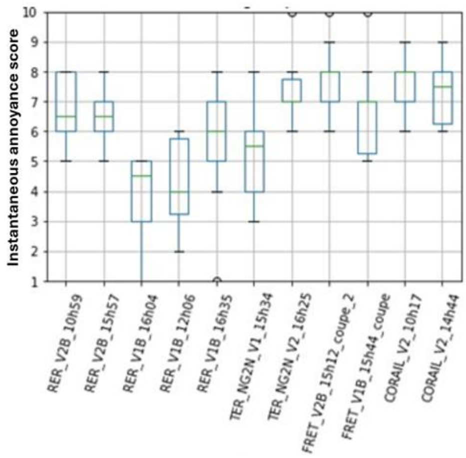

Figure 5 Comparison of the instantaneous annoyance scores per sound sample – Cluster 1 : youngest participants

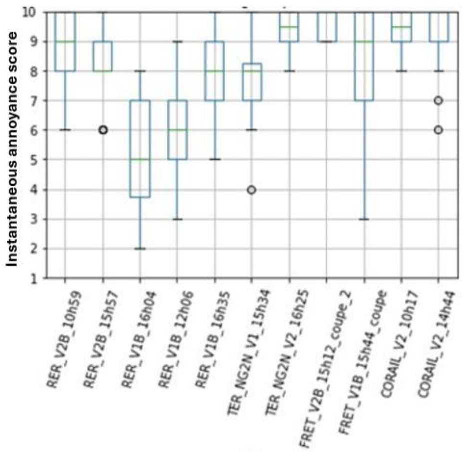

Figure 6. Comparison of the instantaneous annoyance scores per sound sample – Cluster 2 : oldest participants

3.3 Results for equalized sounds

The distribution of instantaneous annoyance score is shown in figure 7. In this situation, the correlation between the instantaneous annoyance scores and the acoustic descriptors decreases significantly compared to non-equalized sounds. For the $L_{Aeq,Tevt}$, the correlation with instantaneous annoyance scores was 0.96 for non-equalized sounds, falling to -0.3 with equalized sounds (Figure 8).

When the levels of the sound samples were equalized, the acoustic parameters best correlated with instantaneous annoyance are then sharpness and tonality, two psychoacoustic indices associated with the spectral content of sounds (correlation coefficient of 0.4 for tonality and 0.7 for sharpness).

Figure 7 presents the instantaneous annoyance ratings of the same train pass-by events under both non-equalized and equalized listening conditions. The equalized condition demonstrates significantly reduced inter-rater variability, indicating that participant's annoyance judgments became more consistent when level differences were eliminated.

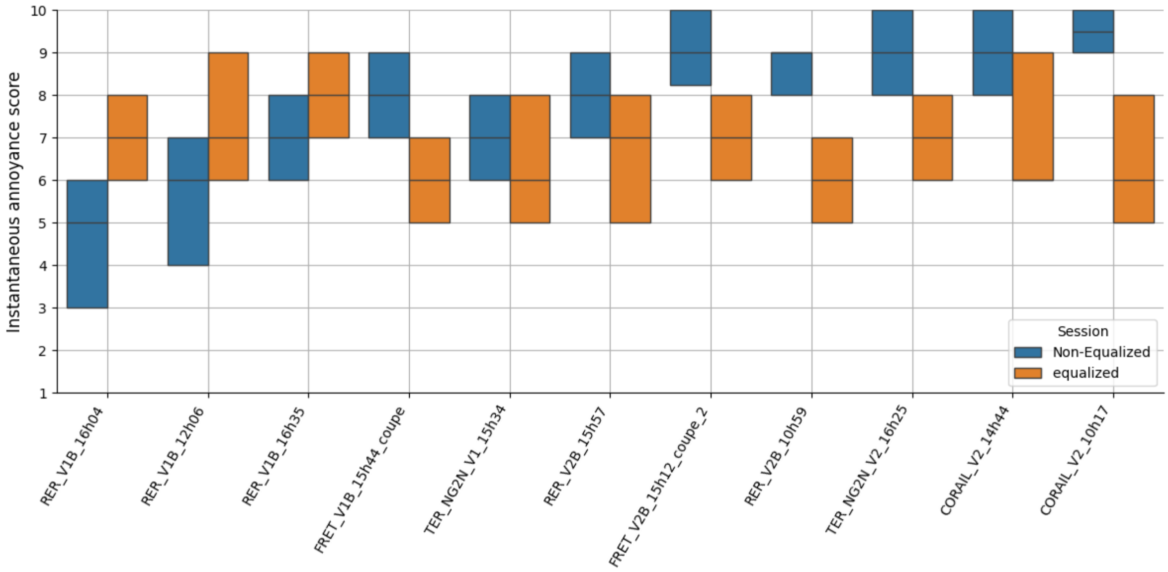

Figure 7. Boxplots of the distribution of instantaneous annoyance scores across different train types for equalized and non-equalized conditions. Only the interquartile range (IQR) is displayed, without individual outliers. The train audio samples are sorted by $L_{eq,evt}$ from quietest to loudest for the non-equalized sessions

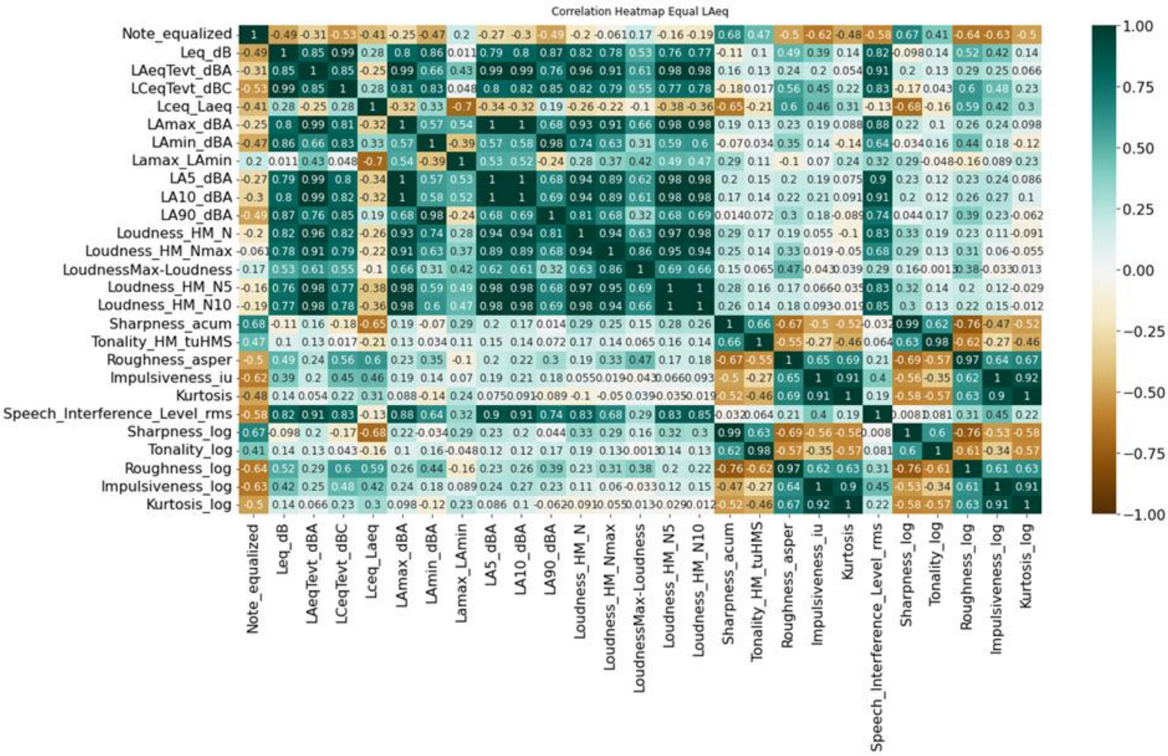

Figure 8. Correlation heatmap between instantaneous annoyance scores and acoustic variables, for non-equalized sounds.

4. DISCUSSION AND CONCLUSION

Correlation between annoyance scores and acoustic descriptors for non-equalized sounds showed that, when listening to accurate sound recordings through headphones, instantaneous annoyance scores are highly correlated ($R > 0.9$) with noise level descriptors such as L_{Amax} or $L_{Aeq,Tevt}$ [3].

This suggests that, for events of comparable duration, sound level is the dominant predictor of instantaneous annoyance, with spectral characteristics playing a comparatively minor role. Furthermore, when the samples are equalized, the variance in annoyance ratings decreases significantly, indicating that frequency-based features (e.g., tonality, spectral shape) have much less influence on perceived annoyance than sound level.

These results are in line with the literature, in which instantaneous annoyance determined by listening to headphones or in laboratory are mainly explained by noise level [4].

Classification seemed to show two distinct groups of “raters”, with age as a potential factor in distinguishing the groups. The youngest participants of the panel tended to give lower instantaneous annoyance scores, with a narrower distribution, than the older participants. This fact has already been observed in other studies [6-7], these showed that the 40-50 and over-60 years old age groups tended to express a greater noise annoyance than other age groups.

An initial assessment of instantaneous annoyance was carried out, in the framed of the GENIFER feasibility study, by the same participants using, at home, a connected remote control [2]. The two methods of assessing instantaneous annoyance - using the remote control in real conditions of exposure at home - and listening to aurally accurate recordings sound samples on headphones – aimed to collect instantaneous annoyance linked to passing trains noise for different conditions and using different means. From the point of view of their implementation, both methods have their advantages and disadvantages. The "remote control" method does not allow a control of the rating conditions and which trains will be rated, but it is carried out in the real conditions of railway noise exposure of participants, which may suggest that it can provide more realistic assessments of the annoyance felt. The "headphone method" allows a control of the rating conditions and which trains will be rated. On the other hand, it places participants in less representative conditions of their railway noise exposure as they experience it at home. The two methods gave different results as shown in Figure 9.

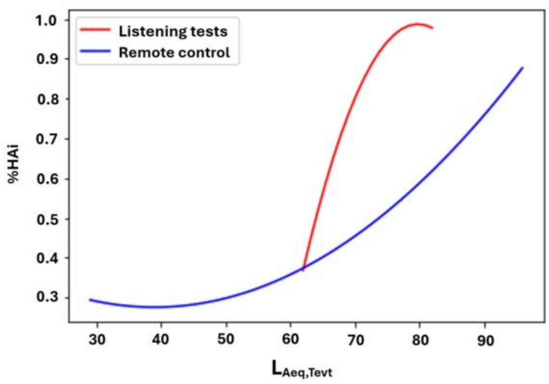

Figure 9. Comparison of the two methods of instantaneous annoyance collecting, with a remote control in the field and with listening tests through headphones. %HAi: percentage of instantaneous annoyance scores exceeding 7

Instantaneous annoyance scores with listening test through headphones were significantly more correlated (> 0.9) with standard acoustic descriptors such as $L_{eq,evt}$, SEL or L_{max} than the instantaneous annoyance scores given "in real-life conditions" with the remote control. The proportion of people highly "instantaneously" annoyed (%HAi i.e instantaneous annoyance scores exceeding 7) was significantly higher when listening to headphones than in a real situation, for equivalent railway noise events.

Listening with headphones is quite similar to listening panels usually conducted under laboratory conditions. The stronger correlation between instantaneous annoyance and sound level is likely due to the greater standardization of the experimental conditions. In real conditions, subjective and environmental factors modulate the relationship between sound level and perceived annoyance.

This study suggests that both methods (listening test and real-world testing) are valuable and complementary. However, findings from standardized listening tests alone may not fully capture annoyance perception, as contextual factors in real environments seem to play a significant role.

5. FUNDING

This project was supported by the French National Research Program for Environmental and Occupational Health (2022-EST-182) of ANSES (French Agency for Food, Environmental and Occupational Health & Safety).

6. ACKNOWLEDGMENTS

The authors would like to thank Nicolas Chouard and Sarah Gagnier-Pannetier from Head Acoustics for their assistance and for the loan of Sqala and Artemis software. They thank the inhabitants of Savigny-sur-Orge for their active participation and involvement in this survey.

7. REFERENCES

[1] M. Sineau, "Improve knowledge of acoustic factors involved in railway noise annoyance", Euronoise Forum Acusticum (Turin, Italia), 2023

[2] M. Sineau, "Improve knowledge of acoustic factors involved in railway noise annoyance - first results of a pilot field survey of sixty-two local residents", Internoise (Nantes, France), 2024

[3] Conseil National du Bruit "Avis du Conseil National du Bruit du 5 octobre 2020 sur les pics de bruit ferroviaires", 2020

[4] Guo-Qing Di et al. "Annoyance and activity disturbance induced by high-speed railway and conventional railway noise: a contrastive case study". Environmental Health 13 (2014), p. 1-10, 2022

[5] Roland Sottek. "A hearing model approach to time-varying loudness". In: Acta Acustica united with Acustica 102.4 (2016), p. 725-744.

[6] Pascal WM Van Gerven et al. "Annoyance from environmental noise across the lifespan". In: The Journal of the Acoustical Society of America 126.1 (2009), p. 187-194.

[7] Keiji Yagawa et al. "Effects of The Number of Noise Events from SHINKANSEN Railway on Annoyance". In: International Congress on Acoustics (ICA Aachen, Germany), 2019.

[8] François Husson, Julie Josse et Jerome Pages. "Principal component methods-hierarchical clustering-partitional clustering: why would we need to choose for visualizing data". In : Applied mathematics department 17 (2010).